telektronikk 3/4.07

Identity Management

Telektronikk
Volume 103 No. 3/4 – 2007
ISSN 0085-7130

Editor:
Per Hjalmar Lehne
(+47) 916 94 909
per-hjalmar.lehne@telenor.com

Editorial assistant:
Gunhild Luke
(+47) 415 14 125
gunhild.luke@telenor.com

Editorial office:
Telenor R&I
NO-1331 Fornebu
Norway
(+47) 810 77 000
telektronikk@telenor.com
www.telektronikk.com

Graphic design:
Design Consult AS (Odd Andersen), Oslo

Layout and illustrations:
Gunhild Luke and Åse Aardal,
Telenor R&I

Prepress and printing:
Rolf Ottesen Grafisk Produksjon, Oslo

Circulation:
4,000

Networks on networks

Connecting entities through networks – in technological, societal and personal terms – enables telecommunication. Networks occur on different levels, form parts of larger networks, and exist in numerous varieties. The artist Odd Andersen visualises the networks on networks by drawing interconnected lines with different widths. Curved connections disturb the order and show that networks are not regular but are adapted to the communication needs.

Per H. Lehne, Editor in Chief

Contents

Identity Management

Kaleidoscope

Guest Editorial

DO VAN THANH

Do Van Thanh is Senior Research Scientist in Telenor R&I

As human beings, we are quite concerned about our identity because it tells who we are and how we want to be perceived by other people. We want to be proud of ourselves and therefore want to keep our identity. Each person has a unique identity. Persons having a problem with their identity are really in big trouble. To have several identities is a malfunction of the brain, in psychiatry called Dissociative Identity Disorder (DID) that needs proper care.

Unfortunately, most people in the modern society are forced to have several identities. Indeed, in the current digital age, our physical world is suddenly extended with a cyber world that keeps growing. From one identity in the real world, we now have a multitude of new identities for this new imaginary world. Some identities are given to us by entities like clubs, alumni groups, etc. Some identities reflect our true identity. Some reveal our aspirations, our dreams, our heroes, etc. On average, a person has five different identities: one for their job's IT system, one for their home broadband connection, one for their mobile subscription, one for their online bank and one for their hobby club. It is hence not surprising that we have an identity management problem. Each identity consists of a user name and password that we have to remember by heart. To maintain the same level of security, we are asked to change our passwords quite often. The simple passwords are not accepted and we are asked to choose long and complicated passwords consisting of lower case and upper case letters and numbers. To write down the password on a piece of paper is not a solution since it constitutes a security breach. The worst thing is that the number of login names and passwords keeps increasing and will be soon unmanageable. The demand for sound and user-friendly identity management systems is obvious.

But, the notion of identity management is not clearly defined, nor is it possible to go and buy a ready-for-use identity management system. It is confusing what an identity management system should comprise. This is not surprising if we contemplate the evolution of computer systems from stand-alone mainframes with their own identity management to computer network systems with external connections. The identity management must evolve at the same pace and meet all the new requirements of the more complex system. The goals of this *Telektronikk* issue are to clarify the notion of identity and to shed light on the field of identity management. It could serve as a clear and concise overview of identity management and its usages.

This *Telektronikk* issue starts with a light review of the notion of identity, from identity of things via personal identity, citizen identity, to digital identity. Next, the concept of identity management (IdM) is studied carefully. All the IdM components are listed and explained. The standards, systems and standardization bodies related to identity management are enumerated and described in one paper, allowing the reader to make further investigation if wanted. Identity management is also explained thoroughly in a paper with a focus on mobile GSM-based systems.

The two current major identity management systems, namely the Liberty Alliance specifications and the Microsoft CardSpace are introduced through two papers written by experts in the respective organizations. The need for a trusted electronic ID is explained in the succeeding paper. Since digital identity is exposed for theft it should be stored securely in smart cards which are portable tamper resistant cryptographic devices. This is examined thoroughly in a paper on Smart cards and digital identity. The next two papers present implementation and usage of identity management systems. One describes the building of a federated identity for education, the other identity federation in a multi Circle-of-Trust constellation.

Without a well balanced and sound business model, the identity management can never be a reality. The next paper addresses this issue by presenting the business scenarios that are both attractive and realistic to telecom operators. The next three papers focus on one central function of identity management, namely authentication. The first one describes the usage of the GSM SIM card in the sign-in to Internet applications using the Liberty Alliance specifications. The second one extends the SIM authentication to CardSpace. The third one reviews all the authentication schemes by mobile phone.

IP Multimedia Subsystem (IMS) is on the way to becoming deployed for both mobile and fixed environments. It is hence necessary to have a way of unifying the different mobile and fixed identities that the user has. One paper is fully devoted to this issue. Last but not least, the access control and privacy facilitated by a role-based identity management system is presented in a paper.

ISSN 0085-7130

For the elaboration and realisation of this *Telektronikk* issue, tremendous efforts and sacrifices have been made. I would take the opportunity to express my gratitude for all the trust and support that I received from the contributing authors. It has been both a pleasure and an honour to work with you. I would like also to thank *Telektronikk*'s editor and staff for their great assistance. Finally, I wish all the readers a great time with this *Telektronikk* issue on Identity Management.

Enjoy your reading!

Prof. Dr. Do Van Thanh obtained his MSc in Electronic and Computer Science from the Norwegian University of Science and Technology and his PhD in Informatics from the University of Oslo. In 1991 he joined Ericsson R&D department in Oslo after seven years of R&D in Norsk Data, a minicomputer manufacturer in Oslo. In 2000 he joined Telenor R&I and is now in charge of a Eureka project called Mobicome that focuses on IMS in a fixed mobile environment. He also holds a professorship at the Department of Telematics at the Norwegian University of Science and Technology in Trondheim. He is the author of over 100 publications at international conferences and journals. He is also the inventor of 19 patents and a dozen pending ones.

email: thanh-van.do@telenor.com

ISSN 0085-7130

The Ambiguity of Identity

DO VAN THANH, IVAR JØRSTAD

"To be or not to be" – Shakespeare

Do Van Thanh is Senior Research Scientist in Telenor R&I

Although identity is a notion related to the existence of all objects on Earth, including human beings and the fundament for the understanding of nature it is rather diffuse and controversial. The goal of this paper is to give an easy but clear elucidation of the notion of identity. It starts with fundamental definitions of identity in logic. The philosophical paradoxes of identity are summarized to illustrate the complexity of the notion of identity. Personal identity is explained in a concise way. Next, the definition of citizen identity is explored. Last but not least, the notion of digital identity is introduced and analysed thoroughly.

Ivar Jørstad is CEO of Ubisafe AS

1 Introduction

Everybody is equipped with the ability to ascertain whether an object is the same or different from other objects. Identity is the relation that states the sameness or identicalness of an object. It is the fundament for reasoning and understanding. It allows individuals to position themselves and to define the relations with other objects in the environment. Identity is a notion that is broadly and intuitively used and its meaning seems to be clear to everyone. Unfortunately, it is not always the case. Indeed, in the same place and time, it is evident that an object is identical to itself but it is not always easy to determine that an object is the same in different places and time. There are also different definitions and opinions about identity which are confusing or even contradictory. The goal of this paper is to give an easy but clear elucidation of the notion of identity. It is not meant as philosophical dissertation about identity but a brief explanation of identity that is useful for understanding the digital identity considered in this magazine issue.

2 Identity of Things

2.1 Identity in Logic

Identity is related to the existence of objects, their uniqueness and distinctness from other objects. It is the fundament for the human being's understanding of nature at the same time as it is quite intuitive [1]. Indeed, perception begins when one recognizes something.

It was first formalised by Aristotle's ***Law of Identity*** as:

A is A – for any A – Everything is itself

At first glance, the Law of Identity appears to be obvious and meaningless because every object is of course itself. It constitutes, however, together with two other laws, the fundaments of formal logic.

Aristotle's ***Law of Contradiction*** states:

Not (A and Not A) – Nothing can both be and not be

The ***Law of Excluded Middle*** states:

A or Not A – Everything must either be or not be

By combining the three laws the following statement can be deduced:

A is not Not A – for any A

This statement is quite broadly and intuitively practiced in the recognition of objects. Indeed, to recognize an object it may be easier to make sure that it is not something else and something else is not this object.

Identity is a ***reflexive*** relation since for every A A is A.

It is ***symmetric*** since for every A and B if only and if A is B then B is A.

It is ***transitive*** since for every A, B, C if A is B and B is C then A is C.

Being reflexive, symmetric and transitive, identity is an **equivalence** relation. This view of identity is the "classical view" characterizing identity as the equivalence relation which everything has to itself and to nothing else [6].

The statements mentioned above are useful in the intuitive recognition process but are not rigorous enough for a systematic recognition. More formal laws are required.

ISSN 0085-7130

The **Identity of Indiscernibles**, also called Leibniz's Law is a principle formulated by Wilhelm Gottfried Leibniz in his Discourse on Metaphysics [2], [3] and states that:

No two distinct substances exactly resemble each other.

Which is again understood as:

No two objects have exactly the same properties.

This law consists actually of two principles that must be distinguished (two equivalent versions of each are given in the language of the predicate calculus) [4] [5].

Principle 1: The Indiscernibility of Identicals

For any x and y, if x is identical to y, then x and y have all the same properties.

$\forall x \forall y[x = y \rightarrow \forall P(Px \leftrightarrow Py)]$

For any x and y, if x and y differ with respect to some property, then x is non-identical to y.

$\forall x \forall y[\neg \forall P(Px \leftrightarrow Py) \rightarrow x \neq y]$

The indiscernibility of identicals states that if two objects are *numerically identical* (the same one), they must have the same properties, i.e. *qualitatively identical*. Numerical identity must imply qualitative identity.

Principle 2: The Identity of Indiscernibles

For any x and y, if x and y have all the same properties, then x is identical to y.

$\forall x \forall y[\forall P(Px \leftrightarrow Py) \rightarrow x = y]$

For any x and y, if x is non-identical to y, then x and y differ with respect to some property.

$\forall x \forall y[x \neq y \rightarrow \neg \forall P(Px \leftrightarrow Py)]$

The identity of indiscernibles says that if two objects have all the same properties, i.e. qualitatively identical, they are numerically identical. Qualitative identity implies numerical identity.

2.2 The Paradoxes of Identity

Combining the two principles, numerical identity is equivalent to qualitative identity. The law of identity seems to be both simple and reasonable but it gives rise to a great deal of philosophical perplexity [2]. The controversy of the law is usually depicted by paradoxes – arguments that apparently derive self-contradictory conclusions by valid deduction from acceptable premises.

2.2.1 Problems with the Indiscernibility of Identicals Principle

a. The Paradox of Time and Change

One of the biggest paradoxes is the problem of changes through time. Consider two photographs of John. In one, John is a little boy and in the other, he is an old man with grey hair. It is absolutely certain that this is the same person. There is hence a violation of Principle 1 because John as the same person must have all the same properties.

Numerical identity does not imply qualitative identity. This paradox shows that the Indiscernibility of Identicals Law cannot be used to determine whether it is the same object in time and space.

Different remedies have been proposed. The most popular is that simple properties such as having or lacking grey hair are actually relations to time. John always has the property of grey hair which was not true when he was young but became true with time. Another popular explanation is that John is an object which is extended over time as well as space. The little boy and the old man are distinct temporal parts of the whole temporally extended John.

2.2.2 Problems with the Identity of Indiscernibles Principle

a. The Symmetric Universe

Let us consider a universe which is perfectly symmetric and consisting of three qualitatively identical spheres A, B and C. Each of them has the same distance, 2 units to the two others [2]. In this case, there is no property which distinguishes any of the spheres from any of the others. According to Principle 2, the three spheres are said to be identical when they are not.

This example is inspired by an objection to Leibniz' Law by Marx Black (1933) who uses a hypothetical in which he conceives two distinct spheres having exactly the same properties.

b. The Infinity Problem

Objects consist of an infinite number of properties which may have infinitely many values and hence be indeterminate. Principle 2 faces here a serious indeterministic problem at two levels. First at property level; an object may have an infinite number of properties and two objects may be "wrongly" concluded as identical based on a given number of considered property. Second, two properties may be declared as identical within a given precision. For example, two persons have the same weight of 80 kilos using a scale with kilos as unit. A more precise scale may

ISSN 0085-7130©Telenor ASA 2007

give different weights like 80.3 kilos for one and 80.1 for the other.

Taking into account the infinity problem, Principle 2 is not usable because it is not possible to determine the identicalness of all the properties of two objects in order to determine their identicalness.

c. The Impact of Quantum Mechanics

Quantum mechanics say that in each state of a system of n particles of the same kind, it is not possible to distinguish one particle from another, i.e. it is not possible to say which particle is which. Although controversial as interpretation, it is a useful heuristic to think of all particles as equal at all the positions they might be in but not determinately at any of them. The particles would seem to be indiscernible although not identical. This is contradictory to Principle 2.

2.3.3 Problems with both Principles

a. The Ship of Theseus Paradox

Consider a wooden ship that is restored by replacing all its planks and beams by new ones [2]. By Principle 1, The indiscernibility of identicals, the "new" ship cannot be the same ship because some of the planks and beams are not the same. Some philosophers like Rea do agree with this conclusion but some others like Plutarch say that the ship remains the same.

The situation gets more complicated when the old parts of the ship are reassembled to constitute another ship exactly like the first. These two ships are clearly two distinct ships. According to Principle 2 the restored ship is the original ship. However, both the restored ship and the reassembled one deserve to be the original.

A similar problem occurs in a brain transplantation from one body to another. It is not possible to determinate which person will wake up after the operation, the brain donor or the receiver.

b. The Paradox of Constitution

Suppose that on day 1 Jones purchases a piece of clay c and fashions it into a statue s_1 [2]. By Principle 2, $c = s_1$ because they have all the same properties.

On day 2, Jones destroys s_1 but not c by squeezing s_1 into a ball and fashioning a new statue out of c. By Principle 2, $c = s_2$ because they have all the same properties.

On day 3, Jones removes a part of c, discards it and replaces it using a new piece of clay, thereby destroying c and replacing it by a new piece of clay c'. By

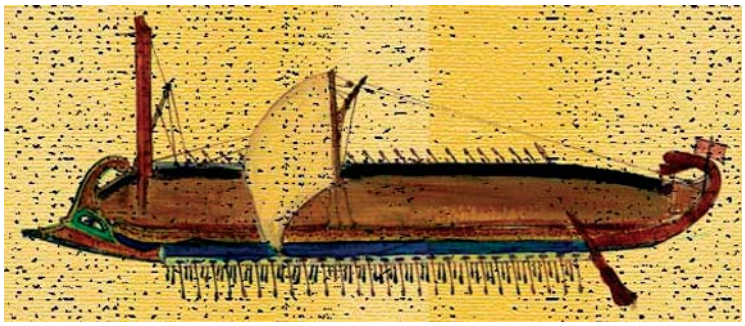


Figure 1 Ship of Theseus

Principle 1, it is possible to say that the statue is still s_2 since it keeps all the properties of the old statue. By Principle 2, $c' = s_2$ because they have all the same properties.

Since identity is a transitive relation,

$$s_1 = c \text{ and } c = s_2 \text{ implies } s_1 = s_2$$

This means that s_1 exists also on day 2, which is not true.

By similar argument, on day 3,

$$c = c'$$

This means that c exists also on day 3, which is not true.

Some philosophers like Gibbard explain that the statue s and the piece of clay c coincide in their entire existence but they are not the same since s may be admired for its aesthetic traits, even long after it ceases to exist, but c cannot be. Other philosophers adopt the doctrine of temporal parts which divide the scenario into separate stages. Since the stages are not identical then transitivity cannot be applied.

3 Personal Identity

As explored in the previous section identity of things is a controversial subject that has raised many debates, but personal identity, i.e. identity of human beings is much more complex. This is not surprising since the human being is the most advanced and complex creature on Earth who can think about himself.

Personal identity is about a person ascertaining the sameness of another person. This issue consists of the following questions:

- *What is it to be a person?* What is necessary and what is sufficient for something to be considered as a person?

ISSN 0085-7130

- *What does it take for one person to persist from one time to another?* What are the necessary and sufficient conditions to say that the same person exists in different times? What sort of experiences a person could survive? What sort of things such as death that end the existence of a person?

- *How do we find out who is who?* What evidence should be used to determine that the person today is the one who was here yesterday?

In addition, since a human being has the ability to think and judge himself, personal identity is also about a person recognizing himself. This issue in its turn contains the following questions:

- *Who am I?* What makes me unique as an individual and different from others? Is it the way I see or define myself?

- *What am I?* What sort of things in terms of metaphysics am I? What am I made of?

- *How do I want to be perceived by other people?* What kind of person do I want to be regarded as?

- *How could I have been?* How different could I have been from the way I actually am? Could I have different parents?

- *Whom can I identify myself with?* Whom do I wish to be? What kind of person do I want to be? Who are my heroes?

3.1 The Persistence Question

As for identity of things the most controversial issue related to personal identity is the persistence of a person over time. It is about numerical identity through time, i.e. to be able to say that a past and a future being are exactly the same and one thing rather than two. If human beings remain unchanged with time then according to Leibniz's Law the qualitative identity can be used to determine numerical identity. Unfortunately, a person changes throughout life both in size, appearance and in many other aspects from an embryo to end up in a persistent vegetative state. Nor is it possible to determine how many differences can be tolerated before it is another person. Qualitative identity can then not be used to determine the sameness of a person. This means that it is not possible to determine the sameness of a person by comparing the properties, and we are faced with an unresolved problem:

Under what possible circumstances is a person existing at one time identical with a person existing at another time?

Three main answers are proposed to the persistence question:

- The Psychological Approach
- The Somatic Approach
- The Simple View

3.1.1 The Psychological Approach

In this approach [8] a certain *psychological continuity* is necessary for a person to persist. A future being inherits the mental features like beliefs, memories, preferences, capability for rational thought of a present being. The present being is in turn the past being whose mental features are inherited by the present being.

Although reasonable for Western Philosophers, the Psychological Approach meets a serious paradox. It suggests that a person would follow his brain if it is transplanted into a different head because the brain contains memories and other mental features. However, this suggestion cannot be verified with full confidence. Now, the brain consists of two hemispheres. If each hemisphere is transplanted to a different empty head then the two resulting persons will be both psychologically continuous to the brain's owner. This implies that both persons are identical with the brain's owner. This again implies that they are the same person. This is absurd because two different persons cannot be the same one.

3.1.2 The Somatic Approach

Another opinion means that personal identity is comprised of some *brute physical relation* [8]. A being is the past or future being that has the same body or the same biological organism. A person survives and per-

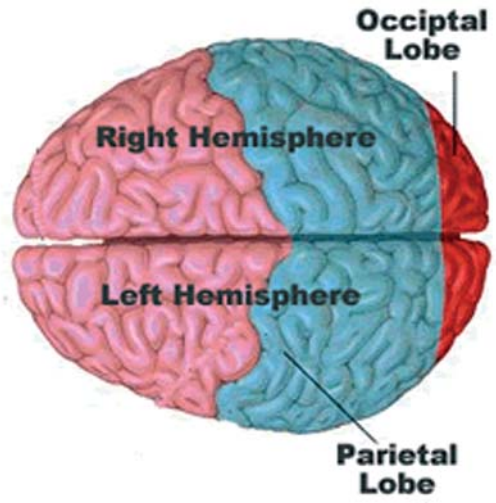


Figure 2 The brain – Cerebrum (Source: Lasker Medical Research Network)

ISSN 0085-7130©Telenor ASA 2007

ishes with his body. A person's identity through time consists in the identity of his body.

The Somatic Approach is unpopular. In the brain transplantation example, the brain's owner will stay behind with his body while the brain transplanted person thinks he is the brain's owner but is not. The Somatic Approach has the virtue of being aligned with the common people's belief. A person is a human animal and has the persistence condition of animals. When someone lapses into a persistent vegetative state, his relatives rarely conclude that their loved one no longer exists, even when they believe that there is no mental continuity of any sort between the human vegetable and the person.

Most people may think that the truth lies somewhere between the two mentioned approaches because a human being needs *both mental and physical continuity to survive*. Unfortunately, it is sometimes difficult to unify the two approaches. In the brain transplantation case, the Psychological Approach will conclude that the body receiving the brain will be the person while the Somatic Approach would say that the empty-headed vegetable is the right person.

3.1.3 The Simple View

Both the Psychological and Somatic Approaches agree that there is something that it takes for a person to persist, i.e. personal identity follows something else than itself. A third opinion denies that mental and physical continuity are evidence of identity but none of them are both necessary and sufficient. The only correct and complete answer is that a person existing at one time is identical with a human being existing at another if and only if they are identical. There are no informative, non-trivial persistence conditions for people. This view is called Simple View [9] and is poorly understood.

3.2 The Evidence Question

The evidence question is quite often confounded with the persistence question but is not the same. The persistence question focuses on finding the requirements necessary to conclude that one person at one time is identical to another one at a different time. The evidence question on the other hand is about proving that the person here now is the one for some time before given the persistence requirements. One source of evidence is memory: one remembers about the event. Another source is physical continuity: the person from yesterday does look just like the one today. Which of these two sources are more fundamental remains for discussion.

4 Citizen Identity

In most countries, in order to ensure the rights and obligations of citizens governments need to establish a citizen identity used in the identification and verification of their citizens. The citizen identity is physically oriented rather than psychologically and focuses more on the *physical continuity* of a person. To avoid confusion, a clearer definition of identity is required [11].

Let us first review the definition of identity and its related notions.

As stated by Leibniz's Law, the identity of a person is defined by all the properties that he has. Since both the number of properties and their values can be infinite it is not appropriate to use all the properties in the identification and verification of the identity of a person. Instead, it is more adequate to use only a restricted set of properties that are *characteristic* to a person. These characteristic properties are called *Attributes* [10].

An **attribute** is a characteristic associated with an entity, such as an individual.

An attribute can be **intrinsic**, i.e. that belongs by nature, or **extrinsic**, i.e. acquired from the outside. Examples of intrinsic attributes include race, eye colour, biometrics (e.g. fingerprints). Example of extrinsic attributes include family name, first name, address.

An attribute can be **persistent** or **temporary**. Examples of persistent attributes include height, eye colour and date of birth. Examples of temporary attributes include address, employer and organizational role. A Social Security Number is an example of a long-lived attribute. Some biological attributes are persistent (e.g. fingerprints); some change over time or can be changed (e.g. hair colour).

To denote, address and identify a person *identifiers* are used.

An **identifier** identifies a distinct person, place or thing within the context of a specific namespace. An identifier is *an attribute that is most representative for an entity within a context*. An identifier is also referred to as *name, label* and *designator*.

For example, an automobile, a bank account and a person each have identifiers. The automobile has a license plate and the bank account has a number. The person may be associated with either the auto or the account through additional information, such as a certificate or ownership, or a social security number.

ISSN 0085-7130

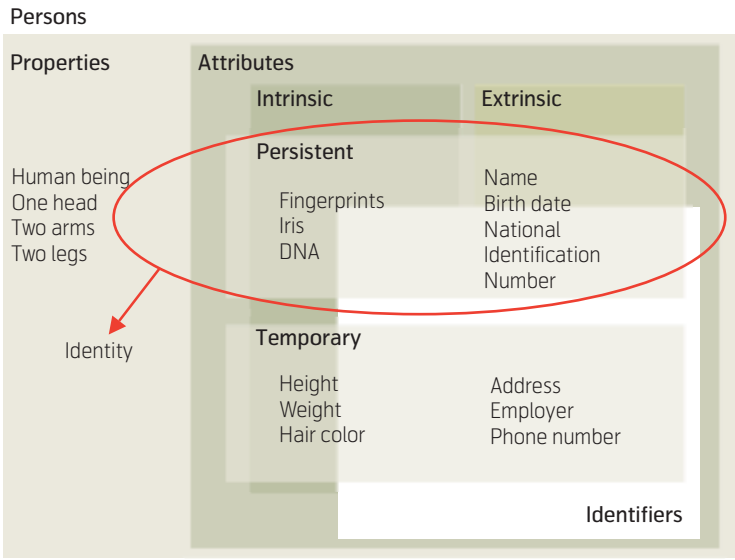


Figure 3 The definition of identity

One identity can have multiple identifiers: A car has a permanent serial number and a temporary license place. Each identifier is meaningful only in a specific context, or namespace, and can reasonably be thought of as having a <thing identified, identifier> pair.

Personal identifiers are unique persistent identifiers associated with an individual human being that are difficult or impossible to change, such as *biometric characteristics* and genetic code.

Biometrics are automated methods of recognizing a person based on a physiological or behavioral characteristic. Among the features measured are: face, fingerprints, hand geometry, handwriting, iris, retinal, vein, and voice [13].

An identity is a *set of permanent or long-lived permanent attributes and personal identifiers associated with an entity such as an individual. With an identity, it must be possible to recognize an invidual.*

In order to avoid misunderstanding it is also crucial to differentiate identification, authentication and authorisation.

Identification is the association of a personal identifier with an individual presenting certain attributes [11]; for example, accepting the association between a physical person and claimed name, or determining the association with a medical record and a patient using physical attributes.

Authentication is proving an association between an identifier or attribute, and the relevant entity. For example, an automobile is identified by its license place, and that is authenticated as legitimate by the database of cars that are being sought for enforcement purposes.

Identity Authentication is proving an association between an entity and an identity. For example, the association of a person with a credit or educational record.

Attribute Authentication is proving an association between an entity and an attribute. Confirming someone's age is an example. This is usually a two-step process, where the association between an entity and an identifier is established, and then a link between identifier and attribute is established.

Authorization is a decision to allow a particular action based on an identifier or attribute. Examples include the ability of a person to make claims on lines of credit, the right of an emergency vehicle to pass through a red light or a certification of a radiation-hardened device to be attached to a satellite.

A citizen identity quite often includes the following attributes:

- Full name (i.e. First, Middle and Last Name)
- Birth Date
- Gender
- Place of Birth
- Parent Names
- National identification number (e.g. social security number, national number, personal identification number, personal number)
- Religion
- Ethnicity
- Citizen status

However, an identity with only the mentioned extrinsic attributes is very difficult to verify since the extrinsic attributes can be copied, falsified and mistaken easily. Intrinsic attributes like face, fingerprints, iris, etc. must be used to ensure the authenticity of a person. In addition, these attributes must be stored in databases that are available and accessible for authentication when requested.

To facilitate identity authentication additional *credentials* or identification document or identity card are introduced.

A **credential** is an attestation of qualification, competence, or authority issued to an individual by a third party with a relevant de jure or de facto authority or assumed competence to do so [12].

An **identification document** or **identity card** is a credential designed to verify aspects of a person's

ISSN 0085-7130©Telenor ASA 2007

identity. Information present on the document might include the bearer's full name, a portrait photo, age, birth date, address, an identification number, profession or rank, religion, ethnic or racial classification, restrictions, and citizenship status. New technologies could allow identity cards to contain biometrics such as photographs, face, hand or iris measurements, or fingerprints.

Although the utility of identity cards may seem obvious, the extensive cost and the potential abuses create a lot of debates. One of the concerns is privacy because electronic ID cards can be used to track anyone's movements and private life. It is also worth noting that in the real physical life the usage of credentials is required only in special occasions like traveling abroad, withdrawing money from a bank, entering some governmental offices, etc. Trust through direct face-to-face communication and authentication is the basis of all decisions and actions.

5 Digital Identity

In the current information age, the real physical world is increasingly extended by a cyber world created by computers. In fact, the access to information and services via the Internet and other computer networks is playing a more and more important role in the individual's life from daily activities, commercial transactions, entertainments to education and governmental services. The amount of time that each individual is spending in the cyber world is increasing. In the cyber world, like in real physical life, for regular actions like web site visits no identification and authentication is needed. But, for more serious errands like paying bills at the net bank, entering governmental web sites, etc. appropriate authentication of the user is required before access can be granted. In such cases a *digital identity* is required.

A **digital identity** is a representation of a human individual's identity in a computer network system like Internet, Corporate Intranet, Home networks, etc. A person does not really exist in the cyber world. Moreover, the communications and interactions in the cyber world are not face-to-face. People do not see who they are dealing with. Consequently, the physical intrinsic attributes like face, hair colour, fingerprint, etc. cannot be used to identify the user. Extrinsic attributes and identifiers like name, pseudonym, etc. are required. Unfortunately, they can be easily copied and duplicated. A *credential* must hence be introduced as an additional attribute to prevent theft of identity. This credential is a secret that only each individual and the corresponding authentication authority know. In its simplest and weakest form, this credential is a password. For higher levels of security, encryption keys and algorithms are used to realize this credential.

A digital identity, as seen by a user consists of a *username* or *login name* and a *password*. However, what other attributes are incorporated within a digital identity varies very much depending on the identity provider.

Some digital identities are issued and certified by established institutions like governmental offices, banks, corporate, etc. These contain all personal attributes and fully identify the individual user.

Other digital identities for social communities like music groups, fan clubs, hobby groups, etc. are self-defined by the user, i.e. the login name is chosen by the user, and may be fully anonymous since no other attribute than the email address is given. In such a case the system is only interested in recognizing the user that has joined the community but not to know who the user really is. These identities ensure the privacy of the user. Another important need that these self-defined identities satisfy while the certified identities do not, is the need for psychological identification. Indeed, for social communities, users may want to choose as login name a hero's name that they identify with. For example, "Batman", "Zorro", "Bart Simpson", etc. are quite popular login names.

The most serious problem with the self-defined identities is the lack of traceability. Criminals or terrorists can hide behind an apparently nice user. The need for an identity scheme for the cyber world that ensures both privacy and traceability is obvious.

Another not less serious problem is the constantly increasing number of login names and passwords that each user is getting. To solve this problem, single sign-on solutions enable the user to sign in once and access all the web sites while simplified sign-on solutions assist the authentication process by managing the user credentials and presenting them to the authentication authorities at request.

6 Conclusions

While the notion of personal identity is not yet fully understood and human beings are still struggling with the identity definition problem, the emergence of disorganized and fragmented digital identities complicates the situation even more. Suddenly, a person can now have a bunch of different identities: some identifying him, some denoting people that he wants to be, some pointing to people that he pretends to be. The paradox here lies in the fact that if a personal identity does not manage to identify a person then it could not

ISSN 0085-7130

be an identity. It is hence crucial that no matter how many identities a user has and who these identities reflect, they must in the end converge to the same and only person. The need for a method to define the links between the identities both in the real world and the cyber world and to navigate between them, i.e. an identity management system for both worlds, is becoming quite urgent. To be successful, such a system needs to have the consensus of not only governments, enterprises, banks but also the human user.

References

1 Blunden, A. 1997. *The Meaning of Hegel's logic*. November 22, 2007 [online] – URL: http://www.marxists.org/reference/archive/hegel/help/mean.htm

2 *The Identity of Indiscernibles – Stanford Encyclopedia of Philosophy*. October 11, 2007 [online] – URL: http://plato.stanford.edu/entries/identity-indiscernible/

3 Loemker, L. 1969, 1956. *Leibniz: Philosophical Papers and Letters*. Reidel.

4 *Identity of Indiscernibles*. October 11, 2007 [online] – URL: http://en.wikipedia.org/wiki/Identity_of_indiscernibles

5 Black, M. 1952. The identity of indiscernibles. *Mind*, 51, 53-64.

6 *Identity – Stanford Encyclopedia of Philosophy*. October 11, 2007 [online] – URL: http://plato.stanford.edu/entries/identity/

7 *Relative Identity – Stanford Encyclopedia of Philosophy*. October 11, 2007 [online] – URL: http://plato.stanford.edu/entries/identity/

8 *Personal Identity – Stanford Encyclopedia of Philosophy*. October 11, 2007 [online] – URL: http://plato.stanford.edu/entries/identity/

9 Chisholm, R M. *Person and Object*. Routledge, 2004. (ISBN 9780415295932)

10 *The National Electronic Commerce Coordinating Council: Identity Management – A White Paper*. October 11, 2007 [online] – URL: http://www.ec3.org/

11 Camp, J. *Identity in Digital Government*. A report of the 2003 Civic Scenario Workshop at the Kennedy School of Government, Harvard University Cambridge, MA, 2003.

12 *Credential*. October 11, 2007 [online] – URL: http://en.wikipedia.org/wiki/Credential

13 *The biometrics consortium – An introduction to biometrics*. October 11, 2007 [online] – URL: http://www.biometrics.org/html/introduction.html

For a presentation of the Do Van Thanh, please turn to page 2.

Dr. Ivar Jørstad received his MSc in Informatics with specialisation in communication systems from the University of Oslo in December 2002. He obtained his PhD in Telematics from the Norwegian University of Science and Technology in 2006. He is currently the CEO of a Norwegian start-up company, Ubisafe AS, which specializes in security solutions for Internet-based services by utilizing security mechanisms of the mobile networks. His fields of interest include service architectures and platforms for mobile services and applications, mobile distributed computing, Service-Oriented Architectures (SOA), XML Web Services, mobile terminal platforms and middleware, personalisation of mobile services and security of distributed services.

email: ivar@ongx.org

ISSN 0085-7130©Telenor ASA 2007

Identity Management Demystified

DO VAN THUAN

Do Van Thuan is Lead Scientist in Linus AS

With the increasing number of identity theft leading to serious damages, it is not surprising that identity management is becoming an area of focus. Unfortunately, it is not clear what identity management is and what domains, processes and technologies it encompasses. It is hence not simple for enterprises to elaborate and deploy a sound and cost efficient identity management. This paper provides a comprehensive and non-exhausting presentation of identity management that can be useful in the planning and establishment of an identity management system. A clear definition of identity management is given and the IdM components are explained thoroughly. To elucidate the concept of IdM the paper presents the enterprise identity management, the partnership identity management, the customer relations identity management and identity management as a commercial service.

1 Introduction

Identity management is a hot area that is experiencing considerable growth. It is predicted that the world-wide sales of identity and access management systems will rise to more than USD 950 million by 2009 [1]. This is not surprising because organisations, supply chains and customers have been tightly connected together in the digital networked economy. Identity management has become the fundament of security. People need to prove they are who they claim to be and their claimed identity must be consistent with the previous knowledge the organisations have on them. The need for a sound identity management is indisputable but the problem lies in the realisation of an efficient and affordable identity management infrastructure [3] [5] [6].

In fact, an identity management infrastructure cannot be bought from a vendor but has to be built from various components in accordance with the specific goals and requirements of the company [4]. This requires technological, economic and strategic knowledge and expertise that are usually missing. The goal of this paper is to give a comprehensive and non-exhausting presentation of identity management that can be useful in the planning and establishment of an identity management system.

2 Definition of Identity Management

There are currently several definitions of identity management which are partially overlapping and sometimes conflicting with each other. Depending on the context, identity management can mean different things to different people. In this paper, Identity management is defined as *a discipline that consists of processes, policies and technologies to manage the complete lifecycle of user identities across the system and to control the user access to the system resources by associating user rights and restrictions. These resources include information, services, process capability, buildings and physical asset* [2].

Identity management (IdM) is also referred to as *Identity and Access Management* (IAM) [4].

3 Business Objectives

The objective of IdM is to achieve the following business benefits:

- **Advanced Enterprise-wide compliance**: Enforcement of consistent business rules and practices; tightening of control over user-to-application interaction; Reporting and auditing of enterprise-wide identity lifecycle events.
- **Lower operational costs**: Automation of lifecycle management for potentially millions of users; Single Sign-On implementation improves user productivity and reduces password reset costs.
- **Enhanced security**: More adequate protection of assets; Elimination of latency in implementation of access privileges due to identity and policy changes; Support for the most granular protection of enterprise resources; Support for multiple authentication schemes.
- **Improved productivity**: Personalised user access to information, services and tools; extended and enhanced access from outside enterprises.

ISSN 0085-7130

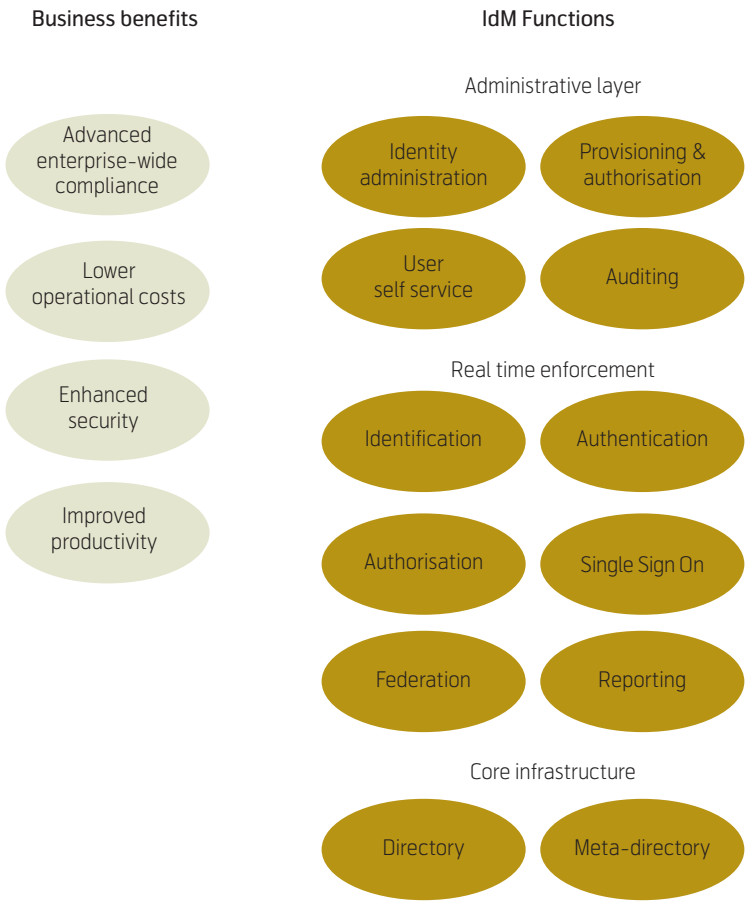


Figure 1 Relations between the IdM functions and business benefits

4 The Identity Management Functions

In order to realise the business objectives mentioned above, an identity management system must be equipped with a myriad of functions such as:

Administrative layer
- Identity administration
- Provisioning & authorisation
- User self-service
- Auditing

Real time enforcement
- Identification
- Authentication
- Access Control
- Single sign-on or Simplified sign-on
- Federation
- Reporting

Core infrastructure
- Directory
- Meta-directory

4.1 Identity Administration

Identity administration includes the complete management of the user digital identity lifecycle in the system. More specifically, it consists of the following:

- Creation and registration of the user identity
- Maintenance and evolution
- Ultimately termination

It includes the issuance, maintenance and revocation of credentials associated to the identity:

A digital identity will be created for every new user. She will receive a private identifier, namely a user name or login name for signing in to the system. A public identifier such as email address can also be allocated for the communication with the new user. The digital identity contains also attributes of her personal identity like name, birthday, address, citizen number or personal number (social security number), photograph, etc. Other attributes related to the system can also be defines and allocated to the users such as role, position, title, skills, etc.

The strictly necessary condition for the creation of the user identity is the identity proofing. This could be done by a physical presentation of the user at a registration desk where she presents an identity card as proof of identity. Copy of the ID card can be made and saved as an attribute of the new user identity. The identity proofing can also be done electronically. The user logs on to the system for the first time by using a one-time password that is obtained via post, email, SMS, etc.

When the identity has been proven, the user will be given credentials for authentication at future accesses to the system. The simplest one could be a password. Stronger credentials include one-time password generator, smartcards, etc.

4.2 Provisioning and Authorisation

Provisioning includes the provision of appropriate resources to each user. Appropriate use of resources (which are typically business assets) is assured through the management and enforcement of permissions, often called *access rights* associated with those resources. Resource provisioning is typically the vehicle for management of such access rights. Permission to access is certainly a permission to be managed, but it is far from the only relevant permission. Other permissions include permission to compare, write, modify, create, destroy, execute, copy, print, forward, delegate, purchase, authorize, approve, sell, sublease, assign, transfer, hire, fire, promote, and so forth.

ISSN 0085-7130©Telenor ASA 2007

The management of access rights associated with access control is typically conducted through the management of access control lists; however, access control lists fall short of being able to express non-access control permissions. Two examples of other mechanisms include attribute certificates and digital rights management.

Permissions are allocated to individuals by an authorization authority in order to authorize the individual to use controlled or protected resources according to the system policies.

Permissions are not identity attributes, though they may in some cases be derived from identity attributes. For example, an organisation may have a policy that states that all employees shall be granted permission to enter the company's building. This policy is essentially a rule which grants "building entry" permission from the "employee" attribute of an individual's identity. In cases in which permissions are derived from identity attributes, it is important to ensure that changes in identity attributes (employee status, age, etc.) are communicated to the person or system in charge of managing permissions, so that permissions can be revised if necessary.

4.3 User Self Service

User self service includes the tools and functionalities that enable the user to participate in the administration of her identity e.g. resetting passwords, obtaining temporary access rights to a particular resource, etc. This will contribute to reducing administrative costs, particularly in the help desk service.

The user can reset forgotten password after authenticating themselves using other previously defined passwords, one-time password sent via other channels like email, sms, etc., or biometrics.

The user can ask for temporary permission to access a certain resource and the manager just needs to approve the request sent by the system. This can reduce the manager's workload and also the time consumption for the accomplishment of the task.

4.4 Auditing

Auditing is an integral part of standard security operations. It includes capturing, archiving, mining and post mortem analysis of audit trails relative to transactions performed within the security infrastructure. In the context of identity management, auditing information is an asset captured by both identity and access management systems which are involved in the lifecycle management of user identities.

Auditing may also include forensic analyses that are aimed at determining whether unauthorized activities have taken place, based on the correlation of various events and their specific attributes. This process involves collecting and analyzing audit logs from systems tracking user activity.

4.5 Identification

Identification is the process of recognizing the user at login. Usually, it is based on a private identifier that the user got issued such as user name, login name, ID code, etc. Identification may also include the identification of a group of users in the same category or having the same role and responsibilities.

4.6 Authentication

Authentication is the process of verifying that a user is who she claims to be. There are many authentication methods of different strengths and based on different mechanisms, also called factors, such as:

- Something the user knows such as a password;
- Something the user has such as a smartcard or SIM card;
- Something the user is born with such as biometrics.

Authentication solutions based on one factor are usually weak and exposed to falsification. Authentication can be considerably strengthened by increasing the number of factors to two. Example of two-factor authentication is the usage of smartcard combined with a password. Authentication can be further strengthened through the use of multiple channels in the delivery of credential factors. For example, while the communication with the smartcard is carried out on IP links, the submission of the password is done via SMS to the system.

Stronger authentication schemes will necessarily introduce higher cost and sometimes higher complexity for both the users and the administrator. The selection of authentication methods must be carried out in a thorough way to find the most appropriate one.

4.7 Authorisation

Authorisation is the real time process that determines what resources a user is allowed to access. For example, a user may be allowed to access her documents, but not those of another user. The information that specifies what individuals are authorized to access may be stored in multiple databases maintained by different administrative units.

While the process is conceptually simple, it is complex to execute. Defining authorisation on a case-by-case basis is extraordinarily time-consuming. Other

ISSN 0085-7130

schemas, based upon an individual's role, organisational structure or policy, are fraught with exceptions. The need to translate complex policies into automated combinations of more basic attributes is an area that is rapidly evolving. There is also the need for immediate change of authorisation in case of attacks and the need for changing back once the situation is again normalized.

4.8 Single Sign-on

Single sign-on is the mechanism that enables the user to sign in just once and have access to all the needed resources. Its benefits include increased productivity and ease of use. Single sign-on is of particular importance because it removes the user's burden of remembering many passwords and the security breeches created when the user writes down her passwords.

Services may require different levels of authentication, and re-authentication may be required before access can be granted to the user. Single sign-on is therefore not always possible and the Liberty Alliance uses the term *"Simplified sign-on"* rather than Single sign-on. With Microsoft CardSpace, *"Reduced sign in"* is used since the user is offered the possibility to select in the appropriate ID card containing the required credentials for authentication.

It is worth noting that with Single sign-on, the user may still have several identities for different services and it is always possible to reverse the process and demand appropriate log-in for each service separately.

A closely related and even more important mechanism is *Single sign-off* that closes the access to all services when the user logs out from one of them. Single sign-off is very important because the user tends to forget to log out and then leaves the access to malicious intruders.

4.9 Federation

Federation is the mechanism that enables the portability of identity attributes across autonomous security domains [7]. The ultimate goal of identity federation is to enable users of one domain to securely access data or systems of another domain seamlessly, and without the need for completely redundant user administration. Identity federation comes in many flavours, including 'user-controlled' or 'user-centric' scenarios, as well as enterprise controlled or B2B scenarios.

In the user-centric solutions, the users will themselves execute the federation of their identities at different security domains. Opaque links between identities will be established to support navigation between identities while anonymity and privacy are preserved.

In the enterprise controlled or B2B scenarios, organisations share identity attributes to enable automated collaborative processes [2]. For example, employees of two collaborating enterprises are automatically granted access to some agreed resources.

4.10 Reporting

Reporting includes the tools that are necessary to generate reports based on information collected by the auditing. It should be able to have report at all levels in the identity management infrastructure, both in the policy administration and in the real time enforcement areas. The potential volume of information is often quite large and it is essential that the reporting tools provide digests of the information that is relevant to an inquiry. They must also provide a range of business and technical views, including the business interpretation of events.

4.11 Directory

Directory is the central function of the identity management system that stores all the identities and information related to all entities in the system from user, user groups, services, service groups, resource, resource groups, etc. It includes not only systems that use the ITU X.500 protocols or the IETF Lightweight Directory Access Protocol (LDAP), but also relational databases, flat files, and data stores of other kinds [5].

Most large organisations have a large number of different systems used as directories and the identity information is distributed across them, often with duplications. The biggest challenge is to keep data consistency, and meta-directory is needed.

4.12 Meta-directory

Meta-directory is the function that provides an organisation-view of information held in several heterogeneous directories and other storages like databases. It is very important for organisations that have a diversity of platforms or multiple directories due to geographical dispersion or for political reasons.

5 Enterprise Identity Management

Enterprises have quite often different computer systems for departments like R&D, products, services, marketing, etc. For each system there are quite often separate administrative entity, separate user identities and different authentication schemes as shown in Figure 2a. Each user will have to remember several usernames and passwords, and for each time she forgets a password the IT support will have to reset it and distribute a new one. This is both an inconvenience for the user and an administration cost for the IT support. A better solution is to have a centralised identity

ISSN 0085-7130©Telenor ASA 2007

management system where each user just receives one pair of username and password. Authentication and access control are done once and the user is permitted to access all the systems according to her access rights and the company's policies.

Single Sign-on is offered between computer systems which are federated into a *Circle-of-Trust*. The federation should be flexible to allow each system to be de-federated when needed. The employee's identity and access rights will be administered in a uniform and consistent way throughout the enterprise's whole computer system.

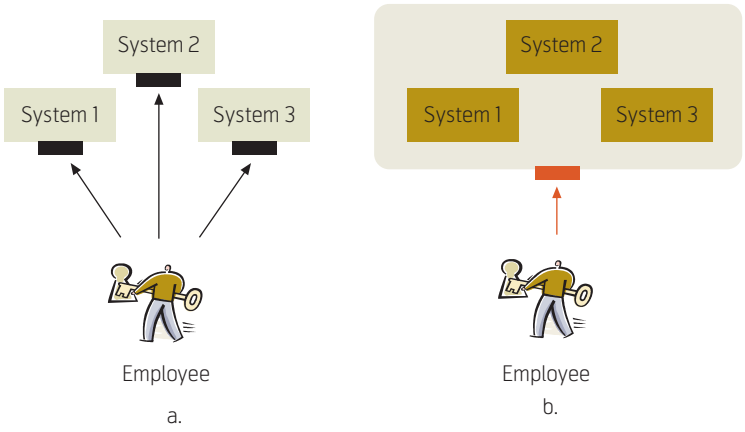


Figure 2 Distributed versus Centralised Identity Management

6 Partnership Identity Management

Nowadays, in order to survive, companies must be agile and flexible to cope with changes in the market. They must be able to establish alliances and collaborations with other companies as fast as they terminate them. In the current digital age, collaboration requires sharing of information between the partners. However, quite often only employees of the partnering companies participating directly in the collaboration are granted permission to access a defined set of resources.

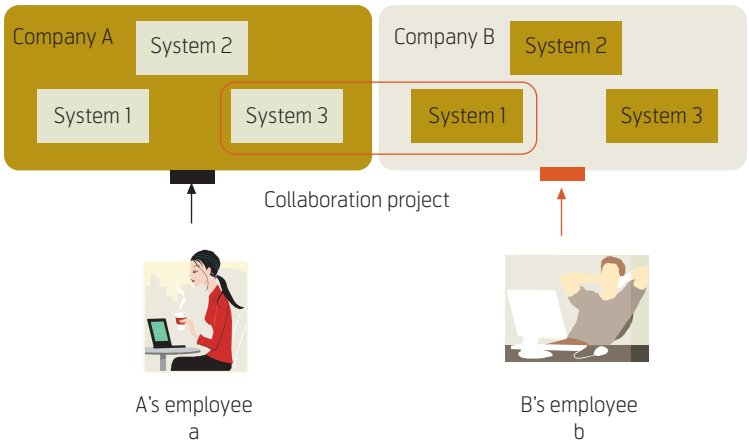


Figure 3 Partnership Identity Management

As shown in Figure 3, A's employee participating in the collaboration will be allowed to access B's System 1, and B's employee participation in the collaboration has the permission to access A's System 3. It should not be necessary to distribute new usernames and passwords to the employees participating in the collaboration. Instead, access to the computer systems should be enabled by the federation between A's System 3 and B's System 1. Single sign-on is provided across company systems. For example, A's employee participating in the collaboration can access B's System 1 without having to log in again. The federated systems must be de-federated rapidly when the collaboration is terminated. It should also be able to revoke the access rights immediately when one employee leaves a partner company.

7 Customer Relations Identity Management

To reduce administrative costs, the customer is encouraged to carry out by herself as much of the administrative tasks as possible. To avoid duplicating data, the customer should be allowed to access certain systems and data. Careful evaluation should be carried out to decide which system should be made available to the customer and additional security measures may require protecting the company's assets. Each customer will be issued an account with a login name and password allowing access to certain user

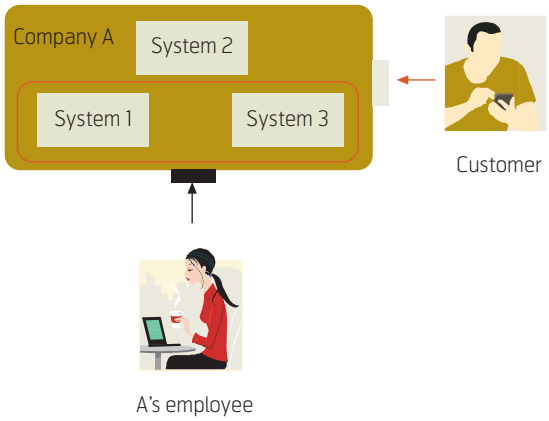


Figure 4 Customer relations Identity Management

data as shown in Figure 4. There should also be a function allowing the users to reset forgotten passwords to reduce administrative costs. Single sign-on is compulsory to provide a simple sign-on for the users.

ISSN 0085-7130

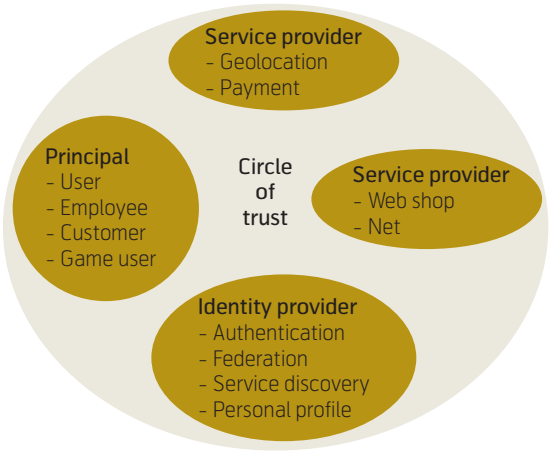


Figure 5 A Circle of Trust

8 Identity Management as Commercial Service

As explained earlier identity management is quite a complex area that requires both knowledge and resources which a regular organisation does not have. Therefore it might be more appropriate for organisations to outsource the identity management to an identity provider.

An *identity provider* is an organisation that directly manages end users. An identity provider is the authoritative source for issuing and validating user identities and network credentials for a set of users; an identity provider "owns the user relationship". For example, many companies act as identity providers for employees, customers, and contractors. Identity providers "vouch" for the user identity and their entitlements in a federated interaction with service providers. So, the "identity provider" role can be thought of as an authentication authority.

A *service provider* provides "services" for end users. They typically do not have a vested business interest in managing the user. Service providers act as a "relying party" to validate credentials issued by a trusted identity partner, on the basis of which they provide services to that trusted identity.

As shown in Figure 5 a *Circle of Trust* consists of one identity provider, several service providers and a large number of users [8][9][10]. The roles of identity provider and service provider are not mutually exclusive. At the moment, many large organisations assume the role of both identity provider and service provider with all the costs and hassles of maintaining an IdM infrastructure. A solution is to have well-known established organisations such as telcos, to offer identity services to enterprises and to hold identities on behalf of consumers. That is, there will be an identity provider and many service providers sharing an IdM infrastructure and providing single sign-on to their customers. The proposition of providing IdM as a commercial service can be explained from several points of view.

Consumers have become far more concerned about invasions of their privacy, and they wish to be as anonymous online as in real life. The idea is that a user will get identities from a company she knows and trusts – for instance, her telephone company or bank. That company becomes her identity provider. The user can then surf the Internet in complete privacy. When a website (a service provider) asking for age verification – for instance, a children chat room or an adult forum, her identity provider can vouch for her without revealing her real identity; at the same time the identity provider also guarantees to the service provider that this is in fact a real person with the correct age. When she wishes to buy something, her identity provider can create a new identity for her – complete with a fictitious name and email address, a coded postal address, and a one-off credit-card number.

This fresh identity is passed, via the online merchant, back to the identity provider, which matches the details with the user's real identity and forwards the transaction details to her on-file credit-card company, which will check and approve the transaction. Meanwhile, the post office is sent a decoded address label, but still a coded name, and ships the goods. No entity knows what is actually going on because each knows only what it needs. Two added bonuses are that, because the fictitious identity is used only once, it is impossible for online marketers to develop a profile of the user – or for criminals to profit from its theft.

This very much simplified scenario is surely an exciting and compelling experience for consumers once the necessary cooperation between different public organisations and enterprises is in place: contracts with the post office or transport firms for decoding addresses; agreements with financial-services firms for payment; etc. There is also the more controversial exchange of user information, such as payment details and buying preferences; between identity providers and service providers.

From the functional point of view, this is actually a separation between authentication and authorisation. An identity provider knows who a user is and can verify this without revealing her true identity. A Service provider does not need to know the user identity but does know what group/role can access resources and thus maintains control of its resources. Naturally, identity providers and service providers must agree on how to exchange credentials and attributes.

ISSN 0085-7130©Telenor ASA 2007

Features	Advantages	Benefits
Rather than having to enroll users into a company's internal identity systems, federated identity management enables service providers to offload the cost of user administration to the identity provider company.	Having an identity provider, the service provider does not have to take on the burden of user administration costs such as user enrollment, identity management, password management, or provisioning and authorisation.	Help to reduce administration and provisioning costs: Managing identities for users can be a manual, cumbersome, and costly proposition that depletes critical IT resources. Give access to stronger authentication methods.

Table 1 Features, advantages and benefits

When organisations create their own IdM infrastructure for their applications, they are spending time and effort building functionality that they should not be responsible for. Not to mention that the cost of such an infrastructure becomes unbearable for most organisations. With trusted identity providers, organisations can deliver a richer experience for users navigating safely on the Internet while saving costs on infrastructure.

Most application developers would like to just define the roles required and map them to a central identity repository. The same would apply to most organisations; they would like to outsource to an "external identity provider" to work through the federation on their behalf and yet retain control over local enforcements. Table 1 summarises the features, advantages and benefits of having an external identity provider.

Implementation-wise, code-based paradigms of authorisation and control are no longer enough. The move is definitely towards a data-based approach, where notions of authority, roles and relationships, and activities and rules come into play. These are just some of the concerns that the application development process should not need to worry about beyond knowing how to hook into those functions (described in the previous paragraphs) that the central IdM infrastructure is providing.

It is clear that there still is a lot of work to do to get the message out. Public and shared IdM is something we firmly believe in as being the way of the future. Airlines, banks and others still worry that this would drive a wedge between them and their customers, that they would lose control. Incumbent telcos should maximizing their neutral position and seize the opportunity to become identity providers and thus provide their subscribers a safer and friendlier Internet.

9 Conclusion

Identity management is everything about how to manage digital identities. However the Holy Grail is not the technology that allows enterprises to manage identities – and thus risk – in a secure and efficient way, but how to convey the necessity of identity management to both enterprises and consumers alike. As we are entering the digital age, it is also crucial to have identity management solutions that ensure a smooth transition between the real world and the expanding cyber world at the same time as integration of private and professional identities is carried out in a uniform and consistent way without sacrificing privacy.

References

1 Stanton, R. Identity crisis – what crisis? *IT-observer*, 31 May 2006.

2 Titterington, G. *Identity Management : Time for action*. Ovum, 2005.

3 The National Electronic Commerce Coordinating Council (NECCC). *Identity Management – A White Paper*. NECCC, KY, USA, 2002.

4 Windley, P. *Identity Management Architecture and Digital Identity*. October 19, 2007 [online] – URL: http://www.oreillynet.com/pub/a/network/2005/08/19/digitaldentity.html

5 Slone, S & The Open Group. *Identity Management – A White Paper*. March, 2004.

6 Oracle. *Reducing Costs and Improving Productivity with an Identity Management suite – A White Paper*. The Radicati group, May 2006.

7 *Federated Identity*. October 19, 2007 [online] – URL: http://en.wikipedia.org/wiki/Federated_identity

ISSN 0085-7130

8 Liberty Alliance Project. *Liberty ID-FF Architecture Overview v1.2*. September 21, 2007 [online] – URL: http://www.projectliberty.org/liberty/content/download/318/2366/file/draft-liberty-idff-arch-overview-1.2-errata-v1.0.pdf

9 Liberty Alliance Project. *Identity Systems and Liberty Specification v1.1 – Interoperability*. Technical Whitepaper, 14 February, 2003.

10 Liberty Alliance Project. *Introduction to the Liberty Alliance Identity architecture, Revision 1.0*. March, 2003.

Do Van Thuan is currently Lead Scientist in Linus AS – a Norwegian consulting and systems house specialising in products for mobile communications and custom development and system integration for telcos. After finishing his studies at the Institute of Informatics, University of Bergen, in 1984, he joined Norsk Data where he worked first with COBOL and FORTRAN compilers, then with OS command interpreters and last with SQL processors for databases. Since then he has been involved with several large projects developing information systems and process/production control systems. These days he is member of two European projects, ADPO – Personalised Adaptive Portal Framework, and Fidelity – Circle-of-Trust based on Liberty Alliance's Specification. His research interests are distributed systems, component technology and software design methods.

email: t.do@linus.no

ISSN 0085-7130©Telenor ASA 2007

Identity Management Standards, Systems and Standardisation Bodies

NGUYEN DUY HINH, SJUR MILLIDAHL, DO VAN THUAN, IVAR JØRSTAD

Nguyen Duy Hinh is Software Developer in Linus AS

Sjur Millidahl is System Developer in Linus AS

Do Van Thuan is Lead Scientist in Linus AS

Ivar Jørstad is CEO of Ubisafe AS

Identity management provides powerful mechanisms to enhance user privacy. The form and speed of development of the Digital Identity industry will be greatly influenced by the adoption of the appropriate standards. While certain standards are clear in their specification and hence their potential impact on the industry, others are less direct. In this paper we will provide short presentations of different projects on identity management with emphasis on the federated approaches, covering related standard specifications, along with security and privacy considerations.

1 Introduction

In the same way as with the definition of identity and the area of Identity Management (IdM), there is a lot of confusion regarding IdM standards and systems. Indeed, there is a plethora of overlapping and conflicting standards that are specified by partially rival organizations. The adoption and usage of standards in IdM systems and solutions are different creating a lot of incompatibility problems. The goal of this paper is to give an overview of all the standards, systems and standardization bodies that are related to Identity Management.

2 IdM related standards

2.1 Liberty Alliance specification

This specification defines a set of protocols that collectively provide a solution for identity federation management, cross-domain authentication, and session management. This specification also defines provider metadata schemas that may be used to make a priori arrangements between providers.

The Liberty architecture contains three actors: Principal, identity provider (IdP), and service provider (SP), see Figure 1. A Principal is an entity (for example, an end user) that has an identity provided by an identity provider. A service provider provides services to the Principal.

Once the Principal is authenticated to the identity provider, the identity provider can provide an authentication assertion to the Principal, who can present the assertion to the service provider. The Principal is then also authenticated to the service provider if the service provider trusts the assertion. An identity federation is said to exist between an identity provider and

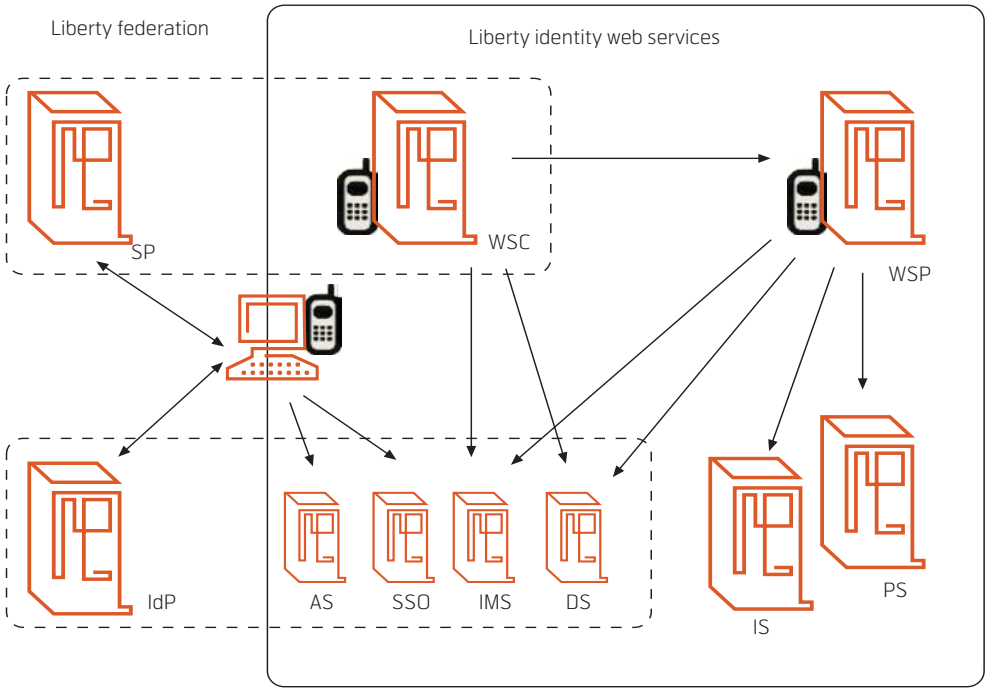


Figure 1 Liberty Alliance actors

ISSN 0085-7130

a service provider when the service provider accepts authentication assertions regarding a particular Principal from the identity provider. This specification defines a protocol where the identity of the Principal can be federated between the identity provider and the service provider. This federated approach does not require the Principal to re-authenticate and can support privacy controls established by the Principal.

This specification relies on the SAML [2] specification. In SAML terminology, an identity provider acts as an Asserting Party and an Authentication Authority, while a service provider acts as a Relying Party.

2.1.1 The Identity Federation Framework (ID-FF)

The Liberty Alliance is developing and delivering the first open architecture and specifications to enable federated identity management. At its core is the Identity Federation Framework (ID-FF), see Figure 2, which facilitates identity federation and management through features such as identity/account linkage, single sign-on, and session management. ID-FF is fundamental to underpinning accountability in business relationships and Web services; providing customization to user experience; protecting privacy; and allowing adherence to regulatory controls.

The Liberty Alliance contributed its federation specifications, ID-FF, to OASIS [9], forming the foundation for SAML 2.0, the converged federation specification that Liberty now recognizes.

2.1.2 The Identity Web Services Framework (ID-WSF)

The Liberty Alliance is also specifying an Identity Web Services Framework (ID-WSF), see Figure 2, that will utilize the ID-FF. This framework introduces a Web Services-based identity service infrastructure that enables users to manage the sharing of their personal information across identity and service providers as well as the use of personalized services. For example, a user may authorize a service provider to access their shipping address while processing a transaction.

2.1.3 The Identity Services Interface Specifications (ID-SIS)

Built on top of the ID-WSF, see Figure 2, is a collection of interoperable identity services, the Identity Services Interface Specifications (ID-SIS). The ID-SIS might include services such as registration, contact book, calendar, geo-location, presence, or alerts. Through Liberty protocols and a standard set of attribute fields and expected values, organizations will have a common language to speak to each other and offer interoperable services. The services defined in the ID-SIS are designed to be built on top of Web services standards, meaning they are accessible via SOAP over HTTP calls, defined by WSDL descriptions, and use agreed-upon schemas.

2.2 SAML – Security Assertion Markup Language

SAML [2] defines an XML-based framework for communicating security and identity (e.g. authentication, entitlements, and attribute) information between computing entities. SAML promotes interoperability between disparate security systems, providing the framework for secure e-business transactions across company boundaries. By abstracting away from the particulars of different security infrastructures (e.g. PKI, Kerberos, LDAP, etc), SAML makes possible the dynamic integration necessary in today's constantly changing business environments. SAML is a product of the OASIS Security Services Technical Committee.

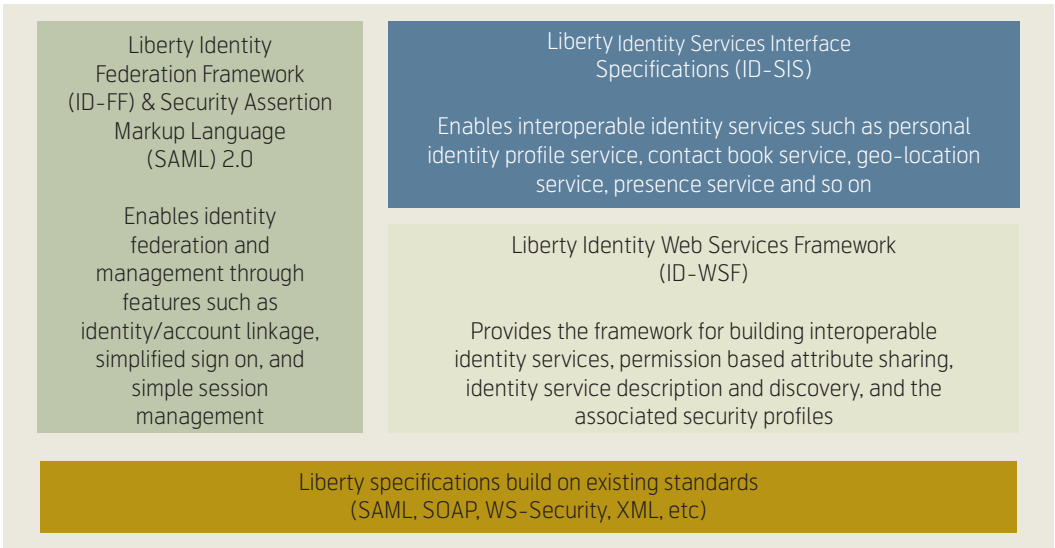


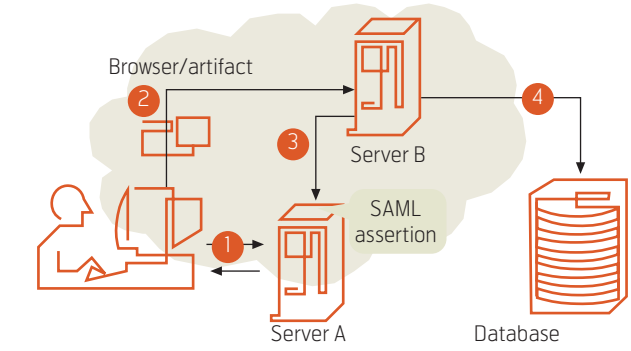
Figure 2 Liberty's Architecture

ISSN 0085-7130©Telenor ASA 2007

SAML does not standardize all aspects of identity management. SAML addresses one key aspect of identity management, namely that of how identity information can be communicated from one domain to another. A full identity management solution will also define mechanisms for, amongst other aspects, provisioning (the establishment and subsequent management of accounts and associated privileges), authentication (how an entity proves their right to lay claim to a particular identity), or access control (how the rules for specifying what individual identities are allowed to do are captured).

SAML is a flexible and extensible standard designed to be used – and customized if necessary – by other standards. The Liberty Alliance, the Internet2 Shibboleth project [27], and the OASIS Web Services Security (WS-Security) Technical Committee have all adopted SAML as a technological underpinning for various purposes.

In practical terms, SAML consists of a set of specifications and XML schemas, which together define how to construct, exchange, consume, interpret, and extend security assertions for a variety of purposes.



1 Authenticated browser user on server A requests access to database on server B. Server A generates URL redirect, which contains SAML artifact, to server B

2 Browser redirects user to server B, which receives artifact pointing to the assertion on server A

3 Server B sends artifact to server A and gets full assertion

4 Server B checks assertion and either validates or rejects user's request for access database

Figure 3 SAML: How it works [3]

2.3 SPML – Service Provisioning Markup Language

Service Provisioning Markup Language (SPML) [4] is an XML-based provisioning standard developed within OASIS, which defines a standard language for exchanging provisioning messages. These messages can be requests to add, modify, or delete user accounts, enable or disable access, grant or revoke access rights, change passwords, and all other types of provisioning tasks. By using SPML, heterogeneous systems can easily participate in provisioning business processes without needing complex and expensive integration.

The goal of SPML is to allow organizations to securely and quickly set up user interfaces for Web services and applications, by letting enterprise platforms such as Web portals, application servers, and service centers generate provisioning requests within and across organizations. This can lead to automation of user or system access and entitlement rights to electronic services across diverse IT infrastructures, so that customers are not locked into proprietary solutions.

SPML is a valuable tool for the enterprise to use to ensure that all of its various systems that need to participate in provisioning business processes can do so automatically. It can be used to seamlessly integrate with the other systems in the enterprise, and particularly with the identity management and workflow tool that implement and control the business provisioning processes.

In addition, SPML is extremely useful to federated networks, because it allows the enterprises participating in the federation to exchange provisioning messages via the common SPML language using web services, instead of needing to directly integrate their disparate systems. SPML can enable the federation to grow as more members join without imposing any integration burden on the existing participants. Figure 4 indicates how SPML can be used to enable enterprises to easily exchange provisioning messages.

For example, a supply partner (Company A) goes to its partner's (Company B) supply chain portal and requests access to its inventory data, which is stored in a back-office system. In response, Company B initiates a request using SPML to communicate with SPML-enabled identity management software. After automatically acquiring the appropriate permissions, Company B grants the appropriate access levels to Company A to gain access to the data it needs.

This process takes place without the need for the portal environment to have an intimate understanding of the back-office environment. In other words, it's all automatic. The prototype encompasses all of the provisions of the proposed SPML standard while also leveraging the benefits of the Security Assertion Markup Language (SAML).

The SPML 1.0 specification supports identifying principles using the OASIS Security Assertion

ISSN 0085-7130

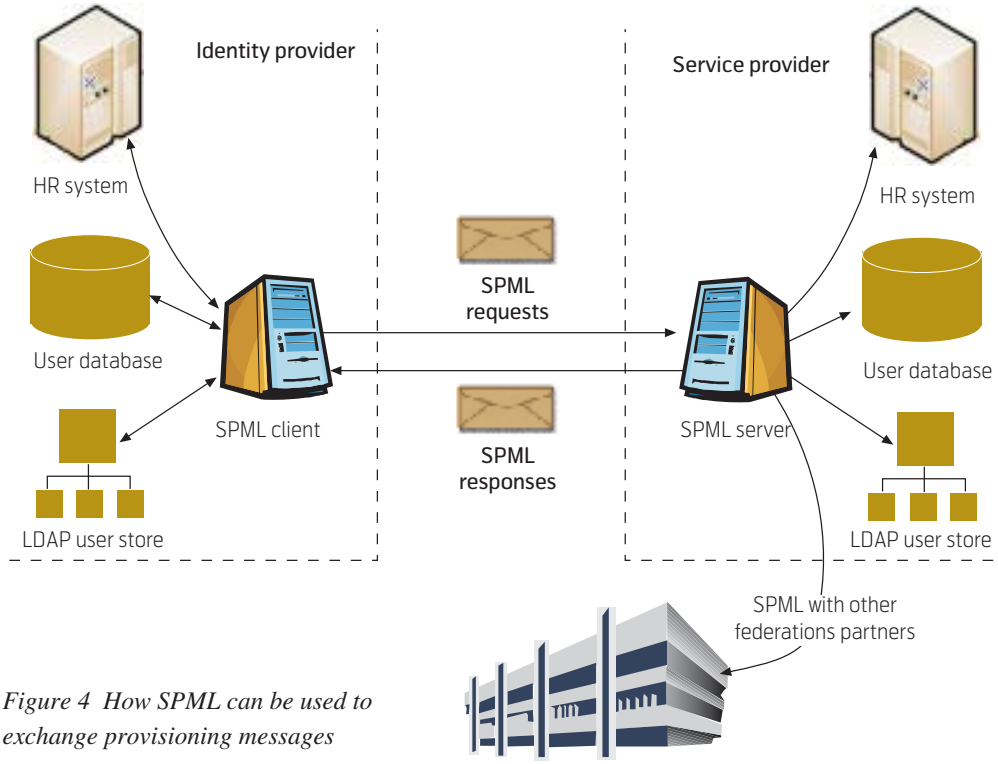


Figure 4 How SPML can be used to exchange provisioning messages

Markup Language (SAML) and Project Liberty standards. Additionally, the SPML 1.0 specification has been designed to accommodate the use of the OASIS Web Services Security (WSS) specification, XML Digital Signatures [14], and XML Encryption [13].

2.4 XACML – eXtensible Access Control Markup Language

In a nutshell, eXtensible Access Control Markup Language (XACML) [5] is a general-purpose access control policy language. XACML is an OASIS [9] standard that describes both a policy language and an access control decision request/response language (both encoded in XML). The policy language is used to describe general access control requirements, and has standard extension points for defining new functions, data types, combining logic, etc. The request/response language lets you form a query to ask whether or not a given action should be allowed, and interpret the result. The response always includes an answer about whether the request should be allowed using one of four values: Permit, Deny, Indeterminate (an error occurred or some required value was missing, so a decision cannot be made) or Not Applicable (the request can't be answered by this service).

The typical setup is that someone wants to take some action on a resource. They will make a request to whatever actually protects that resource (like a file system or a web server), which is called a Policy Enforcement Point (PEP). The PEP will form a request based on the requester's attributes, the resource in question, the action, and other information pertaining to the request. The PEP will then send this request to a Policy Decision Point (PDP), which will look at the request, find some policy that applies to the request, and come up with an answer about whether access should be granted. That answer is returned to the PEP, which can then allow or deny access to the requester. Note that the PEP and PDP might both be contained within a single application, or they might be distributed across several servers. In addition to providing request/response and policy languages, XACML also provides the other pieces of this relationship, namely finding a policy that applies to a given request and evaluating the request against that policy to come up with a yes or no answer.

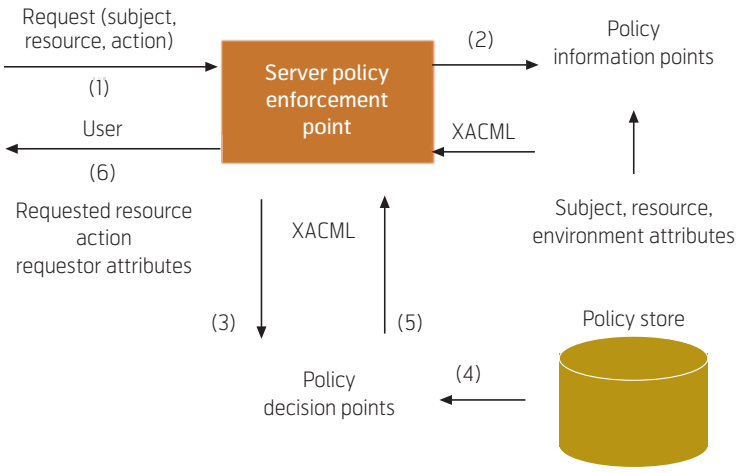


Figure 5 XACML Key Concepts [6]

ISSN 0085-7130©Telenor ASA 2007

2.5 XKMS – XML Key Management Specifications

XML Key Management Specification (XKMS) [7] utilizes the web services framework to make it easier for developers to secure inter-application communication using public key infrastructure (PKI). XKMS is a protocol developed by W3C which describes the distribution and registration of public keys. Services can access an XKMS compliant server in order to receive updated key information for encryption and authentication. XKMS is suitable for use in conjunction with the proposed standard for XML Signature (XML-SIG) [14] developed by the World Wide Web Consortium (W3C) [11] and the Internet Engineering Task Force (IETF) [17] and an anticipated companion standard for XML encryption.

The XKMS comprises two parts: the XML Key Information Service Specification (X-KISS) and the XML Key Registration Service Specification (X-KRSS).

The X-KISS specification defines a protocol for a Trust service that resolves public key information contained in XML-SIG elements. The X-KISS protocol allows a client of such a service to delegate part or all of the tasks required to process <ds:KeyInfo> elements. A key objective of the protocol design is to minimize the complexity of application implementations by allowing them to become clients and thereby shielded from the complexity and syntax of the underlying PKI used to establish trust relationships. These may be based upon a different specification such as X.509/PKIX, SPKI or PGP.

The X-KRSS specification defines a protocol for a web service that accepts registration of public key information. Once registered, the public key may be used in conjunction with other web services including X-KISS.

Both protocols are defined in terms of structures expressed in the XML Schema Language, protocols employing the Simple Object Application Protocol (SOAP) v1.1 and relationships among messages defined by the Web services Definition Language (WDSL) v1.0.

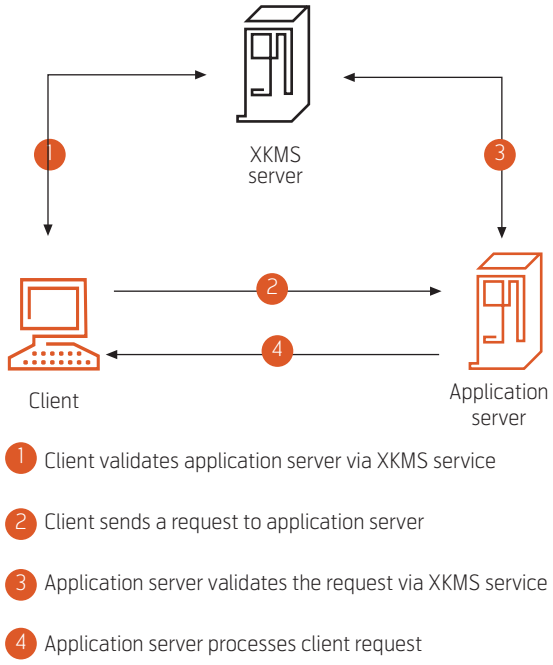


Figure 6 XKMS: How it works [8]

3 Standardisation Bodies

3.1 Liberty Alliance

The Liberty Alliance Project [1] is a consortium of commercial and non-commercial organizations working to "support the development, deployment and evolution of an open, interoperable standard for federated network identity. The vision of the Liberty Alliance is to enable a networked world in which individuals and businesses can more easily conduct transactions while protecting the privacy and security of vital identity information. To accomplish its vision, the Liberty Alliance has established an open standard for federated network identity through open technical specifications that will:

- Support a broad range of identity-based products and services;
- Enable commercial and non-commercial organizations to realize new revenue and cost saving opportunities that economically leverage their relationships with customers, business partners, and employees;
- Provide consumers with a choice of identity provider(s), the ability to link accounts through account federation, and the convenience of single sign-on when using any network of connected services and devices;
- Increase ease of use for *e*Commerce consumers;
- Help stimulate *e*Commerce.

3.2 OASIS

The official OASIS homepage [9] describes "OASIS is a not-for-profit consortium that drives the development, convergence and adoption of open standards for the global information society."

ISSN 0085-7130

The consortium is divided into several groups, in which IDtrust [10] is one. "The OASIS Identity and Trusted Infrastructure (IDtrust) Member Section promotes greater understanding and adoption of standards-based identity and trusted infrastructure technologies, policies, and practices. The group provides a neutral setting where government agencies, companies, research institutes, and individuals work together to advance the use of trusted infrastructures, including the Public Key Infrastructure (PKI).

3.3 W3C

"The W3C (World Wide Web Consortium) [11] develops interoperable technologies (specifications, guidelines, software and tools) to lead the Web to its full potential. W3C is a forum for information, commerce, communication, and collective understanding." The W3C conducts work in the following areas:

- Document Object Model (DOM)
- Extensible Markup Language (XML)
- Hyper Text Markup Language (HTML, HTML2)
- Rich Web Client
- Security
- Semantic Web
- Style (CSS, with the recent release of CSS 3)
- Synchronized Multimedia (SMIL)
- Web Services (SOAP, WSDL)

The W3C member list consists of most of the world's influential ICT-groups and companies, and is headed by Sir Tim Berners-Lee, the primary author of the original URL, HTTP-protocol and HTML.

3.3.1 W3C XML encryption

The XML Encryption Working Group [12] defines its mission in the following way:

> "to develop a process for encrypting/decrypting digital content (including XML documents and portions thereof) and an XML syntax used to repr sent the
>
> - encrypted content;
> - information that enables an intended recipient to decrypt it."

XML Encryption provides end-to-end security for applications that require secure exchange of structured data. XML itself is the most popular technology for structuring data, and therefore XML-based encryption is the natural way to handle complex requirements for security in data interchange applications.

An example of XML vs. Encrypted XML [13] can be seen in the box below:

The last XML fragment is encrypted. The XML Encryption Workgroup does not deal with signatures and key exchange, which is left to other Work Groups in the W3C.

3.3.2 W3C XMLsig – XML Digital Signature

In cryptography, a digital signature or digital signature scheme is a type of asymmetric cryptography used to simulate the security properties of a signature in digital, rather than written form. Digital signature schemes normally give two algorithms, one for signing which involves the user's secret or private key, and one for verifying signatures which involves the

```
<purchaseOrder>
    <Order>
        <Item>book</Item>
        <Id>123-958-74598</Id>
        <Quantity>12</Quantity>
    </Order>
    <Payment>
        <CardId>123654-8988889-9996874</CardId>
        <CardName>visa</CardName>
        <ValidDate>12-10-2004</ValidDate>
    </Payment>
</purchaseOrder>

<?xml version='1.0' ?>
<EncryptedData xmlns='http://www.w3.org/2001/04/xmlenc#'
        Type='http://www.isi.edu/in-notes/iana/assignments/media-types/text/xml'>
    <CipherData>
        <CipherValue>A23B45C56</CipherValue>
    </CipherData>
</EncryptedData>
```


ISSN 0085-7130©Telenor ASA 2007

user's public key. The output of the signature process is called the "digital signature."

The mission of the DSig Work Group [14] is to develop an XML compliant syntax used for representing the signature of Web resources and portions of protocol messages (anything referable by a URI) and procedures for computing and verifying such signatures. This Work Group is a joint effort between W3C [11] and IETF [17].

3.3.3 W3 Platform for Privacy Preferences (P3P)

P3P [15] is another Working Group of W3C. The Platform for Privacy Preferences Project (P3P) enables websites to express their privacy practices in a standard format that can be retrieved automatically and interpreted easily by user agents. P3P user agents will allow users to be informed of site practices (in both machine- and human-readable formats) and to automate decision-making based on these practices when appropriate. Thus users need not read the privacy policies at every site they visit.

The P3P design was officially recommended on April 16, 2002. The P3P 1.1 is now ready for implementation from the Work Group's point of view and awaits interest and acceptance from the browser implementers.

3.4 Web Services Interoperability (WS-I)

WS-I [16] is an open industry organization chartered to promote Web services interoperability across platforms, operating systems, and programming languages.

The WS-I delivered the 'WS-I Basic Profile 1.0' recommendation in August 2003, which has been widely adopted by web services vendors, solution providers and standards development organizations since.

WS-I creates, promotes and supports generic protocols that are independent of any action indicated by a message, other than those actions necessary for its secure, reliable and efficient delivery. The interoperability properties ensure suitability for multiple operating systems and multiple programming languages.

3.5 IETF

The Internet Engineering Task Force [17] is a large open international community of network designers, operators, vendors and researches concerned with the evolution of the Internet architecture and the smooth operation of the Internet. It is open to any interested individual. The IETF are known for a large number of Requests for Comment articles regarding Internet technology and protocols in particular (such as the Transport Control Protocol and the Internet Protocol in particular).

The community is divided into Working Groups focusing on different areas. These areas include:

- Applications Area
- General Area
- Internet Area
- Operations and Management Area
- Real-time Applications and Infrastructure Area
- Routing Area
- Security Area
- Transport Area

The IETF cooperates closely with the W3C and ISO/IEC standards bodies.

3.6 Open Group

The Open Group [18] is a vendor-neutral and technology-neutral consortium whose vision of boundary-less information flow will enable access to integrated information, within and among enterprises, based on open standards and global interoperability.

The Open Group works with customers, suppliers, consortia and other standards bodies to:

- Capture, understand and address current and emerging requirements, and establish policies and share best practices;
- Facilitate interoperability, develop consensus, and evolve and integrate specifications and open source technologies;
- Offer a comprehensive set of services to enhance the operational efficiency of consortia;
- Operate the industry's premier certification service.

Examples of well known and widely adopted standards for computing are Common Operating Environment Platform (COE), Common Object Request Broker Architecture (CORBA), POSIX, WAP and UNIX.

3.7 ISO/IEC

ISO (International Organization for Standardization) [19] is the world's largest developer and publisher of international standards. ISO is a network of the national standards institutes of 155 countries, one member per country, with a Central Secretariat in Geneva, Switzerland, that coordinates the system.

ISSN 0085-7130

While ISO defines itself as a non-governmental organization, its ability to set standards that often become law, either through treaties or national standards, makes it more powerful than most NGOs. In practice, ISO acts as a consortium with strong links to governments.

ISO has formed a close cooperation with IEC (International Electro-technical Commission). The International Electro-technical Commission [20] (IEC) is a not-for-profit, non-governmental international standards organization that prepares and publishes International Standards for all electrical, electronic and related technologies – collectively known as "electrotechnology". IEC standards cover a vast range of technologies from power generation, transmission and distribution to home appliances and office equipment, semiconductors, fibre optics, batteries, solar energy, nanotechnology and marine energy, to mention just a few.

3.8 ITU Focus Group on Identity Management (FG IdM)

The Focus Group on Identity Management was established by Study Group 17 at its 6-15 December 2006 meeting. The objective of the Focus Group is to facilitate the development of a generic Identity Management framework, by fostering participation of all telecommunications and ICT experts on Identity Management. The FG IdM is open to ITU Member States, Sector Members and Associates as well as any individual from a country which is a member of ITU willing to contribute to the work; this includes individuals who are also members or representatives of interested Standards Development Organizations. The FG IdM reports to SG 17.

3.9 Internet2 – (Shibboleth)

Internet2 [27] is the foremost US advanced networking consortium. Led by the research and education community since 1996, Internet2 promotes the missions of its members by providing both leading-edge network capabilities and unique partnership opportunities that together facilitate the development, deployment and use of revolutionary Internet technologies.

By bringing research and academia together with technology leaders from industry, government and the international community, Internet2 promotes collaboration and innovation that has a fundamental impact on the future of the Internet. Internet2 is carrying out the Shibboleth project which is aiming at a standardized open source middleware for Web Single Sign-On.

3.10 ICANN (Internet Corporation for Assigned Names and Numbers)

ICANN is responsible for the global coordination of the Internet's system of unique identifiers. These include domain names (like .org, .museum and country codes like .uk), as well as the addresses used in a variety of Internet protocols. Computers use these identifiers to reach each other over the Internet. Careful management of these resources is vital to the Internet's operation, so ICANN's global stakeholders meet regularly to develop policies that ensure the Internet's ongoing security and stability.

3.11 DDI (Data Documentation Initiative)

The Data Documentation Initiative [21] is an international effort to establish a standard for technical documentation describing social science data. A membership-based Alliance is developing the DDI specification, which is written in XML.

DDI began in 1995 and brings together data professionals from around the world to develop the standard. The DDI specification provides a format for content exchange and preservation of information. Version 3.0 of the DDI standard was recently released.

3.12 IDsec

IDsec [22] is a mechanism that provides a digital identity (also known as Virtual Identity) for users on the Internet. Users may allow Internet service providers to access their User Profile data. As such it can be an alternative to MS Passport.

IDsec presents a generic mechanism for establishing Virtual Identities on the Internet, which standardises protocols and interfaces for exchanging identity information between users and service providers in a secure manner. It enables users to reuse profile information across Internet services and service providers to delegate (part of) their customer information maintenance.

The architecture of IDsec is specified in an IETF Internet Draft.

4 IdM Frameworks & Systems

4.1 Microsoft's CardSpace

Windows CardSpace is Microsoft's latest effort on the identity management front. The client side comes preinstalled with Windows Vista and it is available through .NET Framework upgrades for Windows XP. The solution uses several standards for identity man-

ISSN 0085-7130©Telenor ASA 2007

agement, among them the WS-Trust [26] specification which allows someone to request a security token containing a set of claims from an Identity Provider.

When a user needs to submit personal information to a CardSpace enabled website, the website will demand a set of claims or a token from the user. The CardSpace application will then appear on the user's screen, as shown in Figure 7. It locks the display so that only the CardSpace application is accessible. The user then selects a card, either self-issued or managed, which is used to perform the transaction of security tokens and personal information.

The CardSpace architecture consists of Relying Parties (RP), which are service providers and the Security Token Service (STS), which resembles what has previously been described as Identity Providers.

The major drawback with the CardSpace initiative is that it is tightly integrated with Microsoft products (Windows Vista and .NET Framework); it is rather complex and not entirely trivial to integrate with alternative authentication mechanisms like smartcards, biometrics etc. However, as the system matures and more documentation is released, the weaknesses of CardSpace will hopefully be improved.

Triggering the CardSpace application is also initially only supported through the Internet Explorer Webbrowser. However, third parties have implemented extensions for e.g. Firefox, which allow triggering the CardSpace identity selector application from other browsers as well [24].

4.3 Higgins

Higgins [25] is an open-source framework and collaborative project which among other things develops components that can be used to build the different parts of an identity management system. The project has received technology contributions from IBM and Novell, among others.

The goal is to develop interoperable, protocol- and platform-independent solutions, and this is accomplished by providing developers with a common API for identity management, instead of requiring support for several different identity management solutions. There are two major categories of Higgins components:

Lower-level components can be used to create identity services such as attribute services, token services and relying party Web-sites (i.e. service providers) and services.

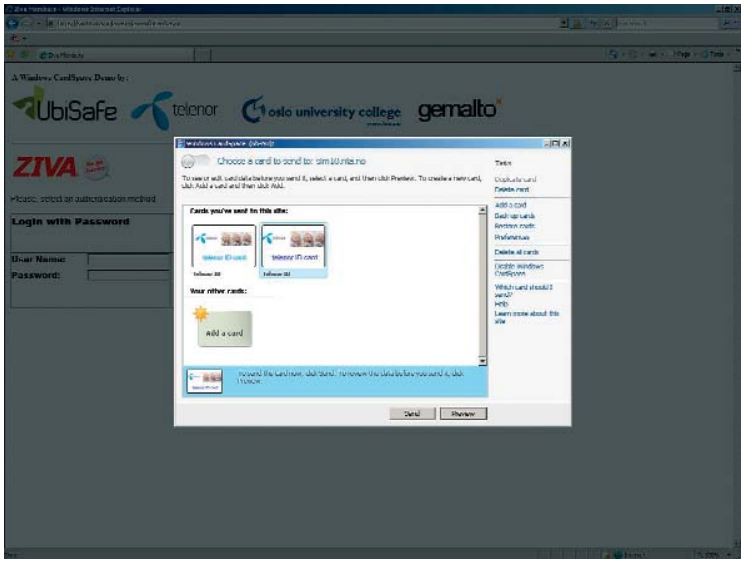


Figure 7 CardSpace appears, locks the display and lets the user pick an identity card to use

Upper-level components can be used to create user-centric applications which allow the user to view, employ and manage his/her various identities (i-cards).

More specifically, Higgins' upper-level components can be used to build identity agents which allow users to accept i-cards from card issuing sites (i.e. identity providers), they can be used to create self-issued cards, manage a user's set of cards and use these cards towards service providers (relying parties) or local applications.

The Higgins project has been working on achieving interoperability with Microsoft CardSpace-compatible card issuing and relying sites (which explains the terminology used) as well as being interoperable with OpenID providers and OpenID service providers (see section 4.4 for more details about OpenID).

Several agent deployment configurations are supported, perhaps most interestingly a Web-based identity agent for Internet Explorer and Firefox (on both Windows and Linux). For Firefox, the architecture includes an extension similar to that of CardSpace, which will allow users to select i-cards when authentication is required towards service providers.

4.4 OpenID

OpenID [23] is an open, decentralized, free framework for user-centric digital identity management. OpenID takes advantage of already existing internet technology (URI, HTTP, SSL, Diffie-Hellman) and realizes that people are already creating identities for themselves whether it be at their blog, photo stream, profile page, etc. With OpenID you can easily transform one of these existing URIs into an account

ISSN 0085-7130

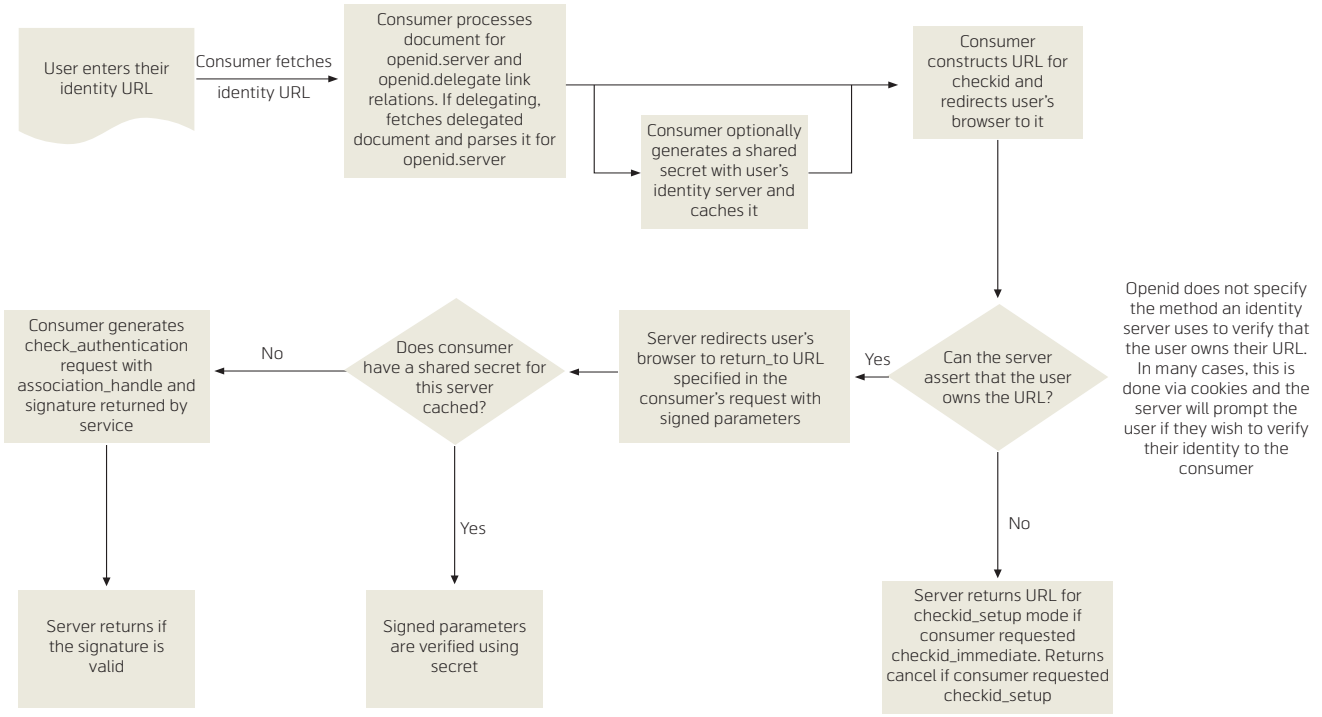


Figure 8 Protocol flow for OpenID, taken from [23]

which can be used at sites which support OpenID logins.

Identities in OpenID are URIs, e.g. http://jorstad.ubisafe.no. Some of the advantages of using URIs are that they are relatively simple, they are pervasive (used a lot) and easy to remember. The authentication process in practice involves verifying that the user owns a certain URI (i.e. their identity).

The authentication process with OpenID consists of five major steps, as illustrated in Figure 8. These steps are:

1 The end-user visits a website, e.g. a Web-shop, forum, blog or similar.

2 The user enters his/her OpenID URI (e.g. http://jorstad.ubisafe.no).

3 The user's Web-browser is then redirected to an OpenID Identity Provider.

4 At the Identity Provider's site, the user enters his/her credentials, in most cases a password.

5 The user's Web-browser is then redirected back to the visited website.

In OpenID terminology, the web site which the user wishes to log in to (i.e. the service provider) is called a consumer.

Anyone can establish an OpenID Identity Provider; there is no need for a particular permission or registration process. Thus, it is possible to establish a personal identity provider, an identity provider for a community or for the general public. Libraries for implementing OpenID identity providers are available for various platforms and languages, e.g. for PHP, Ruby, Perl, Java etc.

One of the main strengths of OpenID is perhaps the delegation of verification. It is possible to use a URI which is not registered by any OpenID Identity Provider as user identity, thus the identity can persist even when identity providers disappear. This is solved with delegation, which is realised by adding a certain code snippet to the HEAD section of the Web-page hosted at the index of the said URI. The part could for example look like in the box below.

The major advantages of OpenID are:

- Highly distributed;
- Flexible – users can keep identity even when identity provider disappears (using delegation with their

```
<head>
    <link rel="openid.server" href="http://www.someidp.no">
    <link rel="openid.delegate" href="http://jorstad.someidp.no/">
</head>
```


ISSN 0085-7130©Telenor ASA 2007

homepage URI as identity to different identity providers);

- Lightweight solution.

4.5 Shibboleth

Shibboleth [27] is standards-based, open source middleware software which provides Web Single Sign-On (SSO) across or within organizational boundaries. It allows sites to make informed authorization decisions for individual access of protected online resources in a privacy-preserving manner.

The Shibboleth software implements the OASIS SAML v1.1 [2] specification, providing a federated Single Sign-On and attribute exchange framework. Shibboleth also provides extended privacy functionality allowing the browser user and their home site to control the Attribute information being released to each Service Provider. Using Shibboleth-enabled access simplifies management of identity and access permissions for both Identity and Service Providers. Shibboleth is developed in an open and participatory environment, is freely available, and is released under the Apache Software License.

4.5 SourceID

The SourceID Identity Platform [28] is an open source project focused on providing solutions for federated identity. SourceID has two Identity Platform projects, SourceID.Java for J2EE Application Server deployment, and SourceID.NET for Microsoft platform deployment. SourceID.Java & .NET provides a comprehensive developer's framework for implementing the Liberty Protocol v1.1 – also known as Phase I, or the Liberty ID-FF Federation Framework. SourceID also has a SAML v1.1 project and has announced a project around WS* – integrating WS-Security, WS-Trust, WS-Policy, and WS-Federation into the SourceID Identity Platform.

4.6 OpenSPML

The SPML Provisioning project develops an open source client code that supports the OASIS Service Provisioning Markup Language (SPML) standard. OpenSPML [29] is a cooperative initiative by independent software vendors and implementers of the SPML version 1.0 specification. Initially developed in Java™, the OpenSPML client code is expected to be available in other languages in the near future.

4.7 DotGNU Virtual Identities

The Virtual Identity Project proposed a "Virtual ID" system [30] which is an integrated solution to Authorization, Customization, Selected Sharing of Private Data. Authentication and access to private information are peer-to-peer to preserve local storage of those things which should remain in private users' hands.

5 Conclusion

This has by no means been an exhaustive overview because there are a lot of activities due to the raised awareness of the crucial role of IdM in almost any digital system. We do hope that we have succeeded in providing you with enough information and pointers to further research.

References

1 *Liberty Alliance Project*. October 22, 2007 [online] – URL: http://www.projectliberty.org

2 *SAML*. October 22, 2007 [online] – URL: http://www.oasis-open.org/committees /tc_home.php?wg_abbrev=security

3 *SAML sets up single sign-on*. October 22, 2007 [online] – URL: http://www.networkworld. com/news/tech/2003/0421techupdate.html

4 *SPML*. October 22, 2007 [online] – URL: http://www.oasis-open.org/committees /tc_home.php?wg_abbrev=provision

5 *XACML*. October 22, 2007 [online] – URL: http://www.oasis-open.org/committees /tc_home.php?wg_abbrev=xacml

6 *Introduction to XACML*. October 22, 2007 [online] – URL: http://dev2dev.bea.com /pub/a/2004/02/xacml.html

7 *XKMS*. October 22, 2007 [online] – URL: http://www.w3.org/TR/xkms

8 *XKMS does the heavy work of PKI*. October 22, 2007 [online] – URL: http://www.networkworld .com/news/tech/2003/0908techupdate.html

9 *OASIS*. October 22, 2007 [online] – URL: http://www.oasis-open.org

10 *IDtrust*. October 22, 2007 [online] – URL: http://www.oasis-idtrust.org

11 *W3C*. October 22, 2007 [online] – URL: http://www.w3.org

12 *XML Encryption Working Group*. October 22, 2007 [online] – URL: http://www.w3.org /Encryption

13 *XML Encryption*. October 22, 2007 [online] – URL: http://www.ibm.com/developerworks /xml/library/x-encrypt

ISSN 0085-7130©Telenor ASA 2007

14 *XML Digital Signature*. October 22, 2007 [online] – URL: http://www.w3.org/TR/xmldsig-core/#sec-Overview

15 *Platform for Privacy Preferences*. October 22, 2007 [online] – URL: http://www.w3.org/P3P

16 *Web Services Interoperability*. October 22, 2007 [online] – URL: http://www.ws-i.org

17 *IETF*. October 22, 2007 [online] – URL: http://www.ietf.org

18 *Open Group*. October 22, 2007 [online] – URL: http://www.opengroup.org

19 *ISO/IEC Information Centre*. October 22, 2007 [online] – URL: http://www.standardsinfo.net/isoiec/index.html

20 *IEC International Electrotechnical Commission*. October 22, 2007 [online] – URL: http://www.iec.ch

21 *DDI Data Documentation Initiative*. October 22, 2007 [online] – URL: http://www.icpsr.umich.edu/DDI

22 *Idsec*. October 22, 2007 [online] – URL: http://idsec.sourceforge.net

23 *OpenID Authentication 1.1*. October 22, 2007 [online] – URL: http://openid.net/specs/openid-authentication-1_1.html, May 2006

24 *CardSpace-Firefox*. October 22, 2007 [online] – URL: http://www.perpetual-motion.com

25 *Higgins*. October 22, 2007 [online] – URL: http://www.eclipse.org/higgins/index.php

26 *Web Services Trust Language (WS-Trust)*. October 22, 2007 [online] – URL: http://www.ibm.com/developerworks/library/specification/ws-trust

27 *Shibboleth*. October 22, 2007 [online] – URL: http://shibboleth.internet2.edu

28 *SourceID*. October 22, 2007 [online] – URL: http://www.sourceid.org

29 *OpenSPML*. October 22, 2007 [online] – URL: http://www.openspml.org

30 *DotGNU Virtual Identities*. October 22, 2007 [online] – URL: http://www.gnu.org/software/dotgnu/auth.html

Nguyen Duy Hinh has an MSc in Informatics from the University of Oslo. He first joined Norwegian National Rail Administration (Jernbaneverket), where he worked on ATC (Automatic Train Control) for the railway signalling system. In 1999 he started work in Tandberg Television (TT) where he worked with embedded programming (C programming) for different TT's MPEG-2 decoder products. Since 2005 he has been working in Linus AS, a Norwegian consulting and systems house specialising in products for mobile communications and custom development and system integration for telcos. He participated in the Eureka Celtic Fidelity project in 2005 and 2006 and is currently participating in the Eureka Mobicome project.

email: h.nguyen@linus.no

Sjur Millidahl received his MSc in Telematics from the Norwegian University of Science and Technology, Trondheim, in 2005. The same year he started work as system developer in Linus AS, a Norwegian consulting and systems house specialising in products for mobile communications and custom development and system integration for telcos. He participated in the Eureka Celtic Fidelity project in 2005 and 2006 and is currently participating in the Eureka Mobicome project. His research interests include telematics, quality of service and convergence in telecommunication networks. His daily work includes server side programming in Java. Other work related activities are systems design and development, database programming and systems integration.

email: s.millidahl@linus.no

For a presentation of Do Van Thuan and Ivar Jørstad, please turn to page 18 and 10, respectively.

ISSN 0085-7130©Telenor ASA 2007

Identity Management in General and with Attention to Mobile GSM-based Systems

TOR HJALMAR JOHANNESSEN

Tor Hjalmar Johannessen is Senior Adviser in Telenor R&I

Identity Management (IdM) has received increasing interest and attention with the growth of telecommunication systems. The reason for this is its impact on security and, consequently, both for network operator and possibly third party business aspects. IdM concerns the protection of business, regulatory aspects, as well as new opportunities. This paper illustrates some of the aspects, decouples some of its complexity and addresses several functional and business issues.

Introduction

Genesis 2.20: *"So the man gave names to all the livestock, the birds of the air and all the beasts of the field."*

Identification systems are as old as the history of mankind. Throughout history they have been furnished in many different ways and with different goals. Apart from the basic need to distinguish people and their belongings in a small community like a family, accounting systems, whether targeting taxation, marriage, heritage, ownership, or payment, etc. etc., IdM relies on the means to pinpoint who to pay, receive, communicate with, marry, divorce, sue, share with, prosecute or bury.

In modern IT systems IdM schemes are furnished to provide the baseline for the registries and administration of employees, subscribers and other customers, organization members, citizens, hospital patients, etc. etc.

Depending on the context, executing technology and the administrational needs, various types of IdM attributes and functionality apply. There is no common standard to cover them all since they mostly rely on independent legacy backgrounds and technical and sociological systems. Therefore different requirements apply, like:

- *Coverage and scalability:* Is the scoping range global or local, e.g. to specific organizations?
- *Security and protection features:* Which assets are associated with the system, e.g. authorizations, access control, monetary transactions, business or military secrets, personal sensitive information etc.? Which security measures are involved (encryption, authentication, key management, password schemes …)?
- *Interoperability:* Does the IdM system follow standardised schemes, functions and procedures?
- *Applicability:* Is the IdM system tailor-made for very specific purposes, or is it furnished for general usage?
- *Purpose:* Which user groups and services are considered: Employees, customers, bank accounts, social security services, passport, roles, and so on?

A modern community is strongly dependent on IT-based identity systems to provide efficient services and high productivity.

The purpose of this article is to shed some light on the IdM concept, especially seen from a telecom operator's point of view. Furthermore, since there has been a growing focus on the role of the SIM, the identity carrier in mobile GSM systems, this article will especially pay attention to the SIM and its possibility also to play a role in non-GSM systems and third party applications by enhancing, aligning and combining relevant IdM systems.

With respect to standardisation, IdM is of growing importance reflected by the fact that large organizations like ITU-T and GSMA[1)] have launched IdM Task force groups to work with and outline the topic to cover the new challenges of interoperability, convergence between fixed/broadband and mobile networking systems, integration of payment and money transfer systems / *e*Banking with network systems and to cover the ever growing number of different SPs that now uses the different networks as their operative marketplace.

Finally, note that the very baseline of the Internet is the IP addressing system. If this had been inflexible and non-scalable the Internet would probably never been a success. Its popularity has shown, however, that the initially immense 32 bits address range covering approximately 4.3 billion different addresses was designed too small in IPv4. GSM telecommunications addressing relies on other schemes, but it is interesting to observe that the

1) *www.gsmworld.com*

ISSN 0085-7130

number of mobile subscribers in 2007 will surpass 3 billion, comparable with the maximum number of different IPv4 addresses.

The big challenge today, however, is not to manage the basic identifiers, but the various security aspects associated with them, and not least all the different attributes that can be associated with the basic identifiers. This touches personal secrecy issues, business perspectives as well as regulatory authority principles.

Definition and Scope of IdM

The short definition of IdM can be given as follows: "IdM is a formal standardized enterprise-wide or community-wide set of processes for managing multitudes of Identities."

This requires, consequently, a definition of "Identity", which turns out to be quite complex and always depending on the context. In principle it is something that characterizes subjects and objects; e.g. a person, an entity or a concept. It may be a common property (like poisonous or green) but in its deepest meaning it is a property that can be used to distinguish something/somebody from something / something else; i.e. uniqueness property.

In information systems, identity management, sometimes referred to as identity management systems, is the management of the identity life-cycle of entities (subjects or objects) during which:

- *The identity is Established:* A name (or number) is connected to the subject or object. Normally, this is done by creating a unique (root) primary identifier (ID) that can be processed by some machine.

- *The identity is Aliased or Re-established:* A new or additional name (or number) is connected to the subject or object for use in different contexts. One root identity can have many aliases, but there must be a unique mapping correspondence to avoid ambiguities.

- *The identity is described or Profiled:* One or more attributes which are applicable to this particular subject or object may be assigned to the identity. Aliases and attributes can also be denoted secondary identifiers.

- *The identity is changed:* One or more attributes which are applicable to this particular subject or object may be altered. Normally, the root identifier is the most stable, while other attributes and aliases may be targets for frequent changes and also "come and go".

- *The identity is destroyed:* This is done when the root identifier is erased. Destruction of secondary identifiers does not destroy the identity, it only changes the descriptors or properties of the identity.

In other words, given a set of parameters like identifiers and associated ID attributes that represent a person or an entity, IdM in the broadest sense encompasses all functions that has some impact on these identifiers and attributes throughout their life-cycle.

In the narrow sense, it means to create, delete, store, rename, associate (or bind), organize and distribute these parameters. In the broader sense, it means to register, verify, protect (e.g. privacy), and also to support user-friendly facilities like Single Sign-on (SSO).

Wikipedia (www.wikipedia.com) introduces three paradigms regarding IdM, reflecting three levels of applicability:

- *The pure identity paradigm* – creation, management and deletion of identities without regard to access or entitlements;

- *The user access (log-on) paradigm* – authentication means username/password or stronger protocols and associated authorizations that a customer uses to log on to a service or services;

- *The service paradigm* – a system that delivers personalized, role-based, online, on-demand, multimedia (content), presence-based services to users and their devices.

The user access paradigm has two sublevels: authentication and authorization, where authentication means to verify a claimed identity (e.g. by running a challenge-response protocol) or password check, while authorization means to equip the user (or the subject that executes on his behalf) with a set of rights to distinguish between what is allowed and what is not. A subscription to some service is one authorization example.

The service paradigm encompasses associated IdM functions like accounting, billing etc.

The paradigms can be structured as a path for a user to achieve services, which may be IT-systems or network services in general or a more specific service provided by some Service provider.

Figure 1 shows where different IdM information for example is applied in a login/access/service sequence.

ISSN 0085-7130©Telenor ASA 2007

First a basic identifier is provided (Identity paradigm), then this is verified by the authentication process, possibly using supplied passwords or other types of authentication information (policy dependent). Then the authorization information for that identity is supplied to be granted service access (User access Paradigm).

This scheme is typical in many cases, but is somewhat generic and general. In a real case the sequence above is often repeated in several stages, e.g.

1 To get access to the local IT-system itself (and the clients it encompasses), whether this is a PC, mobile phone or some workstation;

2 To get access to the network, which may be a GSM network, IP/Internet or whatever (normally assisted by the clients at stage 1);

3 To get access to some specific service provided by a service provider, normally assisted by the clients at stage 1 and potentially IdM services from stage 2.

In the widest sense, IdM may cover elements in all three paradigms. It is therefore necessary to structure IdM with respect to its data structures as well as its functional components in order to describe which parts are relevant for the different tasks.

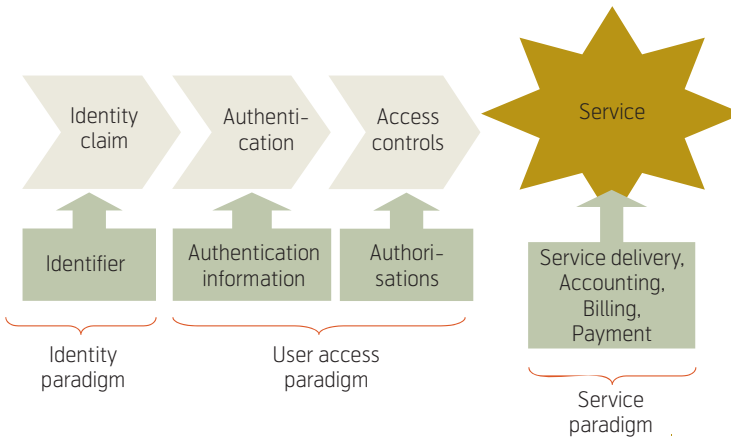


Figure 1 IdM information and IdM paradigms associated with the path to service

Grouping of IdM Attributes

ID Attribute Grouping

Partly following but also extending the different paradigms above, the identity data can be segmented into different groups. Note: In a distributed system, it is likely that different data elements are themselves distributed and may belong to different databases.

1 *The basic ID group*
This group contains Identifiers, aliases and similar numeric or alphabetic literals. An identifier's main property is to be unique within the domain. It also contains status about the identity like when created, validity period, and possibly blocking or revocation information.

Note: An identifier shall be unique to avoid ambiguity. This does not exclude the subject (or person) it represents from being anonymous. This is typical for systems that allow anybody to pick or select an identifier without executing registration operations to verify the user's true identity in a legal sense.

2 *The security group*
This group contains elements used for security enforcement functions like authentication; e.g. crypto keys for authentication, session encryption and similar; furthermore, passwords, PIN, biometrics, PKI certificates and private keys, Liberty Alliance applets, DRM keys, etc.

Which security means and functions that are associated with the parameters often depends on the security risk and threats. Protection of a bank account is somewhat different compared to basic access to the Internet. This also means that some security systems can be reused in other domains, while other cannot, due to different threat scenarios and subsequent security policy.

3 *The authorization group*
This group contains attributes for user profile capabilities like subscriptions, access control lists, physical admissions, possibly roles (i.e. for role-based access control systems), various constraints (like time of day), military classification level, DRM rights objects (RO), and so on.

A general authorization system for distributed multi-domain systems is quite difficult to achieve. This is because of the multitude of services and contexts that are hard to standardize, but also the liability aspects. The most successful distributed authorization systems, however, are found in telecommunications like GSM. But this relies on two aspects:

a Roaming- and SLA agreements to cover the liability aspects of inter-domain services;

b The limited and well-defined set of services offered (voice, SMS, 2G/3G access) and bandwidth for conveying user traffic.

ISSN 0085-7130

4 *Miscellaneous attribute groups*
There are several other ID-relevant items that can be relevant for IdM but that fail to fall into the three major groups above, for example

a Location information, telling where a specific item or person is at a certain time, alternatively, fixed location if not mobile;

b Payment capabilities that can be anything from a bank account ID to a counter in some electronic purse or prepaid telecom account;

c On-line status and accounting information (for accessed services);

d User habits and preference characteristics. This group of user habit attributes is of increasing interest, especially for service providers as *e*Commerce on the internet grows: Statistics of user habits like what, when, how etc. of their preferences of their consuming activities are basic for tailoring offers and advertisements aimed at consumers with different attitudes.

The distinction of the groups indicates their sequencing for a typical login to some service provider service through the network. The different groups above reflect increasing scoping levels of IdM that also implies increased applicability and complexity.

Level 1 – "Basic IdM": Basic (and unsecured IdM) comprises elements from group 1 only.

Level 2 – IdM with security features: IdM with security comprises group 1 enhanced with elements from group 2. This level provides e.g. authentication services.

Level 3 – IdM with authorization features: This level comprises level 1 and 2 and includes elements that can be used for access control policies (i.e. different capabilities depending on service to access). Note: It is not likely that authorization can be achieved without some kind of authentication in advance. This level provides means for access control services.

Level 4 – IdM level 2 or 3 with miscellaneous attributes enhancements. This level provides enhancements for various business application services.

The increasing complexity can be depicted as in Figure 2.

The complexity also reflects an increasing context dependency and thus an increasing interoperability problem. This is partly caused by the difficulty achieving standardized contexts and interpretations. Another issue is the fact that the concept of an Identity Provider (IdP) is many-faceted, and not straightforward to define, neither with respect to the variety of roles it may have, nor to which granularity it provides its services.

Levels 1 and 2 can be standardized to achieve global compatibility. Examples of global level 1 systems are are easily found, e.g. the IETF IP protocol and ID system for the Internet, telecom numbering systems, whether fixed or mobile networking (ISDN/MSISDN). The identifiers are typically structured in a unique fashion, but are system specific. In order to bridge the GSM and the IP world, the UICC may also carry IP-oriented identifiers via the ISIM application: These identifiers (IMPI/IMPU) are handled through the IMS/HSS system e.g. for VoIP purposes.

Level 2 also reflects global standards, partly system specific, like the GSM algorithms A3/A5/A8 for 2G and Kasumi for 3G. Standardised algorithms and protocols apply, many of which are registered in the OID system, originally defined by ITU-T recommendation X.208 (ASN.1), with registries managed by IANA, ANSA and BSI.

Level 3 concerns access to network, systems, service providers etc. These elements are very context/system dependent, ranging from employee roles in some enterprise to subscription based roaming systems in mobile networking. The latter case provides global interworking and access to communication, while the former often has very local operability. The reason is obvious: a local authorization role like "manager" has as many interpretations (i.e. mapping to the set of functions that this authorization is meant to cover) as there are systems.

Level 4 has the same system dependency, but depending on applicability and acceptance in the communication community can be made very interoperable.

Examples are different methods for payment and location based services like GPS. A GPS reference and a bank account/*e*Purse can be associated to a spe-

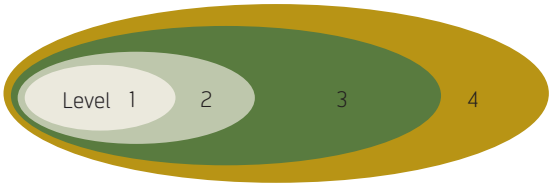


Figure 2 Scoping Levels of Identity Management

ISSN 0085-7130©Telenor ASA 2007

cific user ID in some IdM system; hence, they are targeted by IdM functions and mapping. Likewise, group 4d of user habit characteristics is very dependent on the commercial service in question. This group is also very disputed by regulatory authorities because of its association with spam.

The ID Attribute Interdependencies

Some of the attributes are more basic than others. The basic ID group is a prerequisite for the others. If the basic (or root) identifier is deleted, all other parameters become invalid and target for the garbage collector.

In a similar way, if authorization (which distinguishes between what is allowed and what is not) shall not be easily violated, it must rely on some security features like authentication. The various blocks are therefore organized into a stack.

Figure 3 shows that the basic ID group is a prerequisite for the other groups and parameters. While Authorisation attributes may rely on security attributes as a result of functional necessity, miscellaneous attributes may build on any level.

One example is that a user is given an ID (basic) and password (security) before authorization to some domain or portal (authorisation). Miscellaneous attributes like sex, age, habits and location may be built anywhere depending on policy. A bank account ID would normally be put on the very top above authorization or at least at the same level because of the risk assessment.

The ID Attribute Life Span Problem

While a social security element may be static for the whole life-time of a person, other elements can be more or less short lived. This has several reasons: A PKI certificate or crypto keys should be exchanged within a few years for security reasons. If not changed, the keys may be apt to successful attacks. The risk increases with time.

The other aspects have to do with sociological aspects: Authorization given to an employee (e.g. username/ password to the enterprise IT system) should only be revoked when the employee quits. Other constraints may be "for the day" if the user is connected through a specific access point / system. Some personal IdM attributes may have long term validity, e.g. professional degree, age above 18; and some are situation dependent, not necessarily including a time span: e.g. "enough money in the *e*Purse/ bank account for debit tranfers" or in possession of an access token.

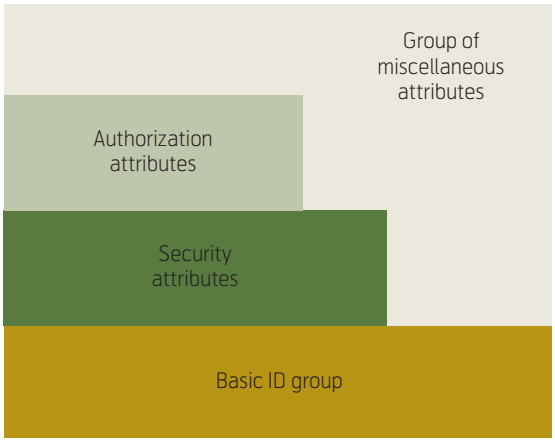


Figure 3 ID attributes grouping and inter-dependencies

ID Databases

Identity elements and attributes are stored in more or less distributed databases and registries like the X.500 (ITU), HLR/HSS (GSM), etc that can be accessed by servers through convenient protocols like LDAP (IETF), SS7/MAP, SOAP, HTTP/S etc. All managed identities belong to a "Domain", which can be commercial, governmental, organizational etc. At the user side the identity attributes are also stored in corresponding agents or clients.

An IdM system may address a single-domain only, or may be intended to cover multiple domains. In the latter case it is important to segment the information so that information sharing is limited to what is agreed between the domains. In order to provide different agreements, the different IdM attribute groups must be operated separately. In order to hide the true identity of an entity or a user, the basic ID group of identifiers may provide aliases that are revealed to entities outside the home domain.

Note: Regulatory authorities normally have comments when registries are interconnected. One should not overlook that this may raise problems for some business perspectives.

Compound and Self-proving Identifiers — Digital Certificates

A digital X.509 certificate contains not only basic identifiers Distinguished Name (DN), but also extension fields that can be equipped with profile information like roles, organizational connections etc. Furthermore, it also contains the public key of the subject or user it is assigned to, a pointer to where revocation information can be collected, and is itself a digitally signed object (by the issuing Certificate Authority (CA)). It is thus a powerful object that not only contains primary and secondary identifiers and attributes, but also means for a receiver to verify its

ISSN 0085-7130

own correctness and possibly also to verify digital objects that have been signed by the corresponding user.

A Public Key Infrastructure (PKI) is needed for the certificates to be operative and to provide scalable usage. A complete PKI contains a CA, a Registration Authority (RA), validation functions, and revocation services (CRL or OCSP).

A certificate can represent a person, a server or an organization. Several quality classes apply, e.g. the Qualified Certificate for persons as defined by ETSI TS 101 456. Certificates are used mainly in the three functional fields:

- authentication;
- electronic signature (content commitment/non-repudiation);
- mutual (symmetric) session key establishment and distribution between two communicating parties or for local object encryption. Integrity protection provisioning is intrinsic in all three cases.

The most common use today is to provide SSL/TLS-based Web server authentication (HTTPS). Functional and operational frameworks are defined by RSA, IETF, and ETSI.

Trust and Verification

Registration and Enrolment – Levels

"Shit in, simply implies Shit out"

The following story was overheard in a jail: "We had the best artists, colours, technology and paper available to produce forged money, but the 100 dollar bill we counterfeited was false".

If an IdM system is to be trusted, the process of enrolment through registration must be carried out through certain routines. NIST[2)] recommends four registrations levels, each corresponding to a 4-level authentication model:

At the lowest level (1) no requirements apply. At Level 2 and higher, the applicant supplies his or her full legal name, an address of record, and date of birth, and may, subject to the policy of the registration authority, also supply other individual identifying information. The types of ID cards and tokens needed for levels 2-4 are precisely described in the NIST Guidelines Table 1. In-person application is needed for the two highest levels (3 & 4).

All IdM systems must have a policy for registration. In order to be interoperable, e.g. in a federation case, different IdM systems must align to a common set of principles that in reality sustain a certain minimum level and adjust their discrepancies. The policy that regulates the level depends on several principles, e.g.:

- Risk assessment – which assets are at stake in the system;
- Liability – what are the liability perspectives for IdM provider, e.g. for banking services;
- Regulatory requirements – forensic and money laundering laws as well as the EU directive like the Data Retention directive[3)] (2005) may set constraints for the minimum level. In Norway it is no longer legal to distribute anonymous pre-paid SIM cards for mobile telephony.

This is also why it is impossible in many cases to federate systems with discrepant IdM registration. All in all, the basic trust of any IdM system relies heavily on the registration policy. One major obstacle for interoperability today is the lack of globally accepted and standardised levels covering all communication systems and IT-based services.

Some countries provide national means for its citizens through national registry services containing social security numbers. This is the case for the Nordic countries, but normally, this is not the case. Many countries abstain. The benefit is that local registrations are convened by check-ups in this national registry, and personal uncertainties can be released by checking whether a person's different claimed identities map to the same social security number. In Norway, the net bank account owners log into their net bank by using their social security number as user name.

Authentication

Authentication is a service that verifies a claimed or asserted identity. One of the main functions of IdM is to provision authentication means. The Security Group of IDs contains the necessary data to provide this.

In the IdM context, the authentication function can be assigned to a dedicated Authentication Provider that can either be one common resource acting on behalf

2) *NIST Special Publication 800-63 "Electronic authentication Guidelines" 2006*

3) *DIRECTIVE 2006/24/EC, 15 March 2006*

ISSN 0085-7130©Telenor ASA 2007

of service providers, even belonging to different domains, or a function local to each service provider's server.

It is necessary to distinguish between entity authentication and user authentication:

User authentication is often a local matter where the physical user provides proof-of-identity to a local client when she wants to initialize a connection like connect to a network, logon to a PC and similar. User Authentication is exercised through a combination of something the user *knows* (like a PIN or Password), *has* (like a bank card, a SIM card or a smartcard), or *is* (biometrics). A combination of two (2-factor authentication) or more items is needed to provide *strong* user authentication. 1-factor is denoted *simple*. After user authentication the user will be granted access to further applications and rights carried out by his local client.

Entity Authentication is when two processing entities with own identifiers authenticate to each other. The process can be single-sided or mutual.

There are several methods with different strengths. The weakest exchange passwords (even hashed), while the stronger methods exercise cryptographically based challenge response methods. Examples of weak methods are typically the default method for logon to a PC or accessing the Internet by username and password. GSM systems provide stronger means. This also reflects the higher assets (actually the business model of mobile operators that depend on charging the subscribers for network usage). Note that 2G or even 3G when SIM cards are used provide only single-sided authentication where the network authenticates the mobile subscriber through the A3 function in the SIM, but not mutually. This implies a certain vulnerability for man-in-the middle attacks. When USIM is used 3G has solved this by providing mutual authentication, and also stronger cryptographic methods and key lengths.

Authentication, Authorization and Accounting (AAA)

AAA is a set of functions necessary to protect a business model for network operators. IdM is deeply involved and designed to support the system in question.

IP-systems

For The IP-based, Internet networks IETF has defined the RADIUS and DIAMETER protocol suites. After IEEE enhanced the LAN-systems with WLAN wireless access protocols (802.11 etc), certain extensions were needed. Basic 802.11 access is based on the MAC addresses that are hard coded in the interface card (actually, many network interface cards have a MAC address whether it is a Bluetooth, Infrared, WLAN or fixed-line). The problems with the MAC addresses is their lack of operative management. This means that seamless hand-over and roaming among multi-domain access points is hard to handle within a business model comparable with GSM. To improve both the security and also to make the management more scalable, IETF extended their RADIUS concept with the EAP methods comprising more than 40 mechanisms. Two of them (EAP-SIM and EAP-AKA) include the usage of GSM SIM/USIM functions and authentication protocol exchange with a mobile operator without using mobile access GSM networks. This opens for the mobile operator (MNO) to become an Identity Provider offering IdM services also for the Internet/IP world.

IdM in GSM-based Systems

GSM System

The GSM world bases its AAA functions on the UICC card (SIM for 2G and USIM for 3G networking), and is designed for multi-domain roaming. The Home Location Registry (HLR) part of the MNO stores and manages all identifiers (Basic ID Group), while the Security Group is handled by the AuC part of the MNO. If IMS is added to the ID management portfolio then the MNO must also be taken care of. The HSS function comprises functions for both HLR and IMS.

IdM in GSM-based systems were built on fixed telecommunications IdM systems but enhanced to handle the roaming aspect where the user may change access network (i.e. visited network). There were two new IdM challenges associated with this:

1 To cover the business model where also a visited network access provider should be able to assure that it got paid for offered services to an unknown visitor that the visited network had no previous knowledge about;

2 To ensure that the routing systems could find a subscriber that was no longer fixed but dynamically changing his connectivity.

The business model relies on a set of bilateral business agreements between operators which comprise the kernel of the Mobile operator industry's trust model: If a home operator of a subscriber guarantees for its subscriber's resource consumption (i.e. network usage) the visited operator will offer its

ISSN 0085-7130

resources to that subscriber and be paid directly by the home operator. The payment of the visiting user is a local matter between him and his home operator.

Identification of the mobile subscriber is fulfilled through the IMSI identifier described below. When the subscriber connects to the network, the visited operator checks whether it has a roaming agreement with the home operator in that country. If so, it sends the IMSI to this home operator. If the subscriber is OK, the home operator returns an OK status including authentication and keying material information, so that the visited operator and the subscriber can perform a security handshake and deduce a session key to protect the radio link. The visited operator may now open for traffic and start accounting.

The callers in the global community address the subscriber by his MSISDN number (described below), so MSISDN is necessary to establish connection.

Basic Identifiers in GSM Systems

Actually, a subscriber can be associated with several identifiers:

IMSI, which is the basic identifier for roaming, also stored in the SIM or USIM. IMSI is structured into three parts: The country code, the operator, and the subscriber number.

In order to minimize the exposure of IMSI during hand-overs between base stations of the same visiting operator domain, the temporarily derived *TIMSI* is used locally.

ICCID, which is a serial number, hard coded into the SIM-card of the SIM-card manufacturer of the UICC chip itself. It corresponds to the MAC identifier.

MSISDN, which exposes the subscriber's ID to the public world, as it is normally entered into telephone directories like White and Yellow Pages. The MSISDN subscriber identifiers are structured according to ITU-T rec 164., indicating nation, area and subscriber ID (maximum 15 digits).

IMEI, which is a hard coded serial number in the handset device, intentionally used to block stolen devices, but due to lack of management hardly fulfilling this scope. An MNO may read it to find out which device type the subscriber actually uses (and where the SIM currently is inserted) when device specific software update is to be exercised by the MNO (through the OTA channel).

IMPI/IMPU are contained in the optional ISIM (that can be both a separate application, or processed within the USIM). These SIP-(URI) oriented identifiers are for use towards IMS-based systems to convey access to Internet services independent of the GSM roaming identifiers. A subscriber can have one IMPI (private identifier) that the IMS kernel may map to several public IMPUs. The idea is to protect the user's true identity (privacy) by providing aliases that are exposed.

Security data:

- OTA keys for end-to-end protection of MNO management operations (e.g. 3DES algorithm);
- Keys for the A3 algorithms that authenticate the client to the network;
- Keys for the A8 algorithm that generates session traffic for the A5. A5 protects the radio link between handset and base-station/base-station controller;
- Keys for the A5 (session keys).

ID Carriers in Clients

- The UICC stores the hard-coded ICCID
  Note: The UICC is also the platform for the SIM, USIM and ISIM applications.
- SIM and USIM (on the UICC) stores the IMSI and the MSISDN. SIM/USIM also stores the temporary TIMSI.
  Note: USIM may comprise the ISIM application.

  Security data:
  - Keys and vectors needed for the A3 and A8 algorithms;
  - One or more OTA keys.
- ISIM (on the UICC) is carrier for the IMPI/IMPU;
- The mobile hand-set stores the IMEI.

  Security data:
  - Session key for the A5 algorithm.

MNO ID Management Server Entities and their Basic ID Capabilities

- HLR stores IMSI, MSISDN and ICCID. HLR may also collect IMEI and store this for various usages. HLR stores the OTA key(s), one dedicated for each distributed UICC card.
- AuC stores security data:
  - Authentication vectors and keys for roaming access.

ISSN 0085-7130©Telenor ASA 2007

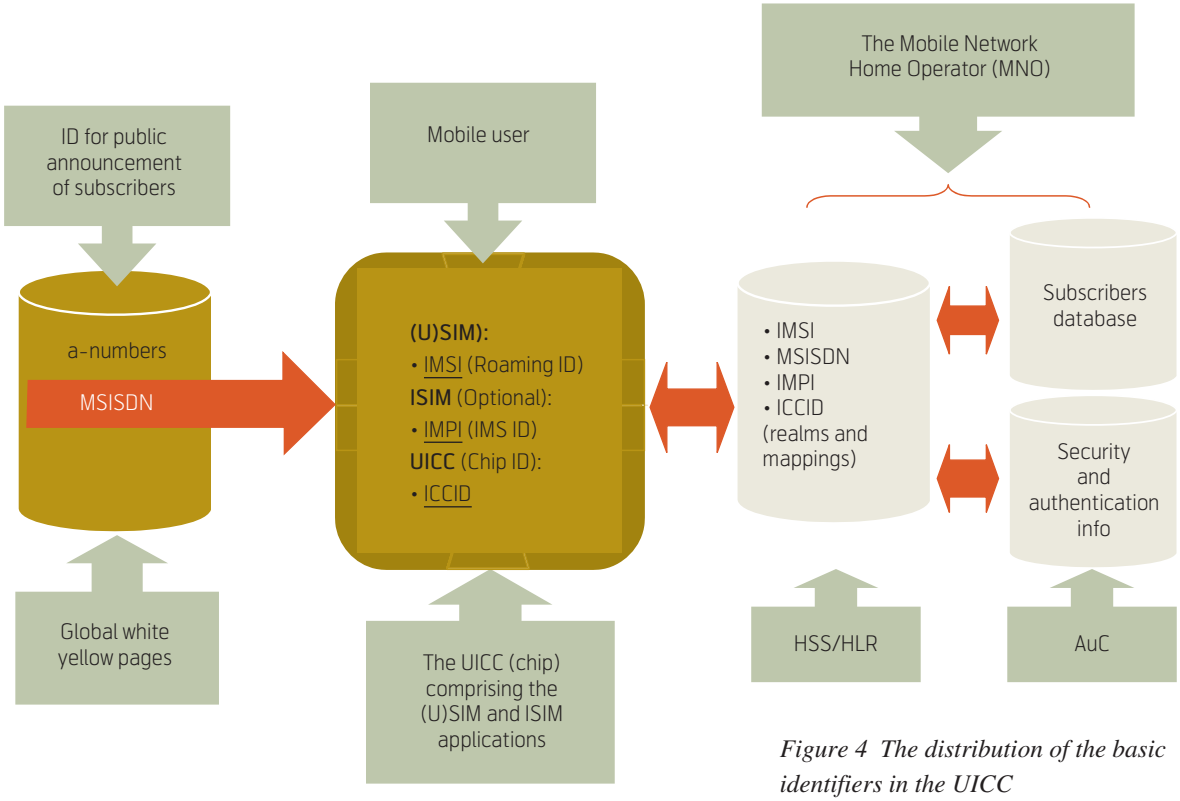


Figure 4 The distribution of the basic identifiers in the UICC

- HSS is an enhancement of the HLR, also to handle IMS systems; stores the IMPI/IMPU (and also the HLR identifiers (see above).

- VLR will for each connected roaming subscriber store:
  - IMSI and the derived TIMSI
  - MSISDN
  - Security data: Authentication vectors and session keys for roaming access.

3GPP has in its General Authentication Architecture (GAA) also defined an extended set of security data to use for WTLS protocols in GPRS communications systems, including server and client certificates. The 3GPP GAA framework[4)] is however kept outside the scope of this article. It is anticipated that its IdM role in mobile systems will increase in the years to come, so it is important to keep an eye on it.

Figure 4 describes the distribution of basic identities between the mobile user (or subscriber), the public community, the SIM (UICC) and the mobile network operator (MNO). Only the MNO has full mapping between all identifiers and the specific subscriber. The Visited Mobile Operator Location Registry (VLR) is not shown, however.

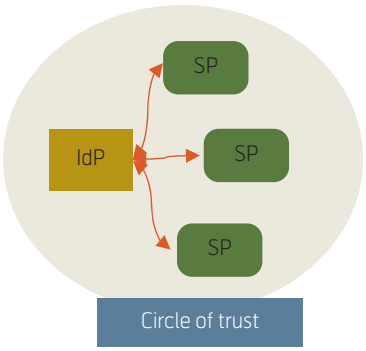
At roaming time, the Visited operator will receive IMSI from the mobile user (the IMSI indicates the MNO ID), to check whether there exists a roaming agreement. If so, the IMSI is sent to the MNO for authentication (AuC function). MNO will return authentication information and keying material to establish an A5 encrypted link with the mobile device. In addition the visited operator receives the MSISDN from the MNO. To avoid unnecessary exposure of the IMSI when the user changes base station for that visited operator, a temporary TIMSI is derived from the IMSI for local domain usage. This is an example of the privacy aspect of IdM, where aliasing is used. Here it is used to protect the true roaming identity (IMSI). Identity mapping is a convenient method for this purpose as long as the system keeps control.

Concerted IdM Concepts

Ideally, it would be fine to have a unified IdM concept managed by a single super-national organization that e.g. could provide single sign-on (SSO) features for service access in the global arena. There are many obstacles to this because of political and business diversities, constraints and interests. One thing is to agree on a unified and standardised set of meta-systems, formats, algorithms, protocols; another thing is to outsource the management of citizens, subscribers, bank account owners, other customer bases, member lists, employees etc. The Microsoft Passport SSO concept featuring global IdM intentions was attacked,

4) *3GPP TS 33.220 "Generic Authentication Architecture (GAA); Generic bootstrapping architecture"*

ISSN 0085-7130



Figur 5 Circle of Trust (I)

e.g. by the EU Commission for potential violations of personal privacy principles, and failed to achieve global acceptance. As a result, the focus was changed to federated IdM systems, where the Liberty Alliance project is in the leading technology position. Liberty Alliance, based on SAML v.2, provides means and architecture to establish Circles-of-Trust (CoT) in the multi-domain area. Federation of Identity Provider (IdP) is a basic property. Microsoft accordingly changed its concept portfolio from "Passport" and now delivers sign-on concepts, but without Circle-of-Trust abilities (Microsoft CardSpace). A third system, WS-Federation, sponsored by IBM, Microsoft, BEA Systems, Inc., RSA Security, Inc., VeriSign, Inc. is also applicable. Liberty Alliance is based on technology from SUN, but the Liberty Alliance comprises a large number of member companies[5)] ranging from vendors, operators, universities, governmental units, service providers etc. in a large number of areas.

There are two levels of CoT complexity:

The simplest level comprises one (outsourced) identity provider that handles IdM and initial authentication of the users on behalf of the SPs within the CoT.

Figure 4 shows one Identity Provider (IdP) executing the IdP functions on behalf of the Service Providers (SPs), which may all belong to different domains. All users who can access the different SPs will after logon to the IdP be equipped with SSO (Liberty) tokens to later perform transparent logon to the different SPs in the CoT.

The more complex level encompasses several circles of trust and corresponding IdPs that are interconnected through federation.

There are no sovereign IdP in this model, but they operate through a federated policy or trust agreement, whereby whoever is authenticated by the other IdP shall be accepted by other IdPs and SP in the interconnected circles. Since SAML may convey authentication method information, it is possible to build policy models that can support multi-level authentication. This means that an SSO logon attempt may be accepted only if the initial authentication complies with certain requirements. Thus, it is possible to build circles where the elements consist of different levels and methods with regard to both registration and initial authentication. If a user presents an SSO Liberty token that is too weak for a certain sub-service according to the SP's policy, multi-tiered authentica-

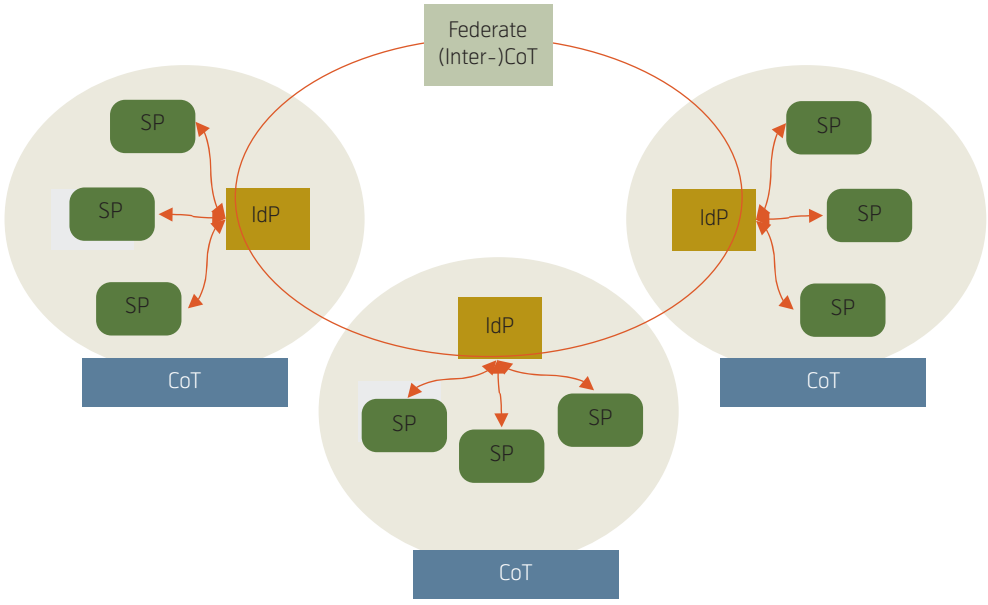


Figure 6 Federated Circle of Trust

5) *http://www.projectliberty.org/liberty/membership/current_members*

ISSN 0085-7130©Telenor ASA 2007

tion may apply, whereby the user must re-authenticate with stronger means either to its local IdP, the remote IdP, or directly to the SP (for that service).

In the federated CoT architecture any of the IdPs may be home IdPs to a set of users to provide SSO towards all interconnected SPs. Depending on the data groups available for the IdPs and the federated policy, the IdPs may provide authentication services, but also provisioning of other attributes like authorization and location information. The Identifier presented to a remote SP may be an alias or pseudonym. This is sometimes useful in order to provide privacy, but also to make the IdM flexible if the user wants to have a different Id towards different SPs. The home ISP, however, is the only entity that knows the "true" identity of the user from a legal perspective due to its control of the registration. Note that by introducing multi-level authentication, a flavour of complexity is added. If also other attributes are to be provided like authorization attributes or even personal attributes like age, customer habits or health status, it may make the whole system very complex. Although technically possible, it is probably best to handle several of these aspects bilaterally between the SPs and the users.

SSO Tokens

The SSO tokens may be signed objects containing various elements for IdM. Liberty Alliance (SAML v.2) issues these as XML documents, as shown in the example in Figure 7.

Note the subsections `AuthnStatement` and `AttributeStatement` in Figure 7 that separates different IdM information from other attributes.

The SSO tokens contain life-time information, and must be renewed after a defined period. It may also contain other attributes and method/level of initial authentication. They are created by the home IdP server but stored in the user's SSO client. When a user attempts to access an SP (e.g. its web page) the SP requests its SSO token. This is transparently transported to the SP for controls, and if OK it allows access. If not capable of concluding itself (especially in the inter-circle federation perspective), it invokes its local IsP to control the received SSO token for assistance. This is because only the IsP has information about all the IsPs in the federated system.

It may even be so flexible that the user client has not established an SSO token before logon to an SP belonging to a remote IdP. In such a case the SP connects to its local IsP, which again signals the home IsP of that user to provide a convenient SSO token. The same routine applies if the SSO token declines to fulfil the needed authentication level. The home IsP then executes an extra protocol with the user resulting in a stronger SSO token. This is one example of multi-tiered authentication.

Single Sign-off

A client can have established several parallel sessions towards different SPs since initial logon. From a security point of view it is important that all these sessions are terminated when the user logs out of his client. This "clean-up" function may be provided by some SSO systems, but it requires that the system (most likely his home IsP) is aware of and keeps track of all established sessions.

Note that Liberty Alliance based on SAML v.2.0 provides Single Sign-off, while most of the competitors do not.

DRM Concepts

Digital rights management (DRM) comprises a concerted IdM system to handle the business problem that only authorized customers may receive and display/play/ reveal some multi-media contents like downloaded TV, music or even ring-tones. A typical DRM concept distinguishes between the Content Object (CO) and the Rights Object (RO). The CO is typically scrambled or encrypted, while the RO contains descrambling information or decryption keys.

To make sure that only authorized customers are able to receive and use the ROs, these objects are typically encrypted and signed. In order to open the RO, the user must have an agent that is equipped with necessary pre-installed keys and potentially digital certificates to validate and decrypt the received RO, which in the DRM case represents the IdM token.

Since DRM comprises basic identifiers, security elements and also authorization information, it belongs to complexity level 3 of IdM as depicted in Figure 2.

Systems based on set-top boxes for video-on-demand and subscribed channels are well known representatives for the DRM technology.

OMA DRM

For mobile systems like GSM, the Open Mobile Alliance (OMA) has provided specifications that cover DRM services on the mobile phone. OMA DRM Ver. 1.0 defines an architecture where the CO and RO can be managed and distributed by different DRM servers. Provided a pre-installed DRM agent in the device (or the SIM card) the mobile device can download COs and access them if a corresponding RO is received. The RO contains administrative

ISSN 0085-7130

```
<saml:Assertion
   xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion"
   xmlns:xs="http://www.w3.org/2001/XMLSchema"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   ID="b07b804c-7c29-ea16-7300-4f3d6f7928ac"
   Version="2.0"
   IssueInstant="2004-12-05T09:22:05Z">
   <saml:Issuer>https://idp.example.org/SAML2</saml:Issuer>
   <ds:Signature
     xmlns:ds="http://www.w3.org/2000/09/xmldsig#">...</ds:Signature>
   <saml:Subject>
     <saml:NameID
       Format="urn:oasis:names:tc:SAML:2.0:nameid-format:transient">
       3f7b3dcf-1674-4ecd-92c8-1544f346baf8
     </saml:NameID>
     <saml:SubjectConfirmation
       Method="urn:oasis:names:tc:SAML:2.0:cm:bearer">
       <saml:SubjectConfirmationData
         InResponseTo="aaf23196-1773-2113-474a-fe114412ab72"
         Recipient="https://sp.example.com/SAML2/SSO/POST"
         NotOnOrAfter="2004-12-05T09:27:05Z"/>
     </saml:SubjectConfirmation>
   </saml:Subject>
   <saml:Conditions
     NotBefore="2004-12-05T09:17:05Z"
     NotOnOrAfter="2004-12-05T09:27:05Z">
     <saml:AudienceRestriction>
       <saml:Audience>https://sp.example.com/SAML2</saml:Audience>
     </saml:AudienceRestriction>
   </saml:Conditions>
   <saml:AuthnStatement
     AuthnInstant="2004-12-05T09:22:00Z"
     SessionIndex="b07b804c-7c29-ea16-7300-4f3d6f7928ac">
     <saml:AuthnContext>
       <saml:AuthnContextClassRef>
         urn:oasis:names:tc:SAML:2.0:ac:classes:PasswordProtectedTransport
      </saml:AuthnContextClassRef>
     </saml:AuthnContext>
   </saml:AuthnStatement>
   <saml:AttributeStatement>
     <saml:Attribute
       xmlns:x500="urn:oasis:names:tc:SAML:2.0:profiles:attribute:X500"
       x500:Encoding="LDAP"
       NameFormat="urn:oasis:names:tc:SAML:2.0:attrname-format:uri"
       Name="urn:oid:1.3.6.1.4.1.5923.1.1.1.1"
       FriendlyName="eduPersonAffiliation">
       <saml:AttributeValue
         xsi:type="xs:string">member</saml:AttributeValue>
       <saml:AttributeValue
         xsi:type="xs:string">staff</saml:AttributeValue>
     </saml:Attribute>
   </saml:AttributeStatement>
```

Figure 7 A Liberty Alliance SSO token example

information like validity period, number of display/play times etc. The received RO and content are locked to the receiving device.

In version 2.0 a super-distribution flexibility is added. The CO can be transferred to another device and displayed there. This is achieved by

a Grouping subscriptions to comprise a set of devices which can independently request the corresponding RO for that group;

b If the RO agent resides in the SIM card (UICC), this card can also be physically transferred to another device or work station that may be more convenient for displaying or playing or whatever.

The planned OMA DRM v2.1 enhances the flexibility further. OMA DRM is a good example of the increasing role of the UICC (or (U)SIM card) as a security element for storage of security group data like keys, and execution of security functions that belong to a third party, i.e. a non-mobile operator. Payment operators like credit card companies are likewise showing increasing interest for their mobile payment applications.

ISSN 0085-7130©Telenor ASA 2007

Payment Systems – Mobile "Real Estate" and NFC

There are several payment systems available in the world today. Most of them are applicable for the micro and intermediate payment level. Payment systems are always targets for fraud because of the assets involved. IdM is needed for authentication and authorization reasons, not only to protect the payment operator, the user and the service provider's interests, but also because Regulatory Authorities acts, e.g. for money laundering transactions.

Some of the solutions include a payment buffer or *e*Purse application that can be filled up with a limited amount for later use in smaller payment situations.

It is typical that the payment providers regulate their own terms and systems. A typical application is the POS (Point of Sale) terminal for credit or debit card based payments. These are typically accredited to perform EMV (EuroPay, MasterCard and Visa) specified payment functions. Several other systems exist on the web. The new IdM challenge is to handle the growing interest of executing payment on the mobile phone – *m*Payment. One way to implement this is for the user to use plain SMS to signal transactions between his payment provider and the SP. Although supported by some security features, this method is believed to be very vulnerable. However, it is not worse than provisioning your credit card number on the web. Normally, such systems are carried out through SSL protected HTTPS sessions towards the SP.

International payment providers are paying increased interest to implementing their payment agents (like EMV) on the mobile phone, and especially on the security element – the UICC or SIM-card. The new term for the UICC is the "Real Estate", where the payment operators can rent a security domain to store their application. For contact-less operation, Near Field Communications (NFC) technology is included in the concept. NFC can operate in both power-on and power-off mode. In the latter case it operates similar to a magnetic stripe card, or RFID cards. In power-on mode, NFC operates bi-directionally, and can thus convey payment in both directions in addition to executing more secure protocols.

The IdM challenge is to manage the system that includes two different IdM concepts; the mobile operator and the payment operator.

There are several methods for this:

a The UICC cards are entered into the UICC at manufacturing time and later activated.

b The application is downloaded OTA (Over The Air) to the UICC card and then activated.
Note: OTA is a dedicated and end-to-end protected

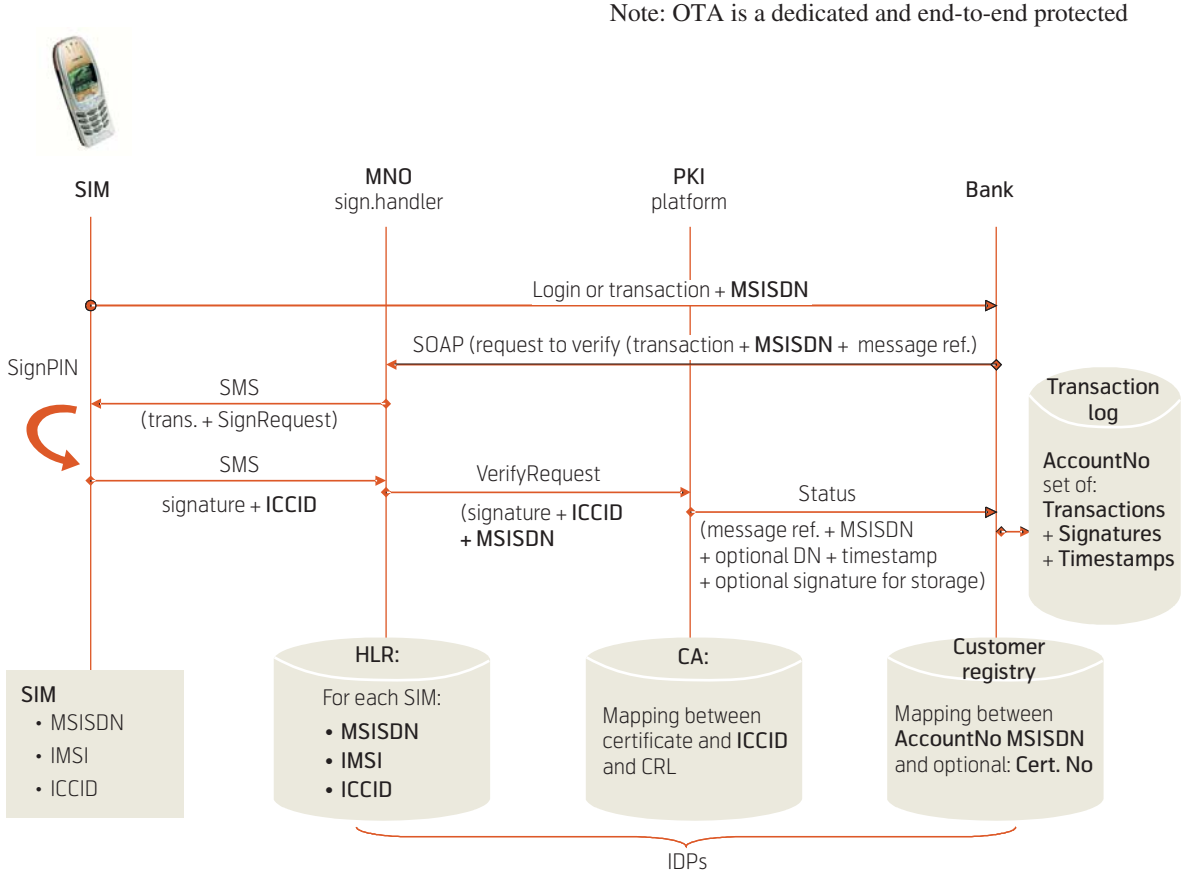


Figure 8 The IdM transactions for the BankID application (at run-time)

ISSN 0085-7130

channel for operation and maintenance of SIM-based applications, especially the SIM Application Toolkit (SAT).

c At runtime the mobile operator mediates the transactions in agreement with the payment provider through special protected interfaces. For *e*Purse systems, this method can be split into two parts where the first part is to fill up the *e*Purse, e.g. from a bank account (i.e. payment provider that is involved at this stage), and the second part is local payment where the bank is not involved, only the user and the SP.

d At runtime the transactions are handled by the user and the payment provider and transparent to the MNO.

In case *a* the IDM provider is the MNO that takes the role of a "Real Estate" provider and Real Estate Manager (activation and deactivation of an application), but not involved in handling of specific parameters.

In case *b* and *c* the MNO in addition operates as an Authentication provider. In case *d* the MNO operates as Real Estate provider. It is in the interest of the MNO to always be a Real Estate Manager, especially if an application is harmful and must be closed down or deleted.

Mobile Banking Systems – the Norwegian BankID Concept

In 2005 Telenor and Norwegian financial interest groups agreed on establishing a PKI signature function on the SIM for *m*Banking and also convey security in web/net banking.

Figure 8 indicates the complexity of IdM protocols when several IdPs with different functionality and domains cooperate.

The UICC (SIM Card) as a Security Element for IdM

The UICC smart card platform for the SIM (2G) and USIM (3G) applications is specified by ETSI SCP to handle IdM functions for GSM roaming mobile networking. It provides HW protection to its keys and security functions. The UICC is a stand-alone processing system with its own CPU, ROM, RAM (flash memory), operational system including a file system. It also has drivers to communicate with the environment, e.g. the handset. The traditional chips have 32-64 kbyte RAM; 128 kbyte RAM cards are available, and new technology can provide megabyte storage capacity chips.

The Identifiers are previously described (IMSI, MSISDN, ICCID); and if the ISIM application (for IMS) is embedded, the IMPI and IMPU(s).

The GSM security functions are the A3 authentication algorithm and the A8 function for session key derivation. The traffic encryption algorithm (A5) is handled by the hand-set and not the SIM.

In addition the UICC (U)SIM provides a channel denoted OTA for operation and maintenance executed by the mobile operator. This channel is typically 3DES end-to-end encrypted.

A typical UICC has eight (some only six) physical pins, of which five are allocated to GSM purposes. ETSI SCP decided in 2006 to use two of the three spare pins (eight pins UICC) for 12 Mb/s USB interface and has finalised the specifications. The last pin is allocated to NFC purposes together with the so-called SWP (Single-Wire Protocol) since the normal 2-wire protocol does not apply to one pin only. One single pin on the chip is the only available resource left for this purpose.

With the GSM allocated IdM elements and functions together with the new interfaces, the UICC has become a powerful tool to handle new IdM tasks including those of a third party. With a segmented operations system that isolates security domains (potentially Java cards) the picture of a Real Estate apartment building becomes clearer; third parties may store their applications in separate security domains and benefit from the various features mentioned.

Examples of UICC Real Estate Renters and IdM Aspects

1 Payment applications may use the OTA facility for maintenance and possibly establishment and the NFC interface for local payments (inbound or outbound). MNO basic GSM IdM functions are involved in initiation and maintenance, while the transactions depend on the third party's IdM system.

2 DRM agents may also be downloaded OTA and also upgraded. At runtime the content can be channelled through the 12 Mb/s USB interface for internal decryption/descrambling by the DRM agent. Although not yet possible, it may benefit from the NFC interface to hand over Rights Objects from one device to another. MNO basic GSM IdM functions are involved in initiation and maintenance, while the transactions depend on the third party's IdM system.

3 PKI agents like the BankID application on the SIM uses the OTA for initiation and conveying of public

ISSN 0085-7130©Telenor ASA 2007

keys to become inserted in certificates by using ICCID as a reference. At runtime, the signature function is executed in the SIM (under the user's PIN control) and the signature (signed SMS) is returned to the originating net-bank for verification. The ICCID is appended to the signature for reference purposes (towards the X.500 certificate repository).

Both the MNO and the net bank's IdM systems are interworking for mapping purposes; the bank controls the certificates, PKI verification functions and the bank account and of course the transaction functions. At run-time the MNO's IdM system is involved in routing the SMS challenges (to be signed) and for mapping purposes, while the bank's IdM system is involved in transaction authentication and content verification. The system is planned to be launched Q4 2007.

UICC GSM Functions for WLAN Access and SSO Services Towards IP Systems

The EU project "Fidelity" (finalized 2006) demonstrated the ability to reuse the basic GSM IDM functions and data for convergent networking and SSO.

WLAN Access Using EAP-SIM and EAP-AKA

EAP-SIM and EAP-AKA uses the A3 authentication and IMSI identifier of the MNO's IdM system to trigger a WLAN 802.1X based on a WPA/2 protected session establishment. The RADIUS server involved is collocated to an SS7 gateway to benefit from the MNO HLR and AUC services. In this case the home MNO functions as an IdP for the RADIUS server and the Internet Service Provider (ISP) that allows the WLAN user to the Internet. The IMSI must be mapped to IP-based addresses for that user in order to achieve this. The SIM functions and protocols were made available in two different ways.

1 The UICC was inserted directly into a PC where the WLAN client could access the SIM functions through special APIs provided by Axalto (later GemAlto after merging with GemPlus) that delivered the SIM/UICC functionality.

2 The SIM functions were invoked through a Bluetooth connection from the WLAN client in the PC.

SSO Using Liberty Alliance / Federated SSO

The SSO was an extension to 6.2.1, where the PC client that had been authenticated to the RADIUS server through EAP-SIM/AKA, cooperated with a coordinated Liberty server that furnished the Liberty SSO as a spin-off from the already validated status of the initially presented IMSI. This means that the PC SSO client was equipped with an SSO token used for transparent re-authentication. The benefit is that the initial authentication is denoted "strong" and 2-factor based. The Liberty token could signal this to target web pages and provide access to web sites or portals potentially with strong authentication requirements. The demonstrator yielded access with very short processing time.

The GSM UICC/(U)SIM as a Multi-access Token

Having indicated the possible multitude of access mediating that the UICC/SIM/USIM/ISIM can provide, either in the GSM context, IMS or via the IETF RADIUS or DIAMETER towards external subsystems and non-GSM networks. The associated identifiers are central in the functionality. Figure 4 shows the distribution of the concerned identifiers.

1 The SIM subscriber is announced via the A-subscriber number MSISDN to the global community through public white and yellow pages.

2 The roaming ID IMSI is used for handling the AAA issues for a roaming subscriber, when accessing a visiting access network, and consequently for accounting and billing of GSM network usage.

3 IMSI can also be used for granting access to WLAN (via EAP-SIM/EAP-AKA) IP/Internet services, and also to provide SSO, e.g. in a Liberty alliance context. In this case the HLR operates as an IdP. Since the corresponding authentication level is "strong" (2-factor/ cryptographic protocols), this may also be applied for granting access to high levels of multilevel security portals.

4 In Norway the SIM subscriber is checked against the public social security registry, and because of this association the SIM identifiers may be used to access governmental systems, e.g. through governmental *e*Portals.

5 The BankID (in progress) uses the ICCID to map against the PKI certificates operated by the Norwegian banking systems. This provides access to the user's bank account and financial transactions of different kinds. Since PKI also can target services outside the banking system any PKI-intrinsic subsystem can be targeted. In this case the IdM is provisioned by cooperation between the MNO and the banks, but where the banking system operates the PKI part including certificate validations. This may also be used towards governmental systems, e.g. to reach security portal levels with PKI requirements including Qualified Certificates as defined by ETSI[6)].

ISSN 0085-7130

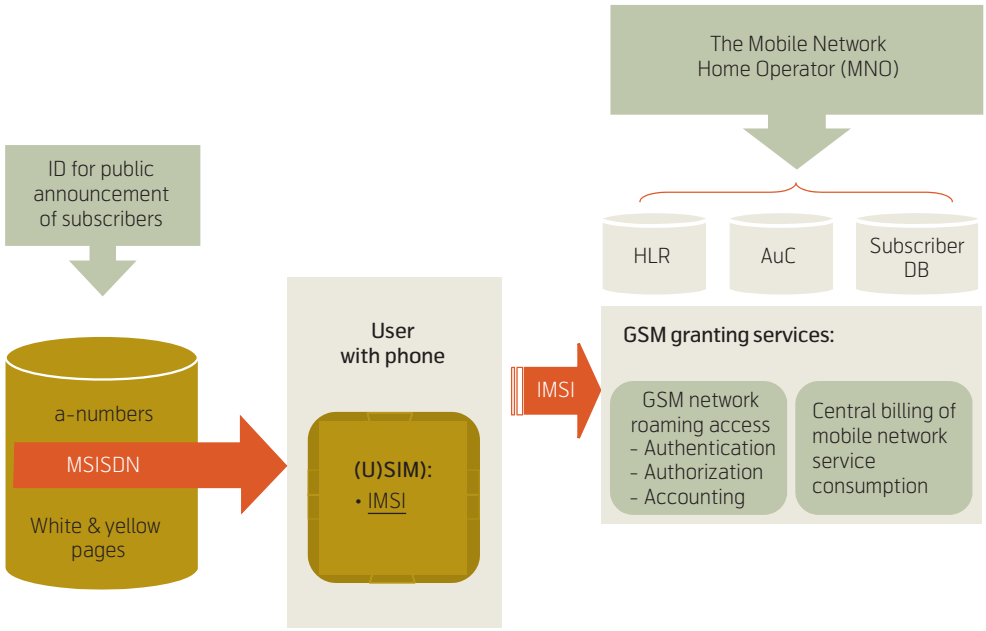


Figure 9 GSM Primary Service access – Identity usages and mappings

6 If the UICC carries the ISIM application that can be a subset of the USIM application, IMS-based IP services can be mediated over the Internet, e.g. VoIP. The HSS enhanced HLR provides the needed IMS enhancement IdP functions in this case. With regard to the ISIM identifiers, the unique IMPI can be mapped to several IMPU addresses in order to conceal the true identity of the user. This type of aliasing is handled by the IMS kernel and represents a parallel to the IMSI – TIMSI mapping in GSM visited locations.

Figure 9 shows the correspondences, usages and distribution of identifiers relevant for GSM primary service accesses. *In order to simplify, the visited network is not shown.*

Figure 10 depicts the secondary usage of the identifiers, i.e. for accessing Internet, IEEE 802.11 Wi-Fi or 802.16 WiMAX access points, Internet IMS-based services and also Liberty Alliance Single Sign-On to web-based service providers. The complexity is reduced, so only the main principles are depicted.

In Figure 10 IMSI is used for granting WLAN access. This can be achieved through the IETF EAP-SIM or EAP-AKA protocols that execute MNO/AuC-based authentication towards IEEE 802.1X/RADIUS based protocols, and where the HLR cooperates with the RADIUS server and an Internet Provider with the IdP role. The same can be achieved with both IEEE 802.11 (Wi-Fi) and 802.16 (WiMAX) access systems since both support RADIUS EAP methods.

The IMSI can also be used as an entry to Liberty Alliance SSO services. In this case the MNO HLR/AuC represents the home IdP of the user, and participates in a Circle-Of-Trust of federated IdP systems: The user Liberty Alliance client authenticates to the MNO/AuC and receives a Liberty authentication cookie for later SSO usage towards web SPs associated with other IdP participating in the same Circle-Of-Trust as the MNO.

If the user wants to access IMS-based services like VoIP, he may provide his IMPI to the IMS subsystem and potentially be mapped to a secondary IMPU identifier and authenticated by the HSS before granted services from the VoIP or other IMS-based service providers.

Telenor has since 2001 provided SIM cards with SIM Toolkit-based PKI signature functionality. The current service is based on cooperation between Telenor and the Norwegian banks. The PKI signature function remains on the SIM while the PKI signature validation functions as well as the certificate catalogues are managed by the banks. The solution represents a "Real Estate" model, where the banks "rent" resident PKI functions on the SIM card. Telenor is involved in the establishment of the PKI keys (on-board generated) and use of the OTA channel for transferring of the public key for certification in the banking domain. ICCID is used as a reference to the certificate in the X.500 catalogue, and the certificate maps to the user's bank account.

6) *ETSI TS 101 456*

ISSN 0085-7130©Telenor ASA 2007

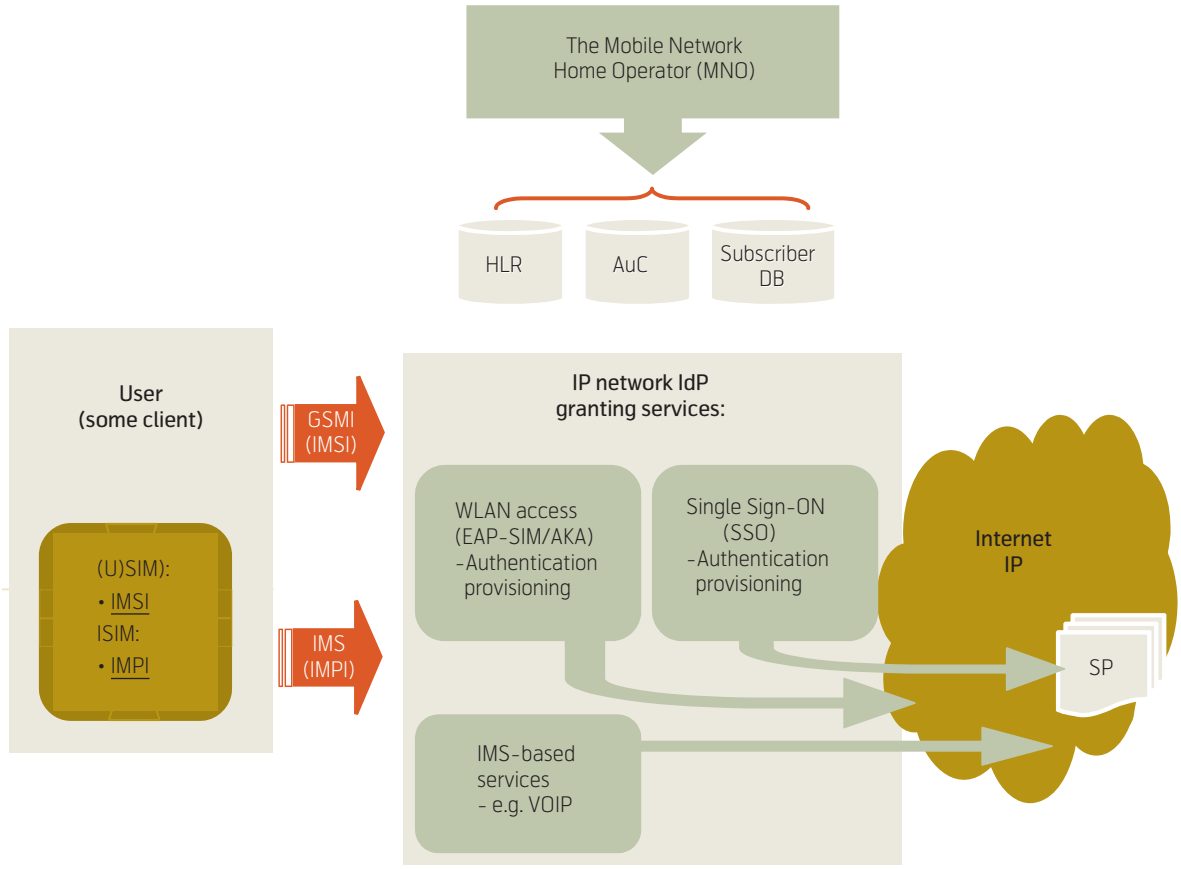


Figure 10 Secondary network and service access – Identity usages and mappings

The signature is executed over an SMS that is pre-formatted and pushed to the user. The user controls the signature by entering a dedicated PIN. The ICCID is appended to the signature and returned to the PKI validation function. IMSI and MSISDN are used only for general GSM network access and routing of the SMS. IMSI is not revealed to the banks, since the PKI functionality is disjoint to GSM roaming. The whole trust is based on specific keys to be under the user control on a specific UICC. The bank must trust MNO in its provisioning of uniqueness of the Identifiers associated with the UICC and the correct mapping. However, if a wrong ICCID is appended, the validation function will fail to find the corresponding certificate and reject the signature.

This model shows that the UICC basic identity ICCID under MNO control binds to the certificate and bank account under the user and the bank's control. It is a relevant example of how two independent IdM systems of different business domains can be connected. It is anticipated to experience more of similar cooperation in the future, i.e. mapping

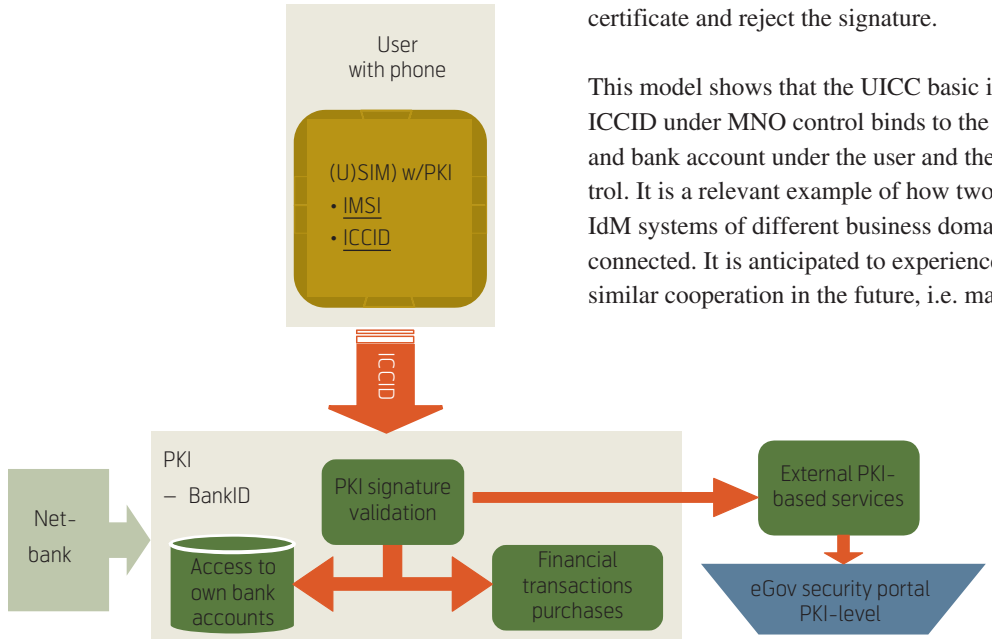


Figure 11 The SIM PKI-based BankID prospect

ISSN 0085-7130©Telenor ASA 2007

between an MNO and third party service providers. It is also important to observe the segregation of liability aspects: banking transactions at the macro payment level that fail can result in large money losses. It is possible to establish cooperative IdM concepts of this kind with clear predefined business agreements that include liability allocation.

The Real Estate Aspect of the UICC/SIM – GlobalPlatform

There is an increasing interest from third party service providers to have their agents on the mobile – and for those with security requirements, on the UICC, that is often denoted "the security element".

The payment providers like VISA have for some years now been working out specifications for smart cards. Also, since the UICC is a smart card – the payment providers have moved their attention towards the UICC.

One of the consortiums in the lead here is the GlobalPlatform[7)] where VISA, MasterCard, SIM-card vendors like Gemalto, Giesecke & Devrient and Oberthur, IBM and a few telcos are kernel members. Although an on-board payment application was the initial idea, GlobalPlatform has a more generalized approach in a sense that the UICC shall house a set of independent MNO and/or third party applications within independent security elements. Figure 13 applies. The figure is taken from GlobalPlatform's white paper on the GPD/STIP (GlobalPlatform Device/Small Terminal Interoperable Platform) solution for Mobile Security. It indicates the real estate aspect of the UICC as a compartmented building with separate apartments, some of which can be rented by third party applications. The technology is based on JavaCard Virtual machine combined with multithread functionality for concurrent operations.

In addition to the STIP architecture, GlobalPlatform has also specified secured channels between the SP and their trusted applications. This relies on the use of the OTA channels that are normally controlled by the MNO. Several models apply depending on the trust relationship between the SP and the MNO.

A simple model is to share the confidentiality with the MNO; i.e. let all information flow unencrypted through the MNO's OTA management system. The more complex model is where the SP only depends on the MNO and OTA for intitiation and general management (the MNO may grant establishment and also block/erase a trusted application on the UICC) but where the MNO has access to the internal application and information flow at run-time. In addition to the OTA usage, GlobalPlatform has also paid interest in contact-less interfaces like NFC and is fairly happy with the ongoing ETSI SCP specifications on NFC and single-wire protocol (SWP) ETSI TS 102 613 (in progress).

Although not obvious, the GlobalPlatform perspectives of the MNO-owned and controlled UICC/SIM-cards have impact on the Identity Management concept itself.

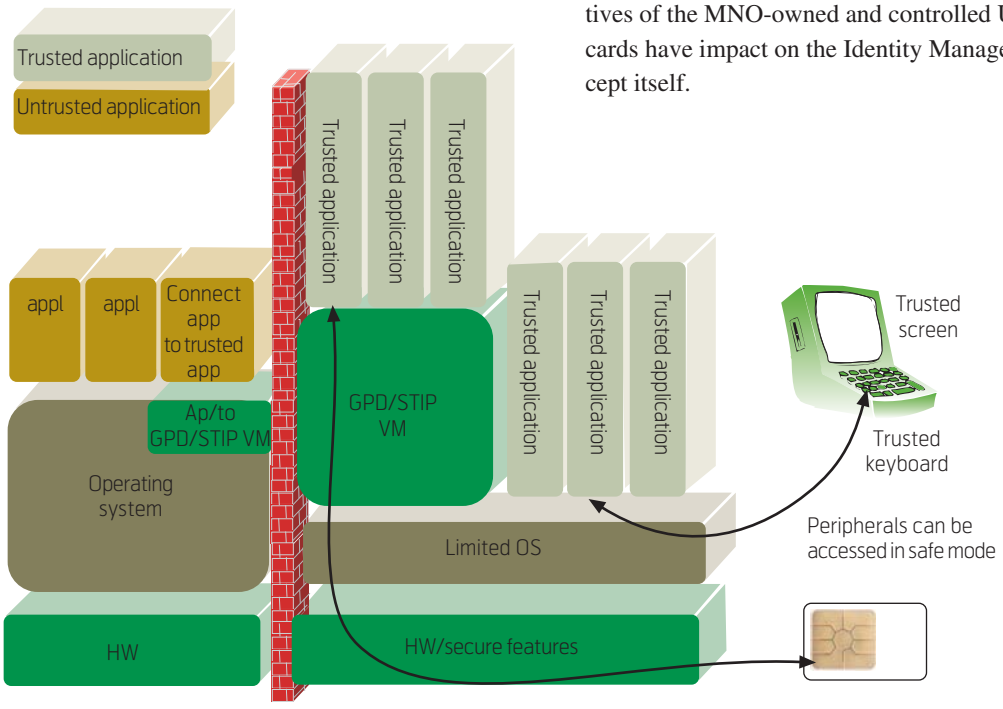


Figure 12 Global Platform architecture for separate secure execution environment

7) *www.globalplatform.org*

ISSN 0085-7130©Telenor ASA 2007

- IdM takes the new roles of Real Estate provider and Real Estate manager. That is, the management of the SIM-card itself becomes an IdM issue, not only the management of the identifiers therein. The UICC is pictured as an apartment building with locations for rent. Typical residents will be third party application agents.

- MNO Real Estate management furnishes the UICC third party compartment by allocating space for a new security element and admitting the SPs (according to business agreements) to establish their agents on the specific UICC and to use the OTA for initialization. Later the MNO may allow for the SP to update their agents, but also forcibly to revoke the compartment on its own behalf when needed. MNO IdM may use its own ID controlled identifiers like MSISDN, IMSI or ICCID for mapping purposes.

- Internal attributes within the secured apartments are under the governance of the corresponding SP and user. This means that any SP specific IdM issues are transparent to the MNO.

- If the SP has confidentiality requirements for its agent, then the end-to-end encrypted channels must be provided. The preferred solution is that the applicable keys are not shared with the Real Estate manager, i.e. the MNO. Several models are suggested according to GlobalPlatform. The most sophisticated concepts rely on on-board key generation facilities for session keys. The models depend, however, on the MNO's OTA functionality for initialization and service. Because of this the management of OTA keys becomes a part of Real Estate Management which again is a part of IdM. The taxonomy of IdM is indeed growing in complexity.

Potential Third Party Real Estate Renters

The architecture of GlobalPlatform is established with payment applications in mind targeting to have the bank or credit card on the mobile/SIM-card. The architecture and functionality is, however, generic in a sense that it can be translated into other types of applications with security needs. In addition to payment applications it is possible to foresee

- DRM agents for control of protected multimedia contents. With the recently ETSI SCP-adopted 12 Mb/s USB interface, it is possible to picture the UICC as a "set-top box" for descrambling/decrypting downloaded TV or music; e.g. OMA DRM v2.0[8)] that specifies DRM on the mobile/UICC.

- Ticket agents. A purchased ticket for logical or physical access can be downloaded by SMS to the ticket agent and stored. The ticket can later be provided locally, e.g. via the NFC interface whether it

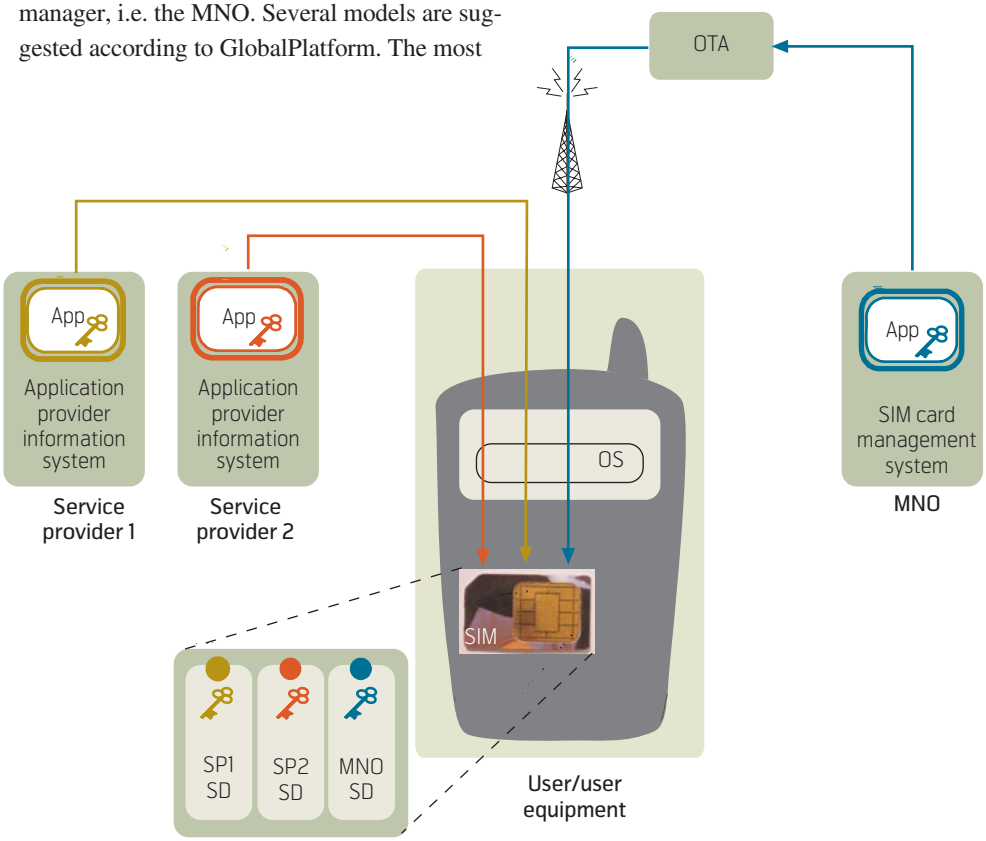


Figure 13 The Real Estate Approach of the UICC (SIM card)

8) *http://www.openmobilealliance.org/release_program/drm_v2_0.html*

ISSN 0085-7130©Telenor ASA 2007

is for transport (similar to the UK Oyster cards) or a cinema or for opening a door. Note that tickets belong to the authorization class of identifiers, ref level 3 in Figure 2.

- Health record agent (personal sensitive data). The UICC is used for storing medical health information and access codes for centrally stored large data amounts.

- Biometric passport (?) to be produced to the passport control equipment via the NFC for validation. Biometric templets yielding the essential elements in a biometric passport may in the future be stored on the SIM. The PKI-signed and encrypted templet is produced to the passport control officer.

- VPN agent for mobile office.

- Any application with security domain requirements.

In the future the UICC may indeed become a common carrier for various security intensive applications and data. The new perspective is that also third party applications are introduced. Note that the BankID concept currently under deployment in Telenor Nordic UICC/SIM cards is one example of introducing third party applications into an MNO's Real Estate management. BankID is however not based on the GlobalPlatform/Javacard approach. BankID depends on OTA functionality for initialisation and re-keying, so conceptual overlaps exist in principle. Figure 14 depicts the Real Estate approach with separate security domains (SD) and dedicated encrypted service channels, including the OTA. The figure depicts a UICC with two separate Service Providers Security Domains (SD) in addition to the MNO SD. The IdM is separated so each SP controls its own application-related internal ID aspects, while the MNO (being the landlord) controls the general management. It is necessary to introduce a new identity element in the MNO's controls in order to address each SD within the UICC. The MNO must also have the information of which SP that corresponds to an SD in a particular UICC. This means that the basic UICC identifiers (ICCID/IMSI/ MSISDN) that are controlled by the MNO/HLR must be extended with SD identifiers, which again are also shared with the corresponding SP. Since the UICC basic identifiers concern user privacy aspects, the issues are GSM policy sensitive for several reasons, and are currently under discussion in GSMA. No conclusions are made so far.

However, when the framework is finalized, the UICC will indeed play a new role, not only in basic GSM services, but in many business areas as stated before.

For the MNO the Real Estate business perspective addresses completely new revenues in the form of rent-of-SD-space income. It is similar to a bank renting out bank boxes in a physically protected area. Not all Real Estate SD services generate GSM traffic however; e.g. transactions over the NFC interface are transparent to GSM accounting mechanisms and hence cannot be billed by the MNO.

Conclusions

Identity Management has received large attention in the global arena of communications. It concerns technology, security, business, organizations, government, and furthermore, legal perspectives such as privacy and other regulatory authority items. It is shown in this article that IdM may include the handling data of several kinds, ranging from basic identifiers, security attributes, tickets, location and personal characterizers, data that can be segregated in different groups. The grouping reflects the different functions and roles of IdM.

While the basic role of IdM is the basic Identity Provisioning role to create, update, distribute, block and revoke identity data, new roles in distributed systems comprise the provisioning of Security, Authentication, Authorization and miscellaneous functions associated with attributes like Location. This gives rise to two challenges; firstly, that it is hard to give an exact definition of IdM, and secondly, the challenge to maintain consistency in the data when the already enormous and distributed IdM databases are continuously changed and updated also to reflect the mobility of the users and the entities IdM covers. The largest and most complex IdM systems of today probably belong to the communications operators in the telecom sector. Especially, the IdM system for GSM was designed for scalability and roaming. GSM operators that partly rely on the SIM-card (UICC) to convey and also protect their roaming systems have achieved a tremendous success. It is shown that the MNO can also play an IdM role outside the traditional 2G and 3G systems, e.g. for achieving IMS services, WLAN access and also SSO among web SPs on the Internet.

The (U)SIM-card has also attracted third party applications as on-board "real-estate" renters. This increases the complexity of the mobile operators' IdM to also comprise Real Estate Management and Real Estate Provisioning. Real estate aspects open

ISSN 0085-7130©Telenor ASA 2007

new strategic business areas like payment and ticket handling for the telcos and also new security and regulatory issues. It is likely that the SIM-card will be an important business enabling element for converging different networks together as well as for being a common carrier of independent applications belonging to different third party business domains. The success depends however on several aspects like the maturity, availability and cost of technology. Furthermore, it depends on the progress of standards, the timing according to available hand-sets and not least on the whole ecosystem that is needed for it to take flight. Three billion SIM-based subscribers today indicate that a massive momentum already exists.

Tor Hjalmar Johannessen is Senior Adviser in Telenor R&I. He graduated from the University of Oslo in 1975 as Cand.Real. After working with military crypto systems at Alcatel Telecom since 1989, he joined Telenor R&I Security Group in 2000. His main interest and occupation has been security in general and deployment of PKI systems, especially in the SIM-card and mobile systems. His activities comprise participation in various EU projects, IETF, EURESCOM, CEN/ISSSWS, GSMA, ETSI SCP and ETSI ESI.

email: Tor-Hjalmar.Johannessen@telenor.com

ISSN 0085-7130

Next Generation of Digital Identity

FULUP AR FOLL, JASON BARAGRY

Fulup Ar Foll is Master Architect in the global software practice of Sun Microsystems, Inc.

Jason Baragry is Lead Architect in SOA/Business Integration in Sun Microsystems, Inc.

While traditional paper based identity and digital identity share the same fundamentals and aim towards similar goals, they nevertheless have some significant differences. This paper provides a high level introduction to the fundamentals of Identity – its technical constituents, its requirements, and its risks. It closes with an explanation of how Project Liberty provides solutions to many of the issues raised by next generation digital identities.

The concept of "Identity" and having to prove your identity is nothing new. A national ID card was introduced in France around 1920, and while we had to wait until 1940 to see its generalisation, it became mandatory for every citizen older than 16 only around 2000. Obviously comparing traditional and electronic identification is a risky business, nevertheless there are some invariants that remain valid whether your ID is paper based or digitally based.

Most identity documents include more information than your basic identification. The first set of attributes is usually information allowing others to determine that you are effectively the person you claim to be – whether this information is a picture, a login/password, a certificate or any other credentials – the final goal is to authenticate you. The second set of information is usually used to apply authorization, e.g. you can drive but only with glasses, you're a French citizen which allows you to move freely within the Schengen area, your name is highlighted in red and thus you have access to confidential documents, etc. The last set of information is dedicated to controlling the validity of the document in the digital world. This will typically be an electronic signature attached to some revocation list. In the paper world, this would typically be the place and date of issuance plus some serial numbers.

Quite surprisingly, the main target of most identity documents is not to authenticate you, but to authorize you. The general mindset is to first perform authentication; this is mostly because traditionally, in order to access needed personal attributes for a given authorization, you first need to answer the question "who is he?" This is not because attributes necessary for the authorization would not be valid outside of an authentication context, but more because the information flow is such that without first doing an authentication you cannot find revealing attributes.

In order to authorize you someone must be in position of controlling the validity of your claim. While information attached to most identity documents is not very visible and thus not very accessible for most users, the way you control the information forces a direct interaction between the user and the controller, thus transforming this operation in something very accessible. The perception of what is acceptable or not depends on the context, for instance, young people versus old, Europe versus North America, etc. English people have trouble accepting any identity document and most drivers will find it perfectly normal not to have any ID while driving in their neighbourhood. On the other hand they find it perfectly acceptable to have video cameras on each street corner or to apply DNA tests to immigrants. In France, the DNA test is unacceptable and was recently refused by the chamber of senators, video cameras are never welcome, but on the other hand most people find it normal for the police to ask for an ID card whenever they want. Furthermore, if you cross the Atlantic, you have to provide your fingerprints to enter the country, show your ID card to buy alcohol, etc. Therefore, identity is something personal that may have ramifications for your private life. Depending on culture and history, some control or cross connection can be either very natural of completely unacceptable. This issue is very often outside of any serious requirement or risk analysis.

Until now identity control was mostly a paper and manual process. For this reason collusion between authorization and authentication was not a serious issue. For instance, for most of us, voting is the only time where we expect real anonymity; we do not mind our doctor knowing who we are, or our car insurance company knowing our home address – but what we don't want is our insurance to know what our doctor knows. In fact, most of us have only very few secrets that we don't want people to know about. Nevertheless, in order to protect ourselves, we want to make sure that personal information about us will not be cross-border from one sector to another. Until very recently those exchanges were only done manually. They where slow, complex, and could not happen on a large scale. Moving from manual to computer-assisted processes completely changed this paradigm. Technology has moved very fast in the past year and what was impossible a few years ago is

ISSN 0085-7130

now something a 1000 Euro computer can do. In fact, when looking back we realise that until very recently limits were more due to technology than to legal regulation, and without doubt, had DNA been available during World War II it would have been used. Today technological limits are pushed further and further every single day. It looks more like a revolution than a simple evolution and both mentality and legal frameworks have a lot of difficulty in keeping up. We are moving from a long period where we were limited by "what is possible" to a new period where we want to limit by "what is acceptable".

The big difficulty with what is acceptable is that, depending on whom you ask the question, you get a different answer. And obviously whatever technologists, politicians, etc may believe, some people will always break legal limits, whether officially or not. Furthermore, the impact of bypassing some of those limits is not visible within a normal human understandable time frame. For instance, many teenagers write very personal things on their blog without realizing that in twenty years from now, when they apply for a new job or a political investiture, that information will almost certainly re-surface. People within rich democracies tend to ignore that our world may still change for better or for worse. They are often ready to sell their privacy for only a few per cent extra discount, allowing airlines, banks, supermarkets, etc to know everything about what they eat, the clothes they like, and the music they listen to. Last but not least, what is acceptable within the current context (i.e. European Nordic countries that use a unique national ID to index almost every single identity record a citizen owns) might not be such a good idea if one day the bad guys take ownership of the country. Implementing privacy aware mechanisms increases complexity, and it is very had to justify when your immediate context does not allow you to contemplate the consequences of not doing so.

Regardless of what technologists may have you believe, today the strongest limit to digital identity is neither technological cost nor availability. In fact, the necessary technology already exists in the mass market, and as of today almost no single commercial company could exist without some form of electronic identity transaction – the cost of ignoring this technology would be so high that it would bankrupt anyone trying this path. Governments do not have the exact same financial constraints that the private sector has; nevertheless they cannot ignore technology anymore and will have to leverage it in order to both reduce cost and increase the quality of services to citizens. In fact the only remaining constraint is complexity, and as of today this is the strongest limitation of digital ID penetration. While digital ID may leverage the same general concepts as traditional identity, it remains very abstract. Indeed, even if basic daily usage is understandable for a significant part of the educated population, as soon something goes wrong, you would very quickly feel like everything is getting out of control. Most of you have already had a refused credit card, usually in an unfriendly environment: for instance a foreign country, the middle of the night, during a train strike, etc. And very rarely can someone explain what actually happened. In fact, most people understand digital ID and electronic transactions like they understand a TV remote control – when it does not work they simply want to change the batteries and if the problem persists they will simply change TV!

Implementing privacy has never been simple, and even in a traditional "paper based" world privacy is most of the time required due to loss of information and not the fact that information is not created. Let's take an example: imagine that you want to buy a book without anyone knowing that you bought it. If you buy it in a shop, even paying with cash, the salesman knows what you have bought. If you ask someone to buy it for you, then that person knows what you bought. In order to break the information chain, you would need something like the following:

- Take an envelope, place a piece of paper into it with the name of the book you are looking for and enough cash.
- Ask someone you trust to go to the shop for you with the envelope.
- The salesman opens the envelope, reads what you want, and takes the money.
- The salesman puts your book in the envelope, closes it, and hands it back to your trusted courier.
- Your partner sends back the envelope to you.

In fact most of the time this is not necessary, and if you buy a book far enough from your home and in a busy place like a central railway station, it might be more than enough for the information to get lost. Nevertheless this shows that in a traditional world privacy is mostly due to loss of information and not the fact that information is not created. This situation has had a significant consequence, it has allowed the mass population to have an acceptable level of privacy except in specific cases, e.g. criminal investigations, which permit the implementation of special processes to ensure that the loss of information is avoided. Obviously the fact that the information is created has limits, and some governments like the

ISSN 0085-7130

old East German one have proved that collecting and manually cross connecting all those small pieces of information was possible.

One of the strongest differences between paper and digital ID is that in the digital world it is very easy to automate processes, and what was once only possible for huge organizations like the Stasi in the recent past, is now possible for almost any student in a garage. Today technology provides almost unlimited capabilities for both storage and processing, making research of correlation between small pieces of information far too simple to guarantee privacy. It is a real risk that regardless of whatever legal frameworks are implemented to protect citizens, that privacy as we know it today, will just disappear. It is obvious that in only a few years from now, we will have the technological capability to determine from our mobile phone camera who everyone is, just by taking a picture and asking the question to whatever will be the next generation Google search engine. The Government cannot prevent rain; it is a good question to ask ourselves if they can limit technological capabilities and, if they can, should they?

Knowing that the cross connection of information will be possible, easy, and unlimited, the only option to protect privacy is not to create the information in the first place. In fact, the next generation of digital ID architecture should do more to limit the creation of information than anything else. Limiting the creation of information is hard but not impossible unless you own the system. It is the only way to guarantee that if the ownership of the system does change, then even if the new owner changes all the rules, they will not get the information they want. This is because it will not exist. We should not forget that the only information no-one can steal is that which does not exist.

As we've seen before, identity information can be spliced in three classes:

- Indexes to search and find the information;
- Attributes containing part of the information about the user;
- Control mechanisms to guarantee the authenticity of the information.

In order to limit the risk of correlation, the first thing to do is to decouple the indexes that connect the principal from the attributes that define himself. If we take as an example your medical records, this is a very private document and in many cases you would not want your employer, nor your insurance company, nor your wife/husband, etc to see it. On the other hand, if we remove your name from this document, then most of us would be more than happy for all the data to be given to a university for research. This is a very good example of what endangers privacy is not the information in itself, but the connection with what we call the principal, e.g. the name and address with the data.

The second action is to make sure that we limit the information itself. The first and most obvious action is to build a process where we only provide necessary information. Most of today's processes request far too many pieces of information. Why does a driving licence require a place of birth? Why would your telephone company need to know your age? Why a hotel your home address? ... Very often when you ask people why they want this type of information they simply cannot provide an answer. In the manual world, collecting the information was very complex and it was somehow understandable that people would ask for more information than needed, just in case they would need that information at a later time. Even if we could argue this, due to the cost of manual correlation being generally unacceptable, the risk to your privacy was probably acceptable. A modern digital ID architecture should allow you to provide only the minimum level of information mandatory for a given action. Furthermore, in many cases the information should even be reduced. For instance, not providing your full date of birth, but just the fact you're an adult. While providing only required attributes for a given service is complex, it leaves the owner of the attribute in a position to choose whether to provide the information of not. Unfortunately, as soon as the attribute is given, you have no idea and no control over how or what for the receiving party will use it. Nevertheless, in most cases you can somehow trust the receiving party to handle the information as you wish, for example they claim not to store it on disk, it should be deleted after six months, etc. This obviously only applies if you have a means of communicating your wishes in a manner that the receiving party will understand. Last but not least, all of this should be done under user control, that is, not by asking a user to renounce any of his rights as is too often the case, but by providing the end-user with real control over what he wants and what he does not want. In fact, a new generation of digital ID architecture would require:

- Separation of identity principal from attributes. Needs a mechanism to get attributes about someone you don't know (i.e. anonymous access to the "adult" attribute only);

ISSN 0085-7130©Telenor ASA 2007

- Detach authentication from authorization and attributes exchange. By coupling all those pieces of information together it creates a new big brother.

- Provide mechanisms to prove an attribute is authentic. Many services create copies of information only because they have no way of trusting other parties.

- Allow users to control the overall system, within an adequate level of complexity.

  - Build a distributed system to both make it scalable to country/continent level and to guarantee that if an element is compromised only a limited number of information pieces/users will be affected.

- A way to handle exceptions, tracking of bad behaviour, terrorism prevention, etc.

Last but not least, what about risk and security? The world is not perfect, no system is fully secure, the world is not as gentle as we would like it to be. Identity attributes are critical pieces of information for many actors: government, intelligence, commercial companies, etc. While the number of risk factors is probably unlimited, the following ones seem to be the biggest.

External attack: Quite surprisingly this is probably the easiest one to deal with. Mail spam and virus are probably the best example of external attack. They create useless trouble and cost money, but at the end of the day they do not kill us. I tend to consider external attack as street graffiti; they create useless damage, without comprising national security. In fact the biggest issue with external attack is that this is the most visible one. It is also the most understood one by IT people and for this reason we tend to allocate too much resources to it, forgetting the other ones.

Internal risk: In my opinion the biggest one. This for the simple reason that in order to make operations possible you need at least some of the employees to get access to some of your critical information. There are two classes of risk. First, human error, where something is leaked unintentionally. Second, criminal, where some bad guy finds a cheaper and faster way to buy someone to get the information from the inside rather than trying to get it by hacking your system from the outside.

Change factor: One risk that today is hardly ever addressed. While we may hope that modern democracies will never be replaced by the bad guys, should it not be taken into consideration? More commonly, when a commercial company changes ownership, how can a given user prevent his personal information being transferred without his consent?

Project Liberty is a technology framework that has been designed to address this class of problems. It might not be perfect, but it is a solution that has been proven to work. Digging deep inside Liberty or SAML technologies is out of scope for this paper, but hopefully even while keeping a high level of abstraction we may explain how it addresses some of these key issues.

Federation: Technically implemented through SAML2 protocols. It is a weak link between different identities, or a given principal. Federation allows us to keep the level of complexity simple enough for end-users to browse seamlessly from one service to another without re-authenticating. Federation is the cornerstone that allows us to decouple the identity of the end-user (principal) from its attributes, allowing a given principal identity to be slotted in many different places in a transparent manner for the end-user.

Identity attributes: Implemented through IDWSF. We have seen that it is very important to separate attributes from the principal identity. Liberty Identity Web Services allow us to discover, request, and retrieve updates of attributes about a given user without needing access to the principal identity. It is designed in such a way that different users may choose different services to hold a given attribute, or even to allow a given user to store a given attribute in two different places with eventually two different values.

Social network: Implemented through people services. There are many cases where you need to group identities, this is either to give special access right to a given group (e.g. allow all parents from a given class to see pictures from a school sport event) or to allow someone to act on behalf of someone else (e.g. an accountant acting on behalf of a company).

User consent: Implemented through Liberty Interaction Service. Allows a given service to request user consent independently of the level of imbrication a given request has.

Identity governance: A Liberty ongoing effort known as IGF (Identity Governance Framework). It is an XACML based layer that, first, allows a service owning an attribute to take a decision on whether or not it can release the attribute to a requesting party, and second, allows a request to be sent attached with the attributes and some metadata for the receiving parties to know how they should/can handle that informa-

ISSN 0085-7130

tion. In many government cases, the fact that you can release or not an attribute is based an a legal framework, and IGF rules should be a flat translation of legal constraints. On the other hand, the receiving party agrees to respect the constraints attached to requested information (e.g. not allow to write on disk, but be deleted after six months, ...).

Obviously technology cannot do everything. Nevertheless it should handle privacy as a first class citizen and propose a framework that handles enough of the problem automatically, thus leaving the remaining part manageable outside of technology by manual and legal mechanisms. Obviously nothing is perfect, but not doing anything because we cannot do everything would be criminal. Not only should a given user own the right to verify and change most of the information held about him, but in many cases he should also own the right to simply forget. How a digital ID will allow someone to be a very bad teenager and then years later to be a very nice respectful father, employee, etc., is only one example of what a modern digital ID framework should implement.

Project Liberty was designed to handle this global problem. Some people complain that it is too complex or too restrictive. However, we had to make it complex enough to support the complexity of our world and we chose to make it restrictive to keep the cost of adoption acceptable.

About Project Liberty

The Liberty Alliance is a global identity consortium formed in 2001 by approximately 30 organizations with the goal of developing open technical, business and privacy standards for federated identity management. Liberty Alliance achieved this goal in 2002 with the release of Liberty Federation and in 2003 released Liberty Web Services, an open framework for deploying and managing a variety of identity-enabled Web Services. Having grown to nearly 150 members from around the world, the Liberty Alliance is currently working toward developing ID-SAFE, the industry's first open framework for deploying and managing interoperable strong authentication. The Liberty Alliance is the only global identity organization approaching identity issues from a holistic perspective, addressing the technology, business and privacy aspects of identity management in order to build a more trusted global Internet for consumers and organizations worldwide. Liberty Alliance background information including a listing of timelines and industry milestones is available for download from http://www.projectliberty.org.

About Sun Microsystems

Since its inception in 1982, a singular vision – "The Network Is The Computer" – has propelled Sun Microsystems, Inc. (Nasdaq: SUNW) to its position as a leading provider of industrial strength hardware, software and services that make the Net work. Sun can be found in more than 100 countries and on the World Wide Web at http://www.sun.com.

Fulup Ar Foll holds a Master degree in Computer Science from the French Military School. Before joining Sun he was a research engineer for ten years on distributed technologies for the French Department of Defence and he taught Internet and Java technologies for six years at South Brittany University. For Sun he has been the lead Internet Architect for many projects related to European Telecom operators (France Telecom, Orange, Vodafone, Telefonica, Turkcell, T-Systems, ...), as well as for other strong identity and security infrastructure users such as banks and governments. In the recent past he helped France and Norway to move toward the Liberty Alliance Federated model. He is currently Master Architect inside Sun global software practice and focuses on high scale federated identity issues. He represents Sun software customer service group inside Liberty Technology Expert Group standardization committee, and inside OMA (Open Mobile Alliance) MWS (Mobile Web Services) group and works as Lead Architect for major identity projects on a world-wide level. He has also been speaker at many international conferences.

email: fulup@sun.com

Jason Baragry holds a PhD in Software Engineering from La Trobe University, Australia. He is based in Norway and is the Lead Architect in SOA / Business Integration for Sun in Central and Northern Europe. He has been an Architecture Advisor for many government and telco projects both in Norway and in the wider Nordic area.

email: Jason.Baragry@sun.com

ISSN 0085-7130©Telenor ASA 2007

Microsoft Windows CardSpace and the Identity Metasystem

OLE TOM SEIERSTAD

Ole Tom Seierstad is Chief Security Advisor in Microsoft Norway

Many of the problems on the Internet today, from phishing attacks to inconsistent user experiences, come from the patchwork nature of digital identity solutions that software makers have built in the absence of a unifying and architected system of digital identity. An identity metasystem, as defined by the Laws of Identity, supplies a unifying fabric of digital identity, uses existing and future identity systems, provides interoperability between them, and enables the creation of a consistent and straightforward user interface to them all. Basing our efforts on the Laws of Identity, Microsoft is working with others in the industry to build the identity metasystem using published WS-* protocols that render Microsoft's implementations fully interoperable with those produced by others.

CardSpace is Microsoft's implementation of an Identity Metasystem that enables users to choose from a portfolio of identities that belong to them and use them in contexts where they are accepted, independent of the underlying identity systems where the identities originate and are used.

Using CardSpace, many of the dangers, complications, annoyances, and uncertainties of today's online experiences can be a thing of the past. Widespread deployment of the identity metasystem has the potential to solve many of these issues, benefiting everyone and accelerating the long-term growth of connectivity by making the online world safer, more trustworthy, and easier to use.*)

The Landscape

In the past three decades, information and communications technologies have transformed the global economy and given hundreds of millions of people new ways to work, communicate, learn, shop and play.

*e*Commerce is growing, with businesses delivering more services and content across the Internet, communicating and collaborating online, and inventing new ways to connect with each other. Greater productivity, more efficient internal processes and new ways of collaborating within organizations and with partners and customers are enabling organizations of all sizes to compete more effectively.

Governments are also taking advantage of these advances to improve the efficiency of their operations and deliver public services more effectively to citizens.

Widely publicized security and data breaches and growing consumer anxiety about identity theft and the privacy of their personal information are eroding public trust in the Internet.

Opportunities and Challenges

But as the value of what people do online has increased, the Internet itself has become more complex and dangerous. Online identity theft, fraud, and privacy concerns are on the rise. And sophisticated practices such as "phishing" are more and more common. Phishing attacks use social engineering to steal consumers' personal identity data or financial account credentials. Social-engineering schemes use 'spoofed' e-mails to lead consumers to false websites designed to trick recipients into divulging personal data such as credit card numbers, account usernames, passwords and social security numbers. Figure 1 shows the number of new phishing sites during the last twelve months (http://www.antiphishing.org).

One root of some of these problems is that the Internet was designed without a system of digital identity in mind. In efforts to address this deficiency, numerous digital identity systems have been introduced, each with its own strengths and weaknesses. But no single system meets the requirements of every digital identity scenario. The reality is that many different identity systems are in use today, with still more being invented. The result is an inconsistent patchwork of improvised solutions at every website.

Open Identity Metasystem

Given that universal adoption of a single digital identity system or technology is unlikely ever to occur, a successful and widely employed identity solution for the Internet requires a different approach – one with the capability to connect existing and future identity systems into an identity metasystem (or "system of systems"). This metasystem leverages the strengths of its constituent identity systems, provides inter-

*) *The paper is based on whitepapers and blogs from Kim Cameron and Michael Jones – Microsoft Cooperation.*

ISSN 0085-7130

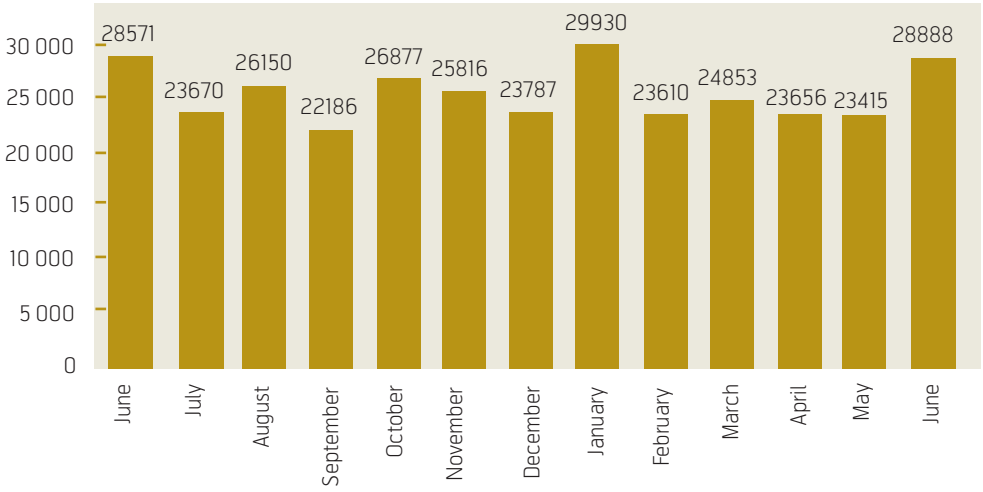


Figure 1 Numbers of new phishing sites (according to www.antiphishing.org)

operability between them, and enables creation of a consistent and straightforward user interface to all. The metasystem enables identities provided by one identity system technology to be used within systems based on different technologies, provided an intermediary exists that understands both technologies and is willing and trusted to do the needed translations.

It is important to note that the identity metasystem does not compete with or replace the identity systems it connects. Rather, it plays a role analogous to that of the Internet Protocol (IP) in the realm of networking. In the 1970s and early 1980s, before the invention of IP, distributed applications were forced to have direct knowledge of the network link, be it Ethernet, Token Ring, ArcNet, X.25, or Frame Relay. But IP changed the landscape by offering a technology-independent metasystem that insulated applications from the intricacies of individual network technologies, providing seamless interconnectivity and a platform for including not-yet-invented networks (such as 802.11 wireless) into the network metasystem.

In the same way, the goals of the identity metasystem are to connect individual identity systems, allowing seamless interoperation between them, to provide applications with a technology-independent representation of identities, and to provide a better, more consistent user experience with all of them. Far from competing with or replacing the identity of a system it connects, the metasystem relies on the individual systems to do its work.

The Identity Metasystem allows users to manage their digital identities, whether they are self-issued or issued by third-party identity providers, and employ them in contexts where they are accepted to access online services. In the Identity Metasystem, identities are represented to users as “Information Cards”.

Maintain the Diversity of Systems

In the offline world, people carry multiple forms of identification in their wallets, such as driver’s licenses or other government-issued identity cards, credit cards, and cards such as frequent flyer cards. People control which card to use and how much information to reveal in any given situation.

Identities can be in or out of context. Identities used out of context generally do not bring the desired result. For example, trying to use a coffee card to cross a border is clearly out of context. On the other hand, using a bank card at an ATM, a government-issued ID at a border, a coffee card at a coffee stand, and a Passport Network account at MSN Hotmail are all clearly in context.

In some cases, the distinction is less clear. You can use a government-issued ID at your ATM instead of a bank-issued card, but if this resulted in the government having knowledge of each financial transaction, some people would be uncomfortable. You can use a Social Security Number as a student ID number, but that facilitates identity theft. And you can use Passport accounts at some non-Microsoft sites, but few sites chose to enable this; even where it was enabled, few users did so because they felt that Microsoft’s participation in these interactions was out of context.

Similarly, the identity metasystem makes it easier for users to stay safe and in control when accessing resources on the Internet. It lets users select an identity from among a portfolio of their digital identities and use them at Internet services of their choice where they are accepted. The metasystem enables identities provided by one identity system technology to be used within systems based on different technologies, provided an intermediary exists that under-

ISSN 0085-7130©Telenor ASA 2007

stands both technologies and is willing and trusted to do the required translations.

It is important to note that the identity metasystem does not compete with or replace the identity systems it connects. Instead, the goals of the identity metasystem are to connect individual identity systems, allowing seamless interoperation between them, to provide applications with a technology-independent representation of identities, and to provide a better, more consistent user experience with all of them. The metasystem relies on the individual systems to do its work.

Principles ("Laws of Identity")

The open identity metasystem is designed to follow a set of principles (also called "The Laws of Identity") that have been developed with ongoing feedback and input from a broad community of people active in the digital identity community.

The principles that an identity system should follow are the following.

- *User Control and Consent*
  Identity systems reveal information that identifies a user only with the user's consent.

- *Minimal Disclosure for a Constrained Time*
  The identity system solution that discloses the least amount of identifying information is the most stable, long-term solution.

- *Justifiable Parties*
  Identity systems disclose identifying information only to parties who have a necessary and justifiable place in a given identity relationship.

- *Directed Identity*
  Identity systems support both "omnidirectional" identifiers for use by public entities and "unidirectional" identifiers for use by private entities, thus facilitating discovery while preventing unnecessary release of correlation handles.

- *Pluralism of Operators and Technologies*
  Identity systems channel and enable the inner workings of multiple identity technologies run by multiple identity providers.

- *Human Integration*
  Identity systems define the human user to be a component of the distributed system, integrated through unambiguous human-machine communications mechanisms that offer protection against identity attacks.

- *Consistent Experience across Contexts*
  Identity systems facilitate negotiation between a relying party and a user of a specific identity. That presents a harmonious human and technical interface while permitting the autonomy of identity in different contexts.

For a complete description of Laws of Identity, please visit Kim Cameron's blog at http://www.identityblog.com.

CardSpace Solution

Windows CardSpace is client software that enables users to provide their digital identity to online services in a simple, secure and trusted way. It is what is known as an *identity selector:* when a user needs to authenticate to a web site or a web service, CardSpace pops up a special security-hardened UI with a set of "information cards" for the user to choose from. Each card has some identity data associated with it – though this is not actually stored in the card – that has either been given to the user by an *identity provider* such as their bank, employer or government, or created by the user themselves.

The CardSpace UI enables users to create *Personal* cards or self-*issued* cards and associate a limited set of identity data. When the user chooses a card, a signed and encrypted security token containing the required information (e.g. name and address, employer's name and address, or credit limit) is generated by the identity provider that created the card. The user, in control at all times, then decides whether to release this information to the requesting online service. If the user approves then the token is sent on to this *relying party* where the token is processed and the identity information is extracted.

CardSpace is an identity selector for Microsoft Windows. Other operating systems have their own identity selector implementations. The architecture upon which CardSpace has been built – consisting of subjects, identity providers and relying parties – is called "The Identity Metasystem". This is not just a Microsoft initiative, but rather it is the shared vision of many across the industry as to how we can solve some of the fundamental identity challenges on the Internet today.

The token is opaque as far as CardSpace is concerned. CardSpace is thus security token agnostic: it can be in any format whatsoever. However, the Identity Provider should provide a plain text version of the token – the display token – so that CardSpace can show this to the user and get the user's consent to give the token to the Relying Party. The user no longer needs a password to login.

ISSN 0085-7130

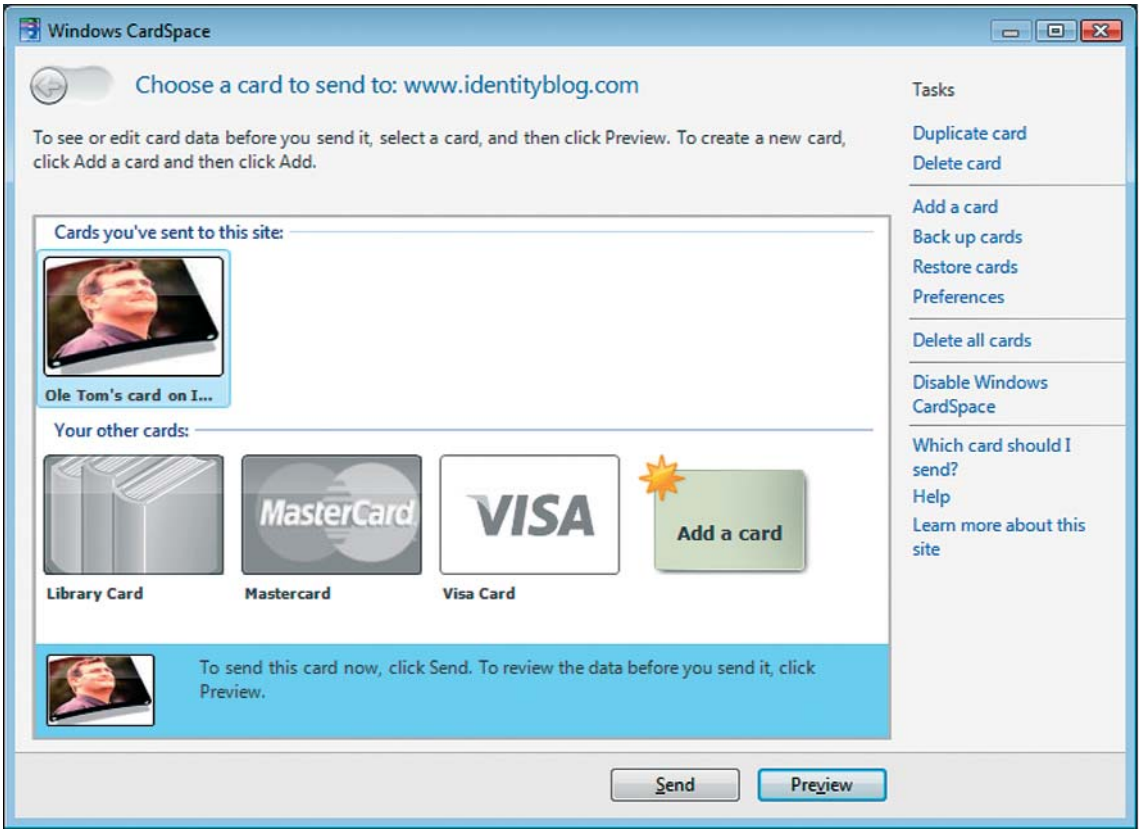


Figure 2 CardSpace solution in Windows Vista

Types of Information Cards

There are two types of Information Cards supported by CardSpace: Managed cards and Personal cards (also called self-issued cards).

Managed cards are cards that an Identity Provider has given to the user, who has imported it into Identity Selector. Identity Providers declare the claims they support in their cards using URIs. Separate Identity Providers can collaborate on the URIs they use to declare their claims, or make up ones specifically for themselves.

Personal Cards are cards that the user is also acting as the Identity Provider, and the user provides all the values for the claims. CardSpace provides the facility for the user to create, edit, export, and import Personal Cards. The data for these cards is encrypted and stored on the user's computer. The claims that a Personal Card can support are fixed, so that Relying Parties can accept a common, consistent Information Card.

Open Identity Metasystem Architecture

This section covers the general architecture of an open identity metasystem.

Roles

Different parties participate in the metasystem in different ways. The following roles within the metasystem are:

Identity Providers issue identities. For example, credit-card providers might issue identities that enable payment, businesses might issue identities to their customers, governments might issue identities to citizens, and individuals might use self-issued identities in contexts like signing on to Web sites.

Relying Parties require identities. For example, a website or online service that uses identities offered by other parties.

Subjects are the individuals and other entities about whom claims are made. Examples include end users, companies, and organizations.

Each person and entity that participate in an identity metasystem can play all the roles, and each person and entity can play more than one role at a time. Often a person or entity plays all three roles simultaneously.

Components

The metasystem is made up of five key components:

ISSN 0085-7130©Telenor ASA 2007

Claim	URI
Given Name	http://schemas.xmlsoap.org/ws/2005/05/identity/givenname
Last Name	http://schemas.xmlsoap.org/ws/2005/05/identity/surname
Street	http://schemas.xmlsoap.org/ws/2005/05/identity/streetaddress
Locality (City)	http://schemas.xmlsoap.org/ws/2005/05/identity/locality
State or Province	http://schemas.xmlsoap.org/ws/2005/05/identity/stateorprovince
Postal Code	http://schemas.xmlsoap.org/ws/2005/05/identity/postalcode
Country/Region	http://schemas.xmlsoap.org/ws/2005/05/identity/country
Phone Number	http://schemas.xmlsoap.org/ws/2005/05/identity/homephone
Other Phone	http://schemas.xmlsoap.org/ws/2005/05/identity/otherphone
Mobile Phone	http://schemas.xmlsoap.org/ws/2005/05/identity/mobilephone
Date of Birth	http://schemas.xmlsoap.org/ws/2005/05/identity/dateofbirth
Gender	http://schemas.xmlsoap.org/ws/2005/05/identity/gender
PPID	http://schemas.xmlsoap.org/ws/2005/05/identity/privatepersonalidentifier
Web Page	http://schemas.xmlsoap.org/ws/2005/05/identity/webpage

Table 1 Schemas

1 A way to represent identities using claims;

2 A means for identity providers, relying parties, and subjects to negotiate;

3 An encapsulating protocol to obtain claims and requirements;

4 A means to bridge technology and organizational boundaries using claims transformation;

5 A consistent user experience across multiple contexts, technologies, and operators.

Claims-Based Identities

Identities consist of sets of claims that are asserted about the subject of the identity. For example, the claims on a driver's license might include the issuing state, the driver's license number, a name, address, gender, birth date, the kinds of vehicles the licensee is eligible to drive, and so on. The issuing state asserts that these claims are valid.

The claims on a credit card might include the card issuer's identity, the card-holder's name, the account number, the expiration date, the validation code, and the card-holder's signature. The card issuer asserts that these claims are valid.

The claims on a self-issued identity (such as a business card) might include your name, address, and telephone number. For self-issued identities, you assert that these claims are valid yourself.

Table 1 shows the claims that are available in Personal Information Cards, along with the URIs that represent each of the claims.

Negotiation

Negotiation enables participants in the metasystem to make agreements required for them to connect with one another within the metasystem. Negotiation is used to determine mutually acceptable technologies, claims, and requirements. For instance, if one party understands SAML and X.509 claims, and another understands Kerberos and X.509 claims, the parties negotiate and decide to use X.509 claims with one another. Another type of negotiation determines whether the claims required by a relying party can be supplied by a particular identity. Both kinds of negotiation are simple matching exercises; they compare what one party can provide with what the other one requires to determine whether there is a fit.

Encapsulating Protocol

The encapsulating protocol provides a technology-neutral way to exchange claims and requirements between subjects, identity providers, and relying parties. The participants determine the content and meaning of what is exchanged, not the metasystem. For example, the encapsulating protocol would allow an application to retrieve SAML-encoded claims without having to understand or implement the SAML protocol.

ISSN 0085-7130

Claims Transformers

Claims transformers bridge organizational and technical boundaries by translating claims understood in one system into claims understood and trusted by another system, thereby insulating the mass of clients and servers from the intricacies of claim evaluation. Claims transformers may also transform or refine the semantics of claims. For example, a claim asserting, "Is an employee" might be transformed into the new claim, "OK to purchase book." The claim "Born on March 22, 1960" could be transformed into the claim "Age is over 21 years", which intentionally supplies less information. Claims transformers may also be used to change claim formats. For instance, claims made in formats such as X.509, Kerberos, SAML 1.0, SAML 2.0, SXIP, and others could be transformed into claims expressed using different technologies. Claims transformers provide the interoperability needed today, plus the flexibility required to incorporate new technologies.

Consistent User Experience

Many identity attacks succeed because the user was fooled by something presented on the screen, not because of insecure communication technologies. For example, phishing attacks occur not in the secured channel between web servers and browsers – a channel that might extend thousands of miles – but in the 70 or 80 centimeters between the browser and the human who uses it. The identity metasystem, therefore, seeks to empower users to make informed and reasonable identity decisions by enabling the development of a consistent, comprehensible, and integrated user interface for making those choices.

One key to securing the whole system is to present an easy-to-learn, predictable user interface that looks and works the same no matter which underlying identity technologies are employed. Another key is making important information obvious – for instance, displaying the identity of the site you are authenticating to in a way that makes spoofing attempts apparent. The user must be informed which items of personal information relying parties are requesting, and for what purposes. This allows users to make informed choices about whether or not to disclose this information. Finally, the user interface provides a means for the user to actively consent to the disclosure, if they agree to the conditions.

WS-* Specifications

As with other features of WCF, the CardSpace technology is built upon a set of open specifications, the WS-* Web Services Architecture. The encapsulating protocol used for claims transformation is WS-Trust. Negotiations are conducted using WS-MetadataExchange and WS-SecurityPolicy. These protocols enable building a technology-neutral identity metasystem and form the "backplane" of the identity metasystem. Like other Web services protocols, they also allow new kinds of identities and technologies to be incorporated and used as they are developed and adopted by the industry.

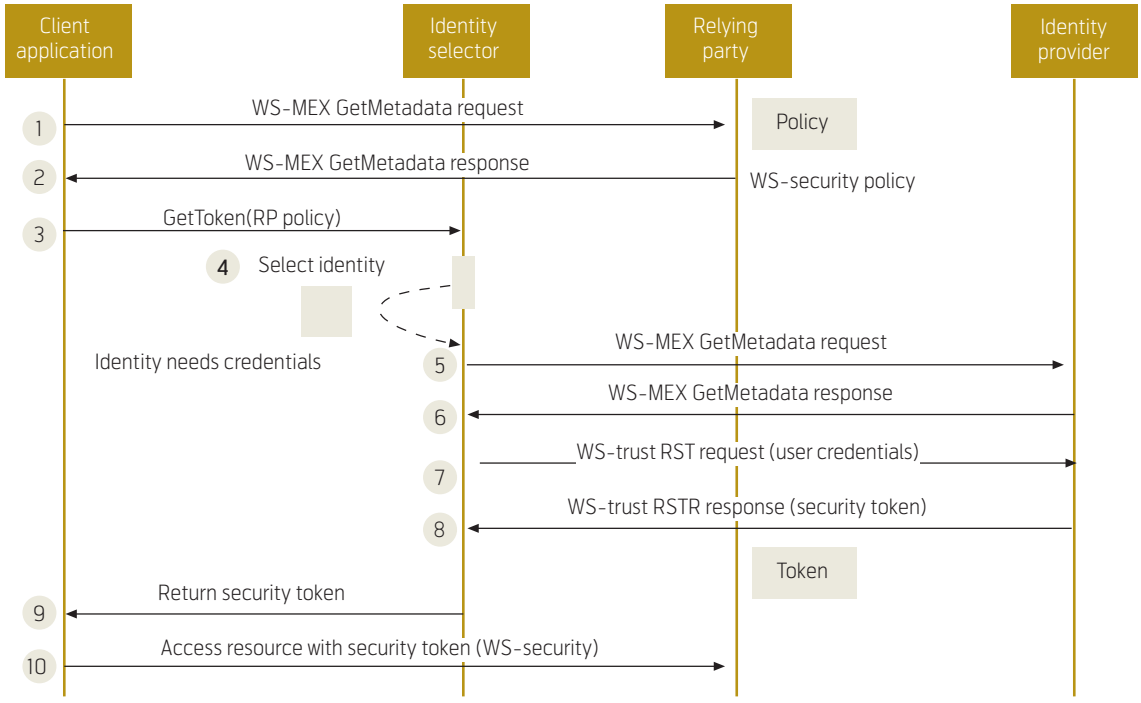


*Figure 3 Dataflow in WS**

ISSN 0085-7130©Telenor ASA 2007

To promote the interoperability necessary for broad adoption, the specifications for WS-* are published and are freely available, have been and continue to be submitted to open standards bodies, and allow implementations to be developed royalty-free. More information about Web Services Specifications is found at http://msdn2.microsoft.com/en-us/webservices/Aa740689.aspx.

Deployments of existing identity technologies can be leveraged in the metasystem by implementing support for the three WS-* protocols above. Examples of technologies that could be utilized via the metasystem include LDAP claims schemas, X.509, which is used in Smartcards; Kerberos, which is used in Active Directory and some UNIX environments; and SAML, a standard used in inter-corporate federation scenarios.

End-to-End Scenario

Figure 4 illustrates the end-to-end processes that occur when you use CardSpace to access a site that requires user validation.

The figure shows information flows through the client machine at the control of the user – this indicates that the metasystem is following law 1.

- In traditional models, identity provider and relying party are confined to the same domain.
- Federated identity allows an organization to consume identities issued by other organizations.
- A metasystem allows identity to be used flexibly and dynamically, with parties negotiating relationships.

Protocol:

1 User is asked for identity.

2 User chooses an identity provider.

3 Identity provider gives user a security token.

4 User passes the token to the requestor.

5 When the user requests a security token they have to authenticate themselves to their identity provider in some way. The IP does not give a token to just anyone who asks, you have to have the right to ask for the token. The four methods of authentication in CardSpace are X.509, Kerberos, username and password, and self-issued token. Any method that can plug in as an X.509 cert via a Crypto Service Provider will work.

6 Token is released to RP; RP reads claims and allows access.

WS-* Metasystem Architecture

Figure 5 depicts sample relationships between a subject, identity providers, and relying parties, showing some of the technologies used by the metasystem and by specific systems utilized through the metasystem.

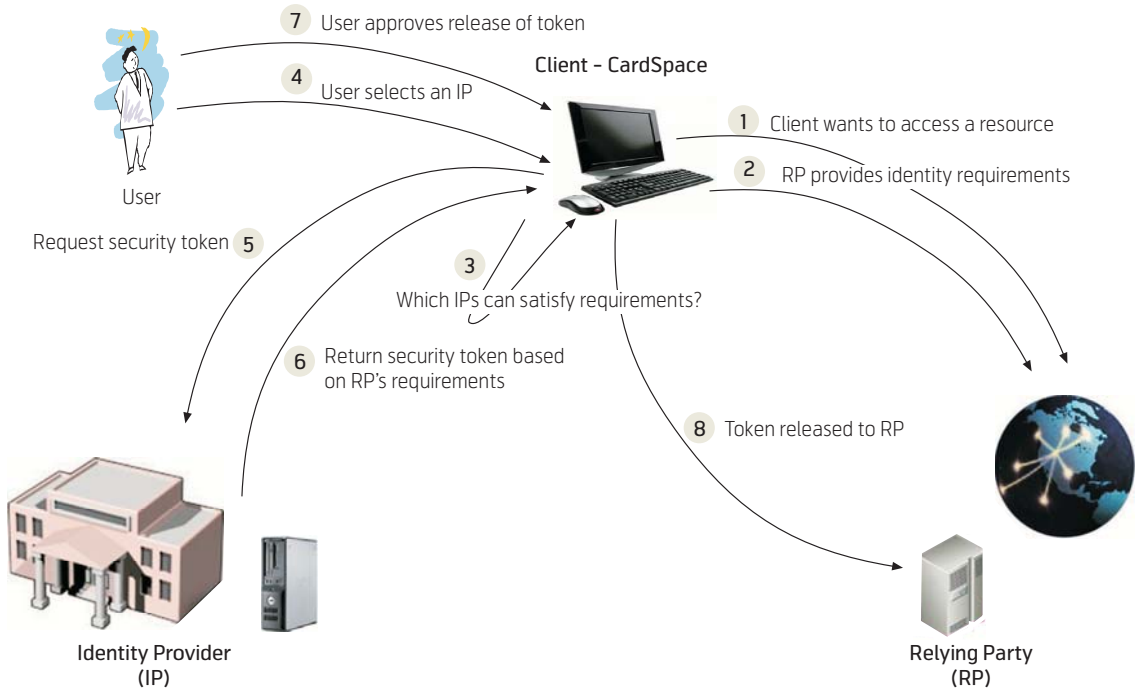


Figure 4 End-to-end scenario

ISSN 0085-7130

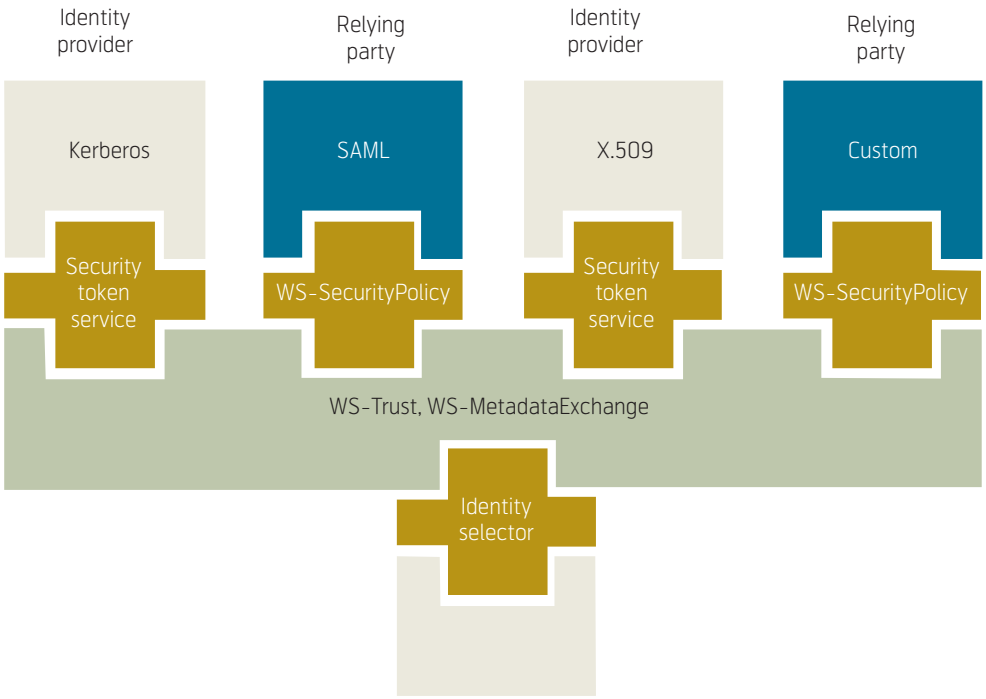


Figure 5

Relying parties express their token requirements via WS-SecurityPolicy (e.g. SAML 1.1 token, X.509-based token).

Identity providers express their token capabilities via WS-SecurityPolicy and do token exchange via Security Token Services (STS) implementing via WS-Trust.

The identity selector pulls everything together, helping the subject match a relying party's requirements to an identity provider's capabilities. After being invoked by an application, it performs the negotiation between relying party and identity provider(s); displays the identities of "matched" identity providers and relying parties to the subject (e.g. the end user); obtains claims; and releases them to the application under the supervision of the subject.

Key Benefits

The key benefits of the identity metasystem are:

- *Greater user control and flexibility*. Users decide how much information they disclose, to whom, and under what circumstances, thereby enabling them to better protect their privacy. Strong two-way authentication of identity providers and relying parties helps address phishing and other fraud. Identities and accompanying personal information can be securely stored and managed in a variety of ways, including via the online identity provider service of the user's choice, or on the user's PC, or in other devices such as secure USB keychain storage devices, smartcards, PDAs, and mobile phones.

- *Safer, more comprehensible user experience*. The identity metasystem enables a predictable, uniform user experience across multiple identity systems. It extends to and integrates the human user, thereby helping to secure the machine-human channel.

- *Increases the reach of existing identity systems*. The identity metasystem does not compete with or replace the identity systems it connects, but rather preserves and builds upon customers' investments in their existing identity solutions. It affords the opportunity to use existing identities, such as corporate-issued identities and identities issued by online businesses, in new contexts where they could not have been previously employed.

- *Fosters identity system innovation*. The identity metasystem should make it easier for newly developed identity technologies and systems to quickly gain widespread use and adoption. Claims transformers can allow new systems to participate even when most participants do not understand their native claims formats and protocols.

- *Enables adaptation in the face of attacks*. New technologies are needed to stay ahead of criminals who attack existing identity technologies. The metasystem enables new identity technologies to be quickly deployed and utilized within it, as they are needed.

ISSN 0085-7130©Telenor ASA 2007

- *Creates new market opportunities*. The identity metasystem enables interoperable, independent implementations of all metasystem components, meaning that the market opportunities are only limited by innovators' imaginations. Some parties will choose to go into the identity provider business. Others will provide certification services for identities. Some will implement server software. Others will implement client software. Device manufacturers and mobile telephone players can host identities on their platforms. New business opportunities are created for identity brokers, where trusted intermediaries transform claims from one system to another. New business opportunities abound.

A benefit we will all share as the identity metasystem becomes widely deployed is a safer, more trustworthy Internet.

Participants in the identity metasystem can include anyone or anything that uses, participates in, or relies upon identities in any way, including, but not limited to existing identity systems, corporate identities, government identities, Liberty federations, operating systems, mobile devices, online services, and smartcards.

References

Kim Cameron's Identity Weblog. September 21, 2007 [online] – URL: http://www.identityblog.com

Microsoft Developer Network. September 21, 2007 [online] – URL: http://msdn.microsoft.com

Cardspace in .NET Framework. September 21, 2007 [online] – URL: http://cardspace.netfx3.com/

Web Services Specifications. September 21, 2007 [online] – URL: http://msdn2.microsoft.com/ en-us/webservices/Aa740689.aspx

Microsoft Open Specification Promise. September 21, 2007 [online] – URL: http://www.microsoft.com/ interop/osp/default.mspx

Acknowledgement

Kim Cameron and Michael B. Jones of Microsoft Corporation.

Ole Tom Seierstad has been with Microsoft Norway for 17 years – his current position is Chief Security Advisor with focus on Microsoft security products and messaging. His previous positions in Microsoft include windows Mobile Evangelist, Technical Support Manager and head of MSN Norway.

email: oles@microsoft.com

ISSN 0085-7130©Telenor ASA 2007

Smart Cards and Digital Identity

JEAN-DANIEL AUSSEL

Jean-Daniel Aussel is Head of the Tools & Application Labs R&D in Gemalto

Smart cards are portable tamper-resistant cryptographic devices that play a key role in digital identity by securely storing the card owner identity attributes and preserving its privacy, and by providing strong authentication of the card owner before releasing identity attributes. Internet authentication has traditionally been performed using Public Key Infrastructure (PKI) and one-time password (OTP) smart cards, mostly for identifying and authenticating corporate users. On the other hand, a huge number of smart cards are deployed by mobile network operators (MNO) to authenticate and identify subscribers to the GSM and 3G networks, and by banks and financial institutions for payment. Large deployments are also on the way for government identification cards or electronic passports. As a result, card issuers like MNOs and banks can reuse their existing infrastructure and act as identity providers to third-party service providers, or service provider can use government cards to identify and authenticate users.

Introduction

Several internet services require the identification of the user, e.g. for home banking, online purchases, voice-over-IP calls, or tax return filing. Identification of the user to access these services involves the disclosure of one or more identity attributes. These identity attributes can be domain-dependent, such as the account number for home banking, or cross-domain, such as the user address. A variable degree of security is required for authenticating the user and protecting her identity attributes. Authentication that results in the proof of identity must be secured to avoid impersonation by a fraudster. Similarly, identity attributes must be securely stored and their access controlled to prevent unnecessary disclosure of identity attributes.

Identification does not always require user authentication. For example, cookies in a browser can be used to store persistent identity attributes without the need to authenticate the user, such as the user's city for a weather forecast site. However, authentication is generally required for identifying users to gain access to valuable services, such as online payment, or services involving privacy such as online consultation of a medical record or tax return.

User name and passwords are traditionally used to authenticate users, and are dimmed as one-factor authentication, i.e. *something you know*. Passwords have two main issues: convenience and security. On the convenience side, the average individual holds several online accounts for online banking, email, social networks, online retailers, etc. It consequently becomes difficult and confusing for consumers to remember all their logins. On the security side, fraudsters have developed several techniques to steal the username and passwords of legitimate users, including key loggers, phishing, pharming and DNS poisoning. Key loggers are Trojan horses maliciously installed on the user computer that record keyboard keystrokes and intercept username and passwords as they are keyed in. In phishing, pharming and DNS poisoning [1], the user is directed to a fraudulent web site that looks like the real service provider site. The user name and passwords are stolen as the user enters them on the fake site.

Stronger authentication is obtained with smart cards, which can provide two-factor authentication, i.e. *something I have* (the smart card) and *something I know* (the personal identification number of the smart card).

The first section of this paper is a quick review of the hardware and software mechanism that makes a smart card tamper resistant against several attacks, and therefore well fit for securely storing identity attributes and providing strong authentication.

The second section presents the conventional smart card strong authentication methods based on Public Key Infrastructure (PKI) or One-Time Password (OTP). PKI and OTP cards are mostly used in corporate environments to secure Virtual Private Networks or intranet web site access. However, more and more PKI-enabled government identity or health cards are deployed around the world and are expected to provide the digital identity of citizens towards online government or commercial services.

A huge number of cards are currently deployed by Mobile Network Operators and financial institutions, who also operate an associated server and cryptographic infrastructure for network access or payment, respectively. Section three presents methods for per-

ISSN 0085-7130

forming strong authentication for web server access control using Subscriber Identity Module (SIM) cards and payment cards.

Finally, the integration of strong authentication in two identity frameworks, Liberty Alliance and Windows CardSpace is discussed. In identity frameworks, the roles of service provider and identity provider are clearly separated. Mobile Network operator or financial institution can operate as identity providers and build value for the end-users and service providers.

Why Smart Cards Are Secure?

A smart card is like an ordinary credit card, except that it has an embedded microchip and metallic contacts. The smart card operates as a very small computer with an embedded operating system that controls application execution, access restrictions and communication with the outside world. The purpose of the smart card is to ensure secure processing and storage of sensitive data and applications. Highly sensitive data such as the user attributes or cryptographic keys to authenticate the user are never released outside the card, and all operations are handled by the operating system of the card.

The security of smart cards is based on a set of components that protect both the physical card and stored data or applications. The first component is the card body. Human-readable techniques (barcodes, holograms, identity pictures) are used to prevent the card body from being physically copied or counterfeited. This offers a first level of security thru visual inspection, which obviously is not relevant for online digital identity.

For protection against physical attacks, functional blocks are mixed, producing what is called a glue logic design. This makes it much more difficult for an attacker to analyze the structure of the logic and locate functional blocks such as the CPU or coprocessors. Buses are scrambled and buried, and thus inaccessible from outside the chip, so that connections cannot be made to recover memory content. Memory is also scrambled, to protect the chip from selective access/erasure of individual data bytes. On top of the physical scrambling, latest chips implement strong ciphering thus preventing the reverse engineering of memory and bus content. A current-carrying protective layer is added at the top of the chip for power supply. If this layer is removed, the chip no longer operates. Finally a set of sensors is activated to detect abnormal variations of voltage, temperature, clock frequency and light.

Side channel attacks on the card are used to recover secrets by monitoring execution time, power consumption or electromagnetic radiation. A well-known class of attacks is based on analysis of smart card power consumption. This class includes Differential Power Analysis (DPA), Simple Power Analysis (SPA) and timing analysis.

Two principles are used for protection against side channel attacks: the first is to reduce as much as possible the power signal and electromagnetic emissions, the second is to add noise, i.e. randomly alter the signals, or add random processor interrupts or change the clock speed.

Fault channel attacks are conducted using a combination of environmental conditions that causes the chip to produce computational errors that can leak protected information. Against fault channel attacks, hardware sensors are used to detect abnormal variations of voltage, frequency, light and temperature. In addition, random delays are added to the code, making it difficult to identify when to inject a fault, and redundancy and consistency checks are implemented to prevent erroneous executions to compromise sensitive functions.

The security of the smart cards against physical and logical attacks has been achieved thru the development of advanced counter-measures, and as a result, smart cards are the de-facto standard for digital security, and as such are the most deployed personal computing device as shown in Table 1.

Public Key Infrastructure Strong Authentication

Public Key Infrastructure (PKI) smart cards are routinely used on personal computers for authentication and identification of users, mostly in the corporate world. PKI smart cards provide two-factor authentication, i.e. *something you have*, the smart card, and *something you know*, the card Personal Identification Number (PIN). The main operating systems have

Personal Computing Device	2006 Worldwide shipments in Millions of Units
Personal Digital Assistants (PDA)	18
Personal Computers (PC)	232
Mobile Phones	1000
Microprocessor Cards	2655

Table 1 Personal computing devices worlwide shipments. Source: Gartner and Eurosmart for Microprocessor Cards

ISSN 0085-7130

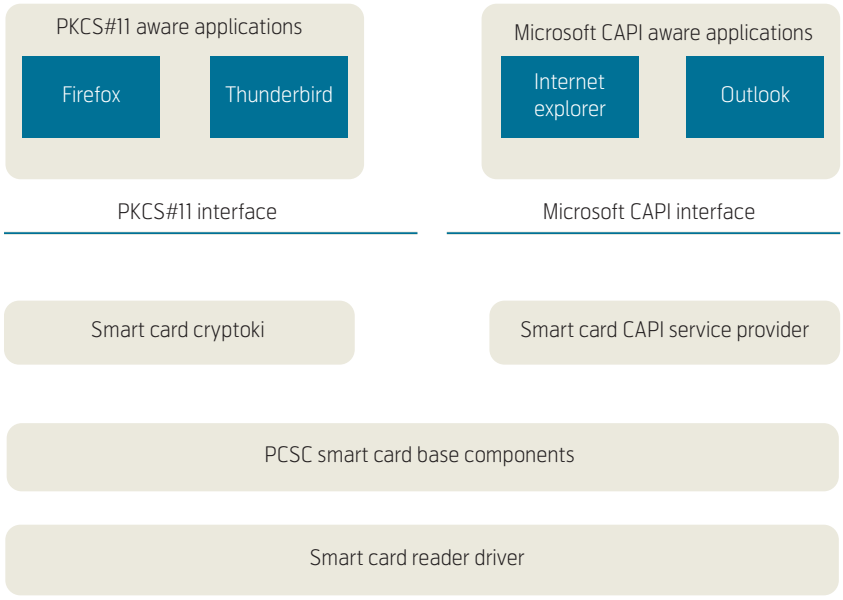


Figure 1 PKI aware applications interface to the smart card using the PKCS#11 or Microsoft CAPI interfaces. Smart card vendors typically provide a PKCS#11 cryptoki library or a Microsoft CAPI service provider

smart card support for establishing VPN connections, most browsers are smart card enabled to perform secure connections using the Secure Socket Layer (SSL) or Transport Layer Security (TLS) protocols, and several applications such as e-mail client are PKI aware and can perform for example digital signature or mail encryption.

PKI smart cards contain one or several X509 v3 [2] certificates. A X509 v3 certificate is a digital certificate containing among other things subject identity attributes, such as the common name, the public key of the subject, the certificate issuer, and the description of the PKI algorithms, such as RSA or DSA, and finally a digital signature of the certificate by the certificate issuer. For strong authentication, the private key of the subject is stored securely inside the smart card. The basic principle for authentication is to ask the user to sign a challenge with her private key, and verify the signature with the public key of the user which is stored in the certificate.

For proof of identity, the user presents her certificate containing the public key to a 3rd party. This certificate presentation is done programmatically by the client software. For example during the establishment of an SSL/TLS connection with a browser, the user certificate is transmitted during the client hello message. The 3rd party then challenges the identity of the subject by requiring the subject to sign a challenge with its private key. Successful authentication is obtained if the signed challenge can be recovered with the public key of the certificate. So far, this only proves that the subject is the valid owner of the certificate, but it does not certify its identity. To do so, the 3rd party checks the certificate validity by verifying that the certificate is effectively signed by the certificate issuer. This verification is done using the public key of the certificate issuer.

The PKI client, e.g. browser, mail client, or VPN client, interfaces to the smart card using two main industry standards: Microsoft CAPI [3] and PKCS#11 [4], as shown in Figure 1. The use of standard APIs allows the plug-in of different implementations of these cryptographic components, called cryptographic service providers (CSP) for CAPI and cryptoki for PKCS#11. PKI smart cards are essentially limited to the corporate environment usage for securing web access or establishing VPN connections. The deployment in the consumer market is limited by the required issuance of smart cards and smart card reader, their associated device drivers, the post-issuance management of the cards, and the certificate and certificate revocation list management.

However, more and more identity smart cards with PKI features are deployed in several countries such as Belgium, Italy, Spain, Estonia, Austria. These identity cards aim to provide to the citizen identification, authentication and signature features, for access to a wide range of online services, such as online tax return. Companies like banks will also use the digital identity of the citizen and the associated strong authentication. The European Committee for Standardization has standardized the signature card [5] and the European Citizen Card [6]. An ISO standardization effort has also started to standardize the cryptographic interface of the applications to the smart card [7].

ISSN 0085-7130©Telenor ASA 2007

One-time Passwords

An alternate method for authentication using smart cards is the one-time password (OTP). An OTP is a generated password valid only once. The user is given a device that can generate an OTP using an algorithm and cryptographic keys. On the server side, an authentication server can check the validity of the password by sharing the same algorithm and keys.

Several software or devices can be used to generate the OTP, including personal digital assistants, mobile phones, dedicated hardware tokens, the most secure means being smart cards which provide tamper-resistant two-factor authentication: a PIN to unlock the OTP generator *(something you know)*, and the OTP smart card itself *(something you have)*. Figure 2 illustrates the three steps required to generate an OTP: the collection of some external data, such as the time for synchronous OTP or a challenge for an asynchronous OTP, a ciphering algorithm with secret keys shared by the device and the authentication server, and finally a formatting step that sets the size of the OTP to typically six to eight digits.

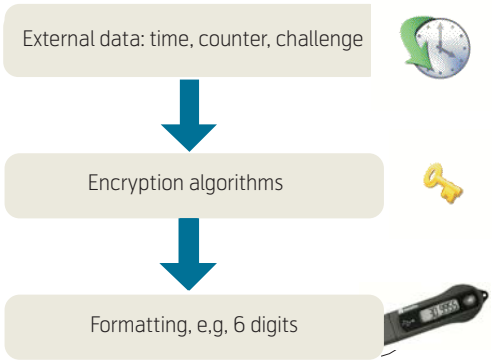


Figure 2 The generation of One-Time passwords generally involves three steps: the collection of authentication information like the value of a counter, the time, or a challenge, a ciphering algorithm applied on this external information, and finally formatting of the OTP to a typical length of 6 to 8 digits

Until recently, OTP solutions were based on proprietary and often patented time-based or event-based algorithms. In 2005, OATH-HOTP [8] was defined as an open standard by major actors in the industry. This open standard allows multi-sourcing of the OTP generating devices and authentication servers from different vendors. The HOTP algorithm is based on a

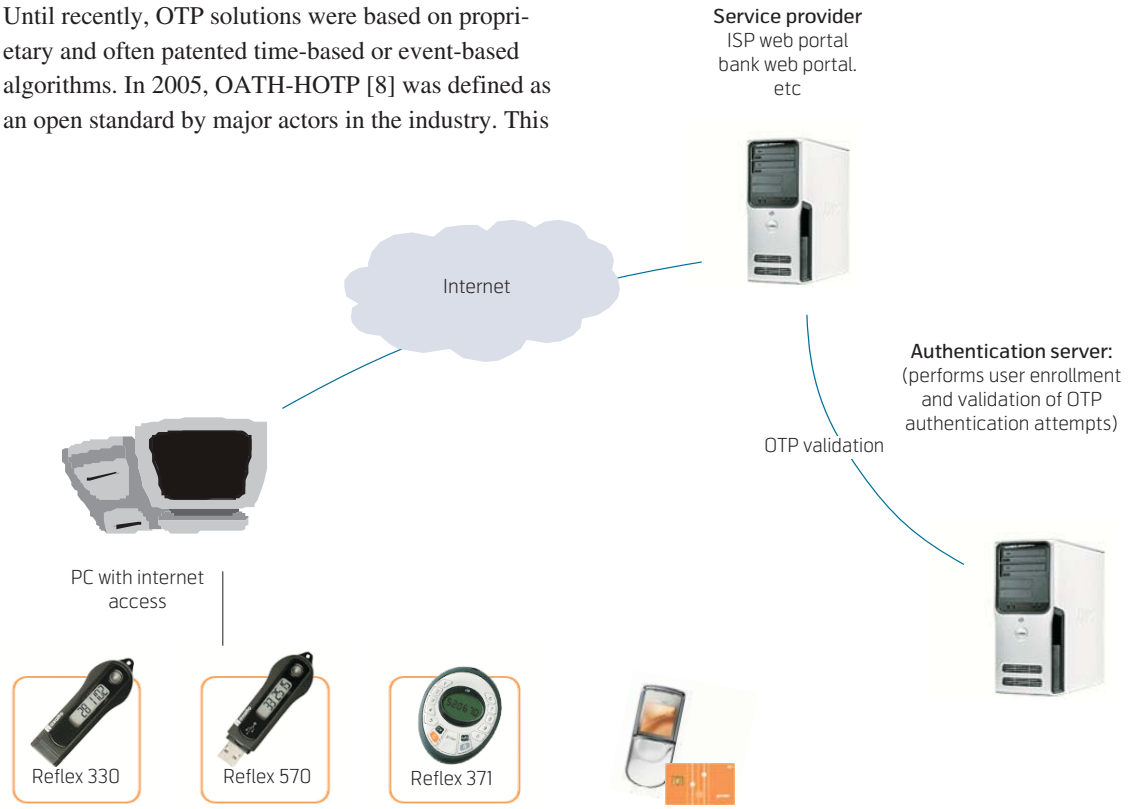


Figure 3 Authentication with smart card based OTP. On the server side, an authentication server validates the OTP passwords entered by the users on the service provider login page. On the user side, several devices can be used to generate the OTP. From left to right, the first two devices are one-factor authenticators, i.e. something I have*, and do not require a PIN. The first device is a smart card with a simple display and push-button, the second device has in addition a USB interface that can be connected to the PC and perform automated form-filling of the password in the browser. The third device is a two factor authenticator, where a PIN or a challenge can be entered on the device to generate the OTP. In the fourth device, the smart card inside the mobile phone equipment generates the OTP and uses the handset display and keyboard using the SIM toolkit programming interface*

ISSN 0085-7130

secret key and a counter shared by the device and the server, and uses standard algorithms such as SHA-1 and HMAC.

OTP has some advantages over PKI in that it does not require the deployment of smart card readers, drivers and PC software. However in terms of features, OTP only provides identification and authentication, whereas PKI provides in addition encryption and signature. OTP being a password-based authentication is also vulnerable to man-in-the-middle attacks, such as phishing scams. Since there is no mutual authentication of the PC and the internet service provider server, an attacker can intercept an OTP using a mock-up site, and impersonate the user to the real internet web site.

Banking Cards and Mobile Phone One-Time Passwords

The identification of the users using strong authentication and smart cards, either OTP or PKI, requires both the deployment of devices to the customers, and the operation of authentication servers. Two industries however already have a wide base of issued smart cards or devices and their associated authentication servers: the financial institutions and the mobile network operators.

Eurocard Mastercard Visa (EMV) smart cards are now the standard in Europe, and are gaining momentum in Asian and South American countries. Financial institutions are turning to two-factor authentication to secure their online services. Mastercard [9] and Visa [10] have develop OTP generation algorithms, in which the cardholders use their smart card payment card and a hardware device to generate the OTP. The OTP generation uses the built-in EMV application of the payment card, with a dedicated EMV key and counter storage. In addition, the devices can implement transaction signature, for example money transfer signature, to prevent man-in-the-middle attacks and render impossible the altering of the transaction parameters on the fly.

Mobile network operators (MNO) have an even wider base of smart cards installed inside the end-user handsets, Subscriber Identity Module (SIM) for the 2nd generation network (2G), or Universal Subscriber Identity Module (USIM) for the UMTS or 3G network. MNO applications can be loaded onto the (U)SIM card and use the handset display and keyboard for user interaction thru the SIM Toolkit standardized interface [11]. As a result, MNOs have started to offer OTP authentication for access to their subscriber services portal. A first solution is to generate the OTP offline and send it using a Short Message (SMS) over-the-air to the card. The second solution is to load an OTP generating SIM Toolkit application on the SIM/USIM card, and to generate and display the OTP using this card application. The on-card generation is more secure, since it can prompt for a challenge or a PIN before generating the OTP.

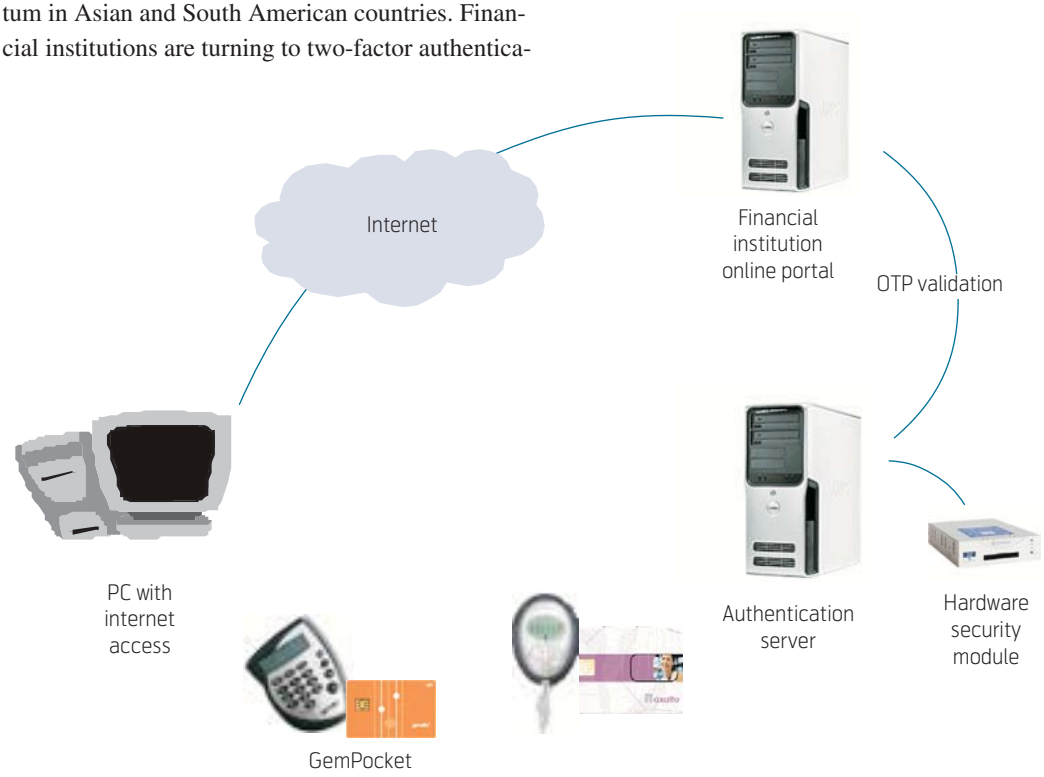


Figure 4 Authentication with OTP generated by EMV banking cards inside a dedicated reader device. The device uses the on-card EMV application and keys to generate the OTP from the EMV transaction counter. The authentication server uses the same hardware security module used for validating EMV payment transactions

ISSN 0085-7130©Telenor ASA 2007

EAP-SIM and EAP-AKA Wireless Authentication

For MNOs, the OTP solution still requires the installation on the operator network of an authentication server managing the validation of the passwords, and managing the identities of the card holders. In the case of OTP generation using OATH-HOTP, the server would have to manage the identities, counter values, secret keys for the users, and the computation of the OTP using the HOTP algorithm. All this is duplicating the existing MNO infrastructure, which already manages a card holder database in the Home Location Register (HLR), with a secret identity key K_i for each user both on the HLR and the SIM card, and an identification GSM algorithm for the network.

This problem is similar to the identification and authentication of users towards wireless network access points, in which initially the authentication protocol was based on Extensible Authentication Protocols (EAP) [12] such as EAP-TLS [13], or EAP-PEAP. EAP-PEAP uses server-side certificates and authenticates the end-user with a login/password encrypted with an SSL/TLS tunnel, and is therefore a weak one-factor authentication. EAP-TLS is based on PKI and mutual authentication, and is a two-factor authentication when using smart cards. However, EAP-TLS requires client and server side certificates. To avoid deployment of a PKI infrastructure, including certificate generation, deployment and management, and the operation of certificate authorities and certificate revocation lists, two EAP protocols have been specified for Wireless LAN authentication: the EAP-SIM [14] protocol, based on the SIM, and the EAP-AKA [15] protocol, based on the (U)SIM. Both protocols have the advantage of using the MNO existing cryptographic infrastructure, i.e. the algorithms and associated keys. The EAP-SIM interface between the PC WiFi network components and the SIM has been further standardized by the ETSI [16] and the WLAN smart card consortium [17].

Figure 5 describes the components of 802.11 authentication to a wireless access point using EAP-SIM. A SIM card is plugged into the personal computer using a smart card reader, which can have a USB token form factor. The smart card issuer provides an EAP-SIM supplicant, which is a system library that interfaces the networking component to the SIM. The supplicant implements the authentication protocol and required calls to the card as per WLAN-SIM specification. On the network side, the WiFi access point sends EAP messages to the authentication server, which is interfaced to an HLR thru an IP/SS7 gateway. The authentication server can in this way request cryptographic data to validate the authentication. Upon successful authentication, the network access point opens the access to the internet.

GSM authentication is based on a challenge/response mechanism. The SIM card and mobile operator server share a secret key K_i. The A3/A8 authentication algorithm that runs on the SIM card is given a 128-bit random number RAND as a challenge, and computes a 32-bit response SRES and a 64-bit key K_c from the challenge and K_i. The challenge RAND, 32-bit response SRES and K_c constitute a triplet. On the server side, the EAP messages are processed by a radius server connected to the subscriber Home Location Register (HLR) thru an IP/SS7 gateway.

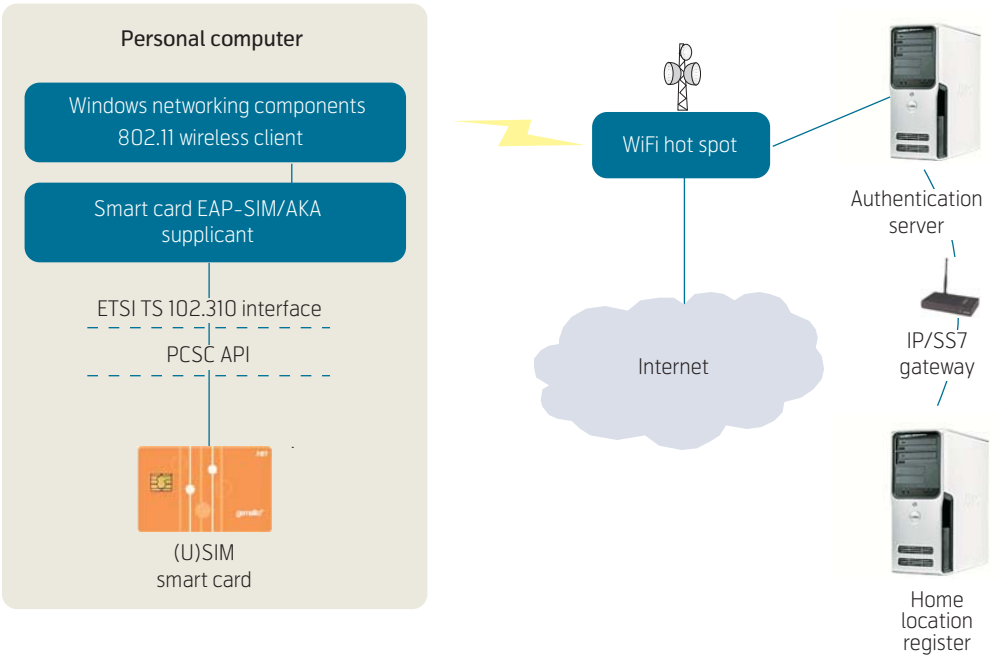


Figure 5 EAP-SIM 802.11 wireless strong authentication with smartcards

ISSN 0085-7130

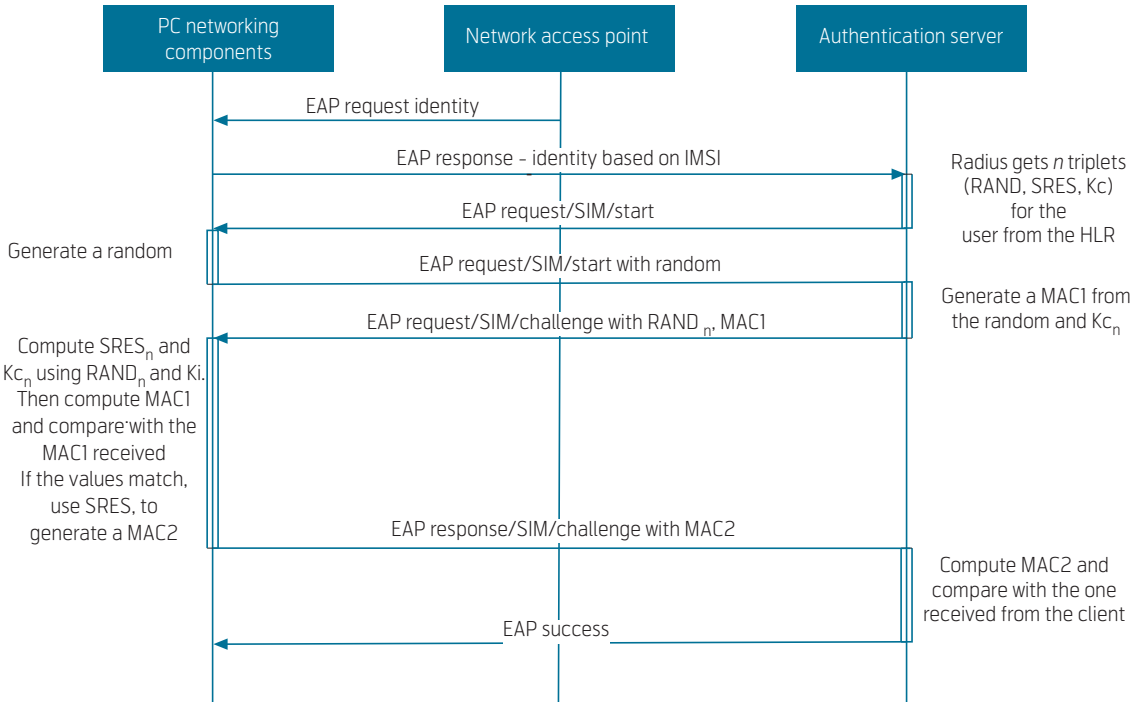


Figure 6 EAP-SIM authentication message flow. The shared keys are only stored securely in the HLR or in the SIM card. The radius server only retrieves a series of triplets from the HLR to optimize network connection. On the PC side, all cryptographic operations are performed by the SIM card, as specified by the WLAN-SIM specification

The radius server can retrieve a set of triplets from the HLR and perform authentication, as described by the message flow of Figure 6.

Multiple authentication triplets can be combined to create authentication responses and encryption keys of greater strength than individual triplets. EAP-SIM also includes network authentication, user anonymity and fast re-authentication.

EAP-SIM and EAP-AKA Authentication for Internet Services

EAP-SIM and EAP-AKA were initially designed to identify and authenticate card holders for wireless network access. This identity is the network identity of the user. However, the identification of the users to access internet services is not necessarily the same as the network identity for several reasons. First, the personal computer could be shared by several users. Second, the internet services to access might be provided by different business units or companies from the internet provider, in which case the identity of the network account cannot be retrieved. Finally, some services require explicit user consent or proof of presence, and require more identity attributes than the simple connection identifier. For these reasons, application level authentications using EAP-SIM or EAP-AKA have been developed to authenticate a user to an internet service [18]. These authentication methods are mostly used for identifying users accessing web servers from a browser, but can be extended to any client protocol, such as the Session Initialization Protocol (SIP) for Voice-over-IP.

For network authentication, the EAP allows for arbitrary authentication methods such as EAP-TLS, EAP-PEAP, EAP-SIM or EAP-AKA. The EAP messages are transported without interpretation over the network components, e.g. the WiFi access point, and are only interpreted by the supplicant and smart card on the PC side, and by the radius server authentication policy. In the case of WLAN-SIM, the EAP messages are even not interpreted by the PC supplicant but just transmitted to the smart card that performs the complete EAP messages processing.

A web extensible authentication framework has been built on this principle, for browser authentication with EAP[13]. The extensible authentication framework components are shown in Figure 7. When connecting to a service provider web site from a browser, the user is directed to an authentication url that holds an EAP gateway java servlet. By accessing the EAP servlet, the browser loads a signed ActiveX for Internet Explorer or a plug-in for Firefox, the Card Access Module (CAM). The EAP servlet and the CAM are then acting as gateways that carry transparently EAP messages between the smart card and the Radius server.

ISSN 0085-7130©Telenor ASA 2007

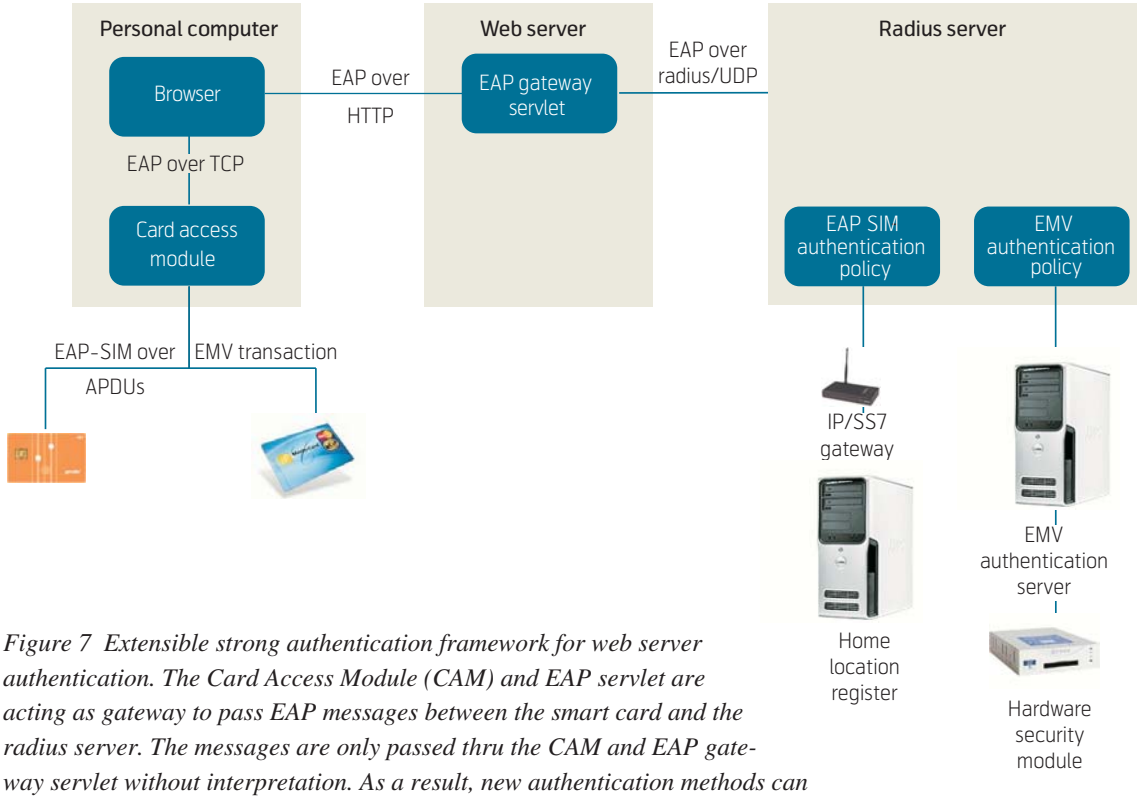


Figure 7 Extensible strong authentication framework for web server authentication. The Card Access Module (CAM) and EAP servlet are acting as gateway to pass EAP messages between the smart card and the radius server. The messages are only passed thru the CAM and EAP gateway servlet without interpretation. As a result, new authentication methods can be implemented by writing the corresponding authentication policies

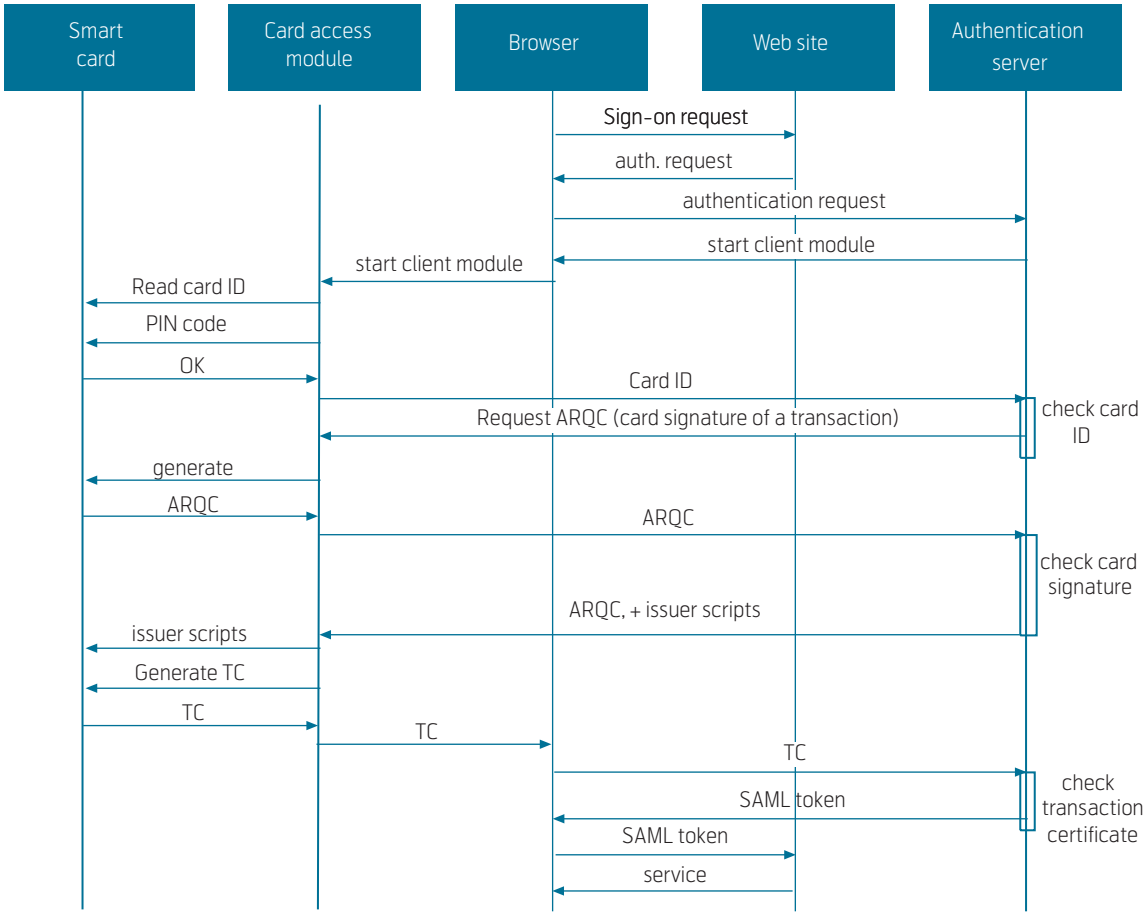


Figure 8 EMV authentication to a web server using the extensible authentication framework. The EMV authentication is performed by completing a zero-amount EMV payment transaction

ISSN 0085-7130

Several authentication protocols can be implemented on top of this framework, the EAP gateway and CAM acting only as message gateways that do not process the EAP messages. Supporting a protocol requires the implementation of an authentication policy plug-in on the radius server. In Figure 7 an EAP-SIM authentication policy is communicating with a SIM smart card thru the network to perform an EAP-SIM authentication, and alternatively an EMV authentication policy is communicating with an EMV smart card [19]. In the case of the EAP-SIM authentication, the messages are the same as defined in the WLAN-SIM specification.

For the EMV strong authentication, a complete payment transaction with a zero amount is performed to authenticate the user, and the message flow is described in Figure 8.

Although most browsers support natively PKI authentication, this extensible authentication framework has the advantage to be open to new protocols, and hence allow the reuse of an existing infrastructure of cards, cryptographic devices and authentication servers. Typically, financial institutions can reuse their issued cards and payment servers by implementing an authentication based on the EMV specifications, or mobile network operators can reuse their existing HLR and deploy SIM cards for PC authentication to their subscribers.

Compared to the OTP authentication, this framework can implement protocols with mutual authentication of the card and server, such as EAP-AKA, and hence avoid man-in-the-middle attacks.

Strong Authentication and Identity Frameworks

The strong authentication methods presented so far assume that the service provider authenticating the user is the card issuer, with the exception of PKI authentication with citizen cards emitted by government or health care.

Deploying a strong authentication solution has a cost: procurement of the identity smart cards and devices such as smart card readers, 24/7 operation of the authentication server, deployment of drivers and middleware for the client PCs, operation of a customer care center and card management system for post-issuance operations, such as unblocking a PIN.

On the other hand, several actors such as MNOs or financial institutions already have a huge installed base of smart cards and devices such as handsets, as well as a server infrastructure and customer care centers. Recently, several identity frameworks have been specified, which formalize the roles of the different actors and allow a clear separation between the identity providers and the identity consumers. This separation brings value to all the identity actors.

For MNOs or financial institutions, operating identity provider services using their infrastructure can provide new sources of revenues not based on sale of air traffic or payment transactions, improve customer loyalty, attract new business customers and strengthen their position by extending the conventional role and values to the internet world.

For service providers, delegating identity and authentication to identity providers provides a higher level of security, cost saving by stopping the operation of the existing authentication schemes, lowering threshold for deployment since the identity provider manages most of the infrastructure, simpler customer management and the ability to reach more customers that are subscribers of the identity provider.

Finally for the end-users, having a centralized identity provider provides a better control and management of their identities, e.g. fewer passwords to remember, better protection and higher level of security with strong authentication, single-sign-on (SSO) with framework that supports it, and universal applicability to various services.

Some of the recent identity frameworks include Liberty Alliance [20], OpenID [21], and Microsoft CardSpace [22].

Liberty Alliance is a consortium of industries that defines a set of specifications for identity federation and single-sign-on. Identity federation in Liberty Alliance is based on the Security Assertion Markup Language (SAML) defined by OASIS [23]. In Liberty Alliance specification, single-sign-on (SSO) is performed using browser redirection, as shown in Figure 9.

When the user is requesting a web page from the service provider that requires authentication, the service provider redirects the authentication request to the identity provider (IDP). The IDP authenticates and identifies the user, and returns upon successful authentication a SAML token to the service provider using browser redirection. The service provider can optionally validate further the token offline, and gives access to the required service if the SAML token is valid.

SSO requires a one-time initialization phase called federation, in which the IDP and service provider

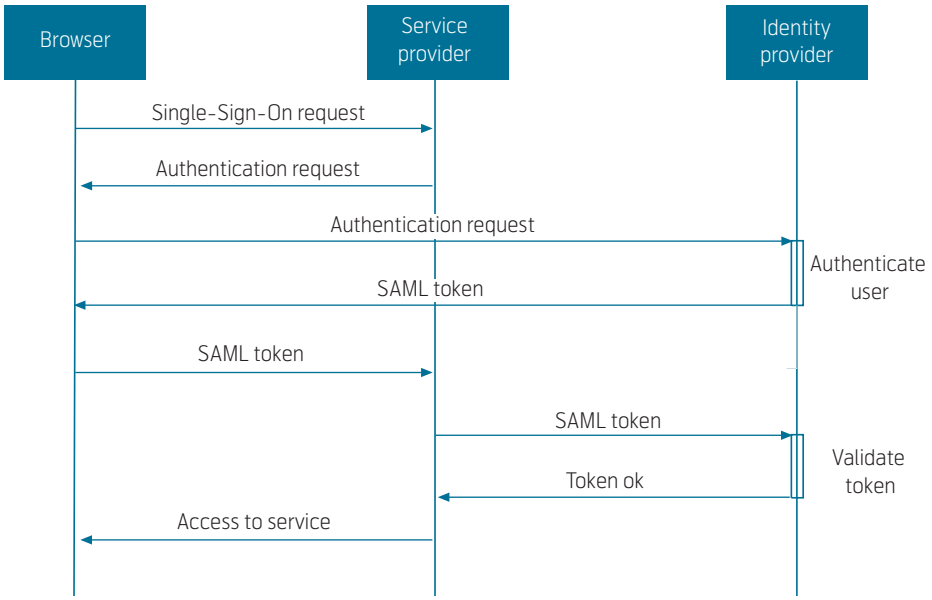


Figure 9 Liberty Alliance single-sign-on data flow

exchange an opaque identifier to the user. This opacity ensures that the IDP and service provider do not share the respective identity of the user.

Liberty Alliance specifications do not specify the authentication methods. As a result, there is no standard strong authentication method implemented in the IDP products of the different vendors, nor is there a framework for plugging authentication methods. As a result, integrating a strong authentication method in Liberty Alliance currently requires case-by-case integration with the different commercial IDP offers. In the scope of the Celtic Fidelity [24] Eureka project, the EAP-SIM/AKA method has been implemented in IDPs from different vendors, and allowed several MNOs such as Telenor, TeliaSonera and Orange to operate pilot IDPs.

An interesting variant of the EAP-SIM strong authentication method for web access control has been designed within the SIMStrong consortium [25], which has the advantage of avoiding the deployment of SIM cards with a USB form factor. In this solution, the Over-The-Air channel (OTA) is used to perform an EAP-SIM authentication between the radius server and the SIM card inside the handset, as described in Figure 10.

In this solution, called SIMStrong-over-SMS, when the end-user is redirected to the IDP for authentica-

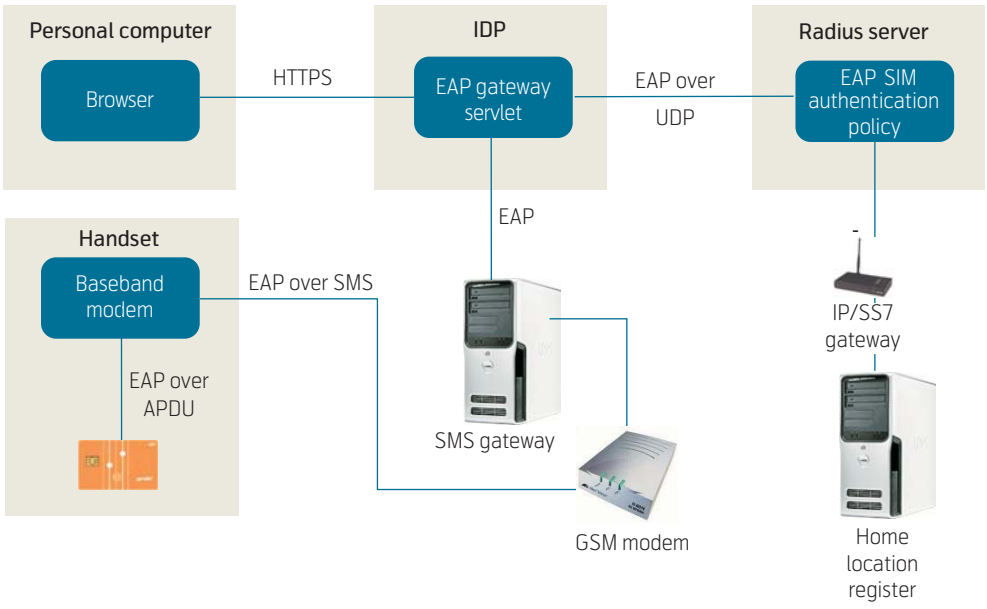


Figure 10 Liberty Alliance strong authentication using over-the-air short-messages

ISSN 0085-7130

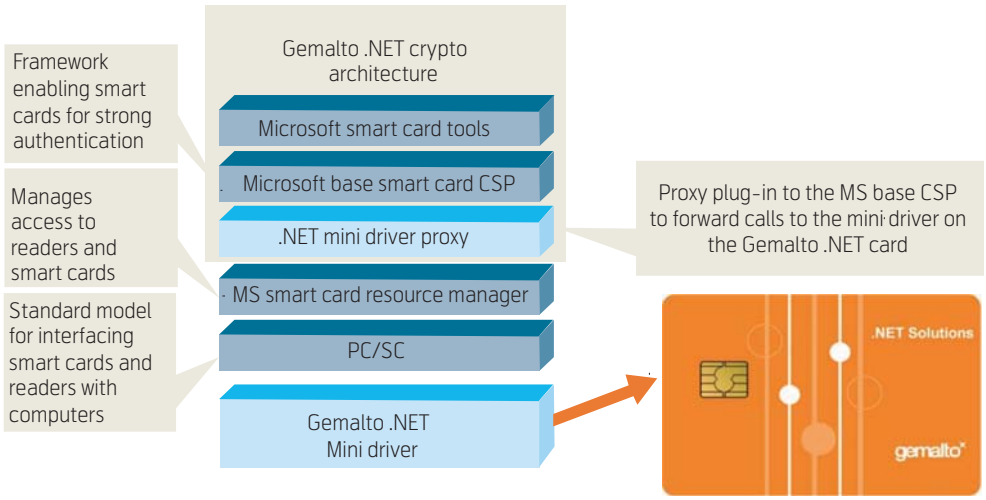


Figure 11 Gemalto .Net Architecture to support Windows CardSpace X509 authentication

tion an end-to-end EAP-SIM protocol is performed between the SIM card in the handset and the Radius server. All EAP messages are exchanged over SMS between the card and the IDP, and as standard radius messages over UDP between the IDP and the radius server. A SIM toolkit applet in the SIM card prompts for user-consent on the handset, and on consent and successful authentication, the SAML token is returned by the IDP to the browser, who is then authenticated towards the service provider.

Windows CardSpace is another identity framework released with Windows Vista and .NET framework 3.0. CardSpace is a claim-based identity management system, in wich a web service provider, called Relying Party in the CardSpace framework, requests identity claims from the user. The user can select a virtual card thru a card selector that provides the required claims. Some cards are self-managed, i.e. the claims are not certified, but other claims are certified and managed by an identity provider. Self-managed cards are like user name/password chosen by a user to access a service, without any verification of the real identity of the user. Managed cards have an identity certified by an identity provider operating a Secure Token Server (STS). To retrieve the claims of a managed card, the card-holder must authenticate to the STS, which returns an encrypted and signed token that can be further presented to the Service Provider.

Windows CardSpace authentication supports login/password, Kerberos and X509 certificates, which limits the possibility of integrating a strong authentication protocol inside the CardSpace selector. Strong authentication with smart cards can be performed using either OTP or PKI.

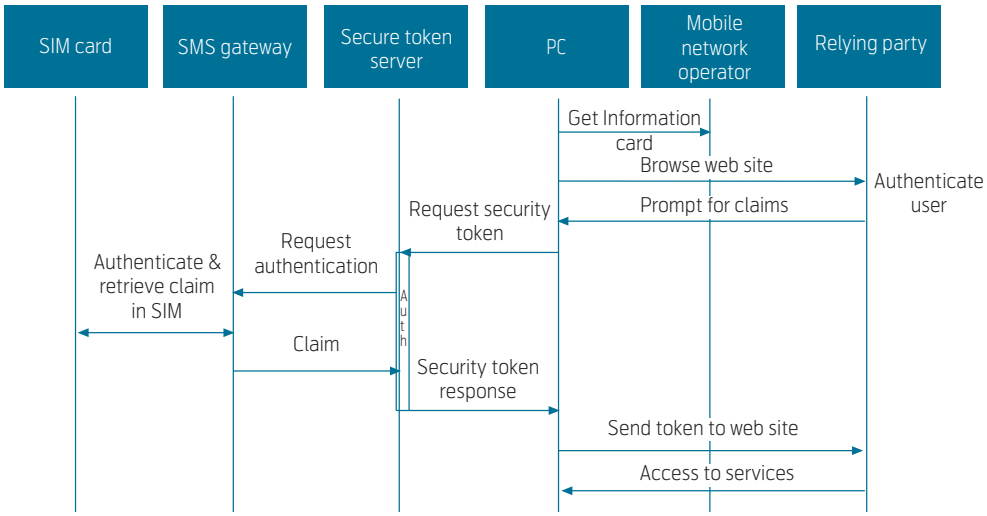


Figure 12 CardSpace strong authentication using the OTA channel. Upon authentication request from CardSpace when selecting the managed card, the STS authenticates the user over the air and retrieves the user's claims inside the SIM card. A SIM toolkit applet prompts the user for consent to publish the identity attributes

ISSN 0085-7130©Telenor ASA 2007

Integrating OTP strong authentication in CardSpace is straightforward: the managed card is a login/password card type, in which the user enters the OTP generated by the smart card device. On the server side, the STS is connected to an authentication server that validates the OTP, and there is no restriction on the type of OTP algorithm.

Windows CardSpace X509 strong authentication is based on PKI, in which the STS authenticates the user using a challenge-response mechanism based on the X509 certificate of the user in the managed card and a private key stored in the smart card. CardSpace client components are accessing the smart card thru a new API, the Crypto API Next Generation (CNG). Smart cards providers typically write a smart card mini-driver [26], also known as a card module, to interface their smart card to the CNG. The CardSpace selector implements the logics to perform the PKI authentication by calling the Base Smart Card Cryptographic Service Provider (CSP). Figure 11 shows the current implementation of X509 CardSpace using a .Net card, which is a smart card with an embedded .Net virtual machine. The base CSP performs the required cryptography with the .Net smart card using the associated mini-driver proxy that forwards the calls to the .Net Mini Driver.

Adding another type of authentication than OTP and X509 to the CardSpace selector is not possible, since the selector is a closed-source component provided by Microsoft. However, using a second channel, such as the over-the-air channel for mobile network operators allow to perform any type of strong authentication in background between the STS and the card. This has been implemented for SMS strong authentication [25] as illustrated in Figure 12.

Conclusion

Smart cards are tamper-resistant devices that can play a key role for storing the identity attributes of the user, or performing strong authentication for proof of identity.

Citizen cards are emerging in several countries, are based on PKI, and can provide identification, authentication and signature services. The electronic identity of these citizen cards is guaranteed by the authorities, and authentication can be performed online using the card issuer certificates without requiring connection to an identity provider.

Financial institutions and Mobile Network Operators have issued a huge number of payment cards and SIM cards and are operating the associated cryptographic server infrastructure. They are as such well positioned to operate identity provider services for end-users and 3rd party service providers.

References

1 *Anti-Phishing working group*. August 31, 2007 [online] – URL: http://www.antiphishing.org

2 Housley, R, Polk, W, Ford, W, Solo, D. *Certificate and Certificate Revocation List (CRL) Profile*. IETF, April 2002. (RFC 3280)

3 Microsoft. *Cryptography API*. August 31, 2007 [online] – URL: http://msdn2.microsoft.com/en-us/library/aa380255.aspx

4 RSA Laboratories. *Cryptographic Token Interface Standard*. June 2004. (PKCS#11 v2.20)

5 CEN/ISSS. *Fundamental specification : application smart card used as secure signature creation device – Part 1 Basic Requirements, Part 2 Optional Features*. European Committee for Standardization, March 2007. (CEN/ISSS EN 14890-1&2)

6 CEN/TC. *European Citizen Card – Part 1 Physical, electrical and transport protocol characteristics, Part 2 Logical data structure and card services*. European Committee for Standardization, April 2007. (Technical Committee CEN/TC 224, Technical Specification 15480-1&2)

7 ISO/IEC. *Integrated circuit card programming interfaces – Part 1: Architecture, Part 2: Generic card interface, Part 3: Application interface, Part 4: API Administration*. 2006. (ISO/IEC 24727-1&2&3&4)

8 M'Raihi, D, Bellare, M, Hoornaert, F, Naccache, D, Ranen, O. *HOTP : An HMAC-Based One-Time Password Algorithm*. December 2005. (RFC 4226, IETF)

9 Mastercard. *OneSmart Authentication*. August 31, 2007 [online] – URL: https://mol.mastercard.net/mol/molbe/public/login/ebusiness/smart_cards/one_smart_card/biz_opportunity/cap/index.jsp

10 Visa. *Dynamic passcode authentication*. August 31, 2007 [online] – URL: http://www.visaeurope.com/aboutvisa/products/dynamicpasscode.jsp

11 ETSI. *Specification of the SIM Application Toolkit for the SIM – Mobile Equipment Interface, GSM 11.14 v. 5.9.0*. 1996.

ISSN 0085-7130©Telenor ASA 2007

12 Aboba, B, Blunk, L, Vollbrecht, J, Carlson, J, Levkowetz, H. *Extensible Authentication Protocol (EAP)*. IETF, June 2004. (RFC 3748)

13 Aboba, B. *PPP EAP TLS Authentication Protocol*. IETF, October 1999. (RFC 2716)

14 Haverinen, H, Salowey, J. *Extensible Authentication Protocol Method for GSM Subscriber Identity Modules (EAP-SIM)*. IETF, January 2006. (RFC 4186)

15 Arkko, J, Haverinen, H. *Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement (EAP-AKA)*. IETF, January 2006. (RFC 4187)

16 ETSI. *Smart Cards: Extensible Authentication Protocol support in the UICC, V6.2.0*. September 2005.

17 WLAN Consortium. *EAP-SIM Handler Specification Version 1.1*. August 1, 2004.

18 Van Thanh, D et al. Offering SIM Strong Authentication to Internet Services. *SIMstrong White Paper*, 3GSM World Congress, Barcelona, February 13-16, 2006.

19 EMVCo. *EMV 4.1 Specifications*. August 31, 2007 [online] – URL: http://www.emvco.com/specifications.asp. (June 2004)

20 *Liberty Alliance Specifications*. August 31, 2007 [online] – URL: http://www.projectliberty.org/specifications__1

21 *OpenID Specifications*. August 31, 2007 [online] – URL: http://openid.net/specs.bml

22 *Windows CardSpace*. August 31, 2007 [online] – URL: http://cardspace.netfx3.com/

23 OASIS. *SAML v2.0 specifications*. August 31, 2007 [online] – URL: http://www.oasis-open.org/specs/index.php#samlv2.0. (March 2005)

24 *FIDELITY – Federated Identity Management based on LIBERTY*. August 31, 2007 [online] – URL: http://www.celtic-fidelity.org/fidelity/index.jsp

25 Van Thanh, D et al. Unified SIM Strong Authentication for CardSpace and Liberty Alliance. *3GSM World Congress*, Barcelona, February 12-15, 2007. Available at http://www.simstrong.org.

26 Microsoft. *Smart Card Minidriver Specification for Windows Base Cryptographic Service Provider (Base CSP) and Smart Card Key Storage Provider (KSP), Version 5.06a*. January 2007.

Jean-Daniel Aussel is Head of the Tools & Application Labs R&D in Gemalto, in the Technology and Innovation division. Gemalto is a provider of end-to-end digital security solutions, from the development of software applications through design and production of secure personal devices such as smart cards, SIMs, e-passports, and tokens to the management of deployment services for its customers. Jean-Daniel holds a PhD from the INSA Engineering School in Lyon, France, and has been working in the smart card industry successively in Bull, CP8, Schlumberger smart cards, axalto, and currently Gemalto, created from the merge of the two smart card market leaders gemplus and axalto. Before working in smart cards and security, Jean-Daniel has been successively working in digital signal processing at the Research Council Canada and Ultra Optec, a small Canadian start-up, and later on designing personal computer and server operating systems at Prologue Software.

email: jean-daniel.aussel@gemalto.com

ISSN 0085-7130©Telenor ASA 2007

Trusting an *e*ID in Open and International Communication

SVERRE BAUCK

Sverre Bauck is Senior Adviser in Brønnøysund Register Centre, Norway

The acronym *e*ID is in common use; related issues like federating *e*IDs from different issuers, Card Space and Open ID are being discussed. But the processing of how to decide to trust an *e*ID in open and international communication has hardly been addressed.

An eID is a secure electronic representation of an ID, and both versions will be trusted and accepted only if the issuer and the identifying documentation are recognized and found reliable. The issuer needs to be identified in a processable way; this means that an identifying attribute according to international standards should be used. Further, the identifying attribute should open for automated checking in the public business register where the issuer is listed; it should also be possible to find out whether the company has been registered as an issuer of *e*ID. However, necessary international standards for such automated processes do not yet exist, and initiatives for their development should be launched.

Standardized identifying attributes for companies would open for secure and automated international communication with verified issuers and issuees, when both are companies. Similarly standardized identifying attributes and corresponding automated processes for individuals need to be agreed and put into use in order to give content for open and international usage to the acronym *e*ID.

Introduction

In this context an ID is a document or card that displays the issuer's name and other attributes, his understanding of the identity of a person and elements that enhance its authenticity. The identity will be presented by a selected set of attributes, such as name, date of birth, citizenship, issuer's unique identifier of the bearer; the latter will be a reference to a register of individuals. Normally, some biometric information, like picture, height and others might be included.

The ID document was designed for evaluation by a human who will examine it for falsification and if the bearer seems to be the rightful one. If the ID is shown in order to give access to a building or transportation, the picture seems to be the most important piece of information on the card. The same document can be used to get access to bank services, but in this case a more profound validation will take place; the card's validity and the bearer's look and basic customer data will be examined.

The set of person identification data has not been subject to a general international standardization; an ID is not a well defined document. However, the structure of widely accepted IDs fits into the following model:

Issuer	Certifies relation to	Subject

The issuer will be recognized by a trustworthy name, logo and visual security mechanisms. Passports are the most accepted ID for general use internationally; their validity can be checked with police authorities in all countries, and they display identification data according to standards processed by ICAO[1)] and ISO[2)]; ICAO represents the interests and competence of air transport companies, and approval by the ISO makes the standards legally binding by all ISO member states.

The picture on the left shows US passport 1500000035 type P issued on 1998-11-16 by the US National Passport Center to Mr Happy Traveler born 1956-09-

1) *International Civil Aviation Organization*

2) *International Organization for Standardization*

ISSN 0085-7130

16 in New Jersey. The lower part is machine readable to enhance validation when the passport is being used. However, the machine readable zone shows the passport's nationality, its number and the bearer's name.

The passport contains security mechanisms that protect it from copying and counterfeiting. One of the protecting mechanisms is the hidden RFID chip. The chip contains an electronic copy of the visual information, and it can be read by use of a reader that extracts the information through a radio frequency, and the system interprets its ICAO-formatted ID data.

The ID shows that Mr Happy Traveler has convinced the US National Passport Center about his name, place and date of birth, and passport 1500000035 type P was issued. Mr. Traveller might be the happy holder of passports from other countries as well, and these might even legally show other names.

IDs are issued by employers, educational institutions, banks, health services and more, but the passport issued by national police authorities is the only one that is strictly personal, issued and formatted according to international standards and agreements.

All issuers of IDs will build and maintain a register of produced IDs and their data, but this means that the selection of data and their quality will vary between them; it might even be difficult to decide whether different registers hold data about the same individual.

Driver licenses are frequently used to prove identity; these documents are issued to prove that the licensee has passed a test, and they are regularly issued to citizens from foreign, but not specified countries. It might therefore be a hard task to validate a citizen's identity by reading a person's driver license.

*e*ID

Providers of personalized services, such as banks and tax authorities, have learned that they can produce better and more efficient services on the Internet. However, they need to know that they are servicing the right person, especially if accounts and funding or information protected by privacy regulations are involved.

Clients have been given user names and passwords to access secure services on the Internet; in combination with other security mechanisms these have worked rather well. However, an increasing number of Internet services are providing users with a confusing number of user names and passwords. Electronic services are trusted and are operating growing values; this implies that access codes are interesting objects for theft, phishing[3)] and abuse. Obviously, better solutions than user name and password are needed, and *e*ID technology offers helpful improvements.

One common *e*ID is a small file with protecting mechanisms that inhibit changes to its content by making it useless when hampered with. And the issuer can revoke it for access to services he is supporting, but not to other ones. The *e*ID file can be stored on:

- a hard disk on the user's computer
- the issuer's central system
- memo stick
- smartcard
- the SIM card of a mobile phone,

and be activated by codes entered by the user. The codes can be traditional PINs, dynamic ones created by a calculator, distributed as an sms from the issuer on demand or biometrically, combinations of these are also in use.

For the user, a generic electronic ID – *e*ID – for use on the Internet could appear easy and convenient, but also threatening if it directly gives general access to his personal information in the systems of several service providers that he is using. It might give a feeling of using the same key for his home, his car, his office and his safe. The running and maintenance of a good *e*ID based system is an expensive and demanding task, and such a specialized service could certainly be welcomed by public service providers, if it meets their requirements on security and cost efficiency.

A strictly personal ID, like passport or citizen card, can be verified when used by comparing the picture and the person. For an *e*ID user on Internet this implies the use of biometric data representing the user's facial picture, finger prints, or iris to prove that he is the rightful user of the token[4)]. Technologies giving such solutions exist and are in use, but not in open systems.

Devices for the scanning of biometric data use algorithms that create character strings representing the result of the scan. So far, there are no standards for representation of biometric data, and the outcome

---

3) *Phishing is a criminal activity using social engineering techniques. Phishers attempt to fraudulently acquire sensitive information, such as user names, passwords and credit card details. (Wikipedia)*

4) *A security token (or sometimes a hardware token, authentication token or cryptographic token) may be a physical device that an authorized user of computer services is given to aid in authentication.*

ISSN 0085-7130©Telenor ASA 2007

from scans done by devices from different producers cannot yet be compared directly.

There are also regulatory challenges to open international use of biometric data when it comes to storing and re-use. Should such data be used for validation of *e*ID only, or could they be used for other purposes as well?

Biometric systems are in use in many systems around the world, and necessary standards, protocols and solutions for wider use can be foreseen in future; *e*ID with biometry will offer new opportunities.

The model:

Issuer	Certifies relation to	Subject

is valid for the electronic ID as well. The issuer, his certified relation to the subject and the subject itself are three important elements, and aspects of these will be addressed one by one:

Issuer

When an *e*ID certification is presented to a service provider, he can choose to accept it as is, or he can decide to verify its content. The latter will be of importance if the service involves or leads to financial transactions.

Verification of an *e*ID requires a processable identifying attribute of the issuer; it should be possible for a computer system to verify that the issuer is listed in a business register, and it should be possible to check an *e*ID with the issuer and originator of identifying attributes.

The service protocol OCSP[5)] specifies how data in an *e*ID certificate can be verified, and it is used within the user group of issuers. Data of an *e*ID will be compared with those registered by the issuer. Full and open use of the security mechanism would only be achieved if the actual public business register information of the *e*ID issuer could be verified as well. Technically this is possible, but standards and ascerting protocols for the use and exchange of identifying attributes for companies issuing *e*IDs have not yet been implemented.

The current situation can be exemplified by the issuer of *e*ID Telenor; Telenor ASA is the official name of the main company. Another identifying attribute is its business registration number 982463718; neither name nor number uniquely identifies the business register that should be used for real time verification of basic company data.

Certifies relation to

An *e*ID is a token in which the issuer certifies some registered pieces of information about the identified subject; the selection and quality of data reflects the purpose of the token.

Passports and citizen cards, and their electronic versions, document that the subject is listed in the country's passport and population registers.

Bank cards are issued to named persons, and in some countries they even display the person's picture and signature; the cards certify that the subject has a bank account. Most bank cards can be used for electronic payments by keying name, account number, expiry date and security code to prove that the user has the card in his hands. An increasing number of banks are introducing *e*ID systems to enhance security and reduce losses, but both old and new tokens work until expired or revoked. The bank does neither know nor care whether the named person is using the token as long as it has not been reported lost or stolen. The certified relationship is one of access to services that can be used by the identified subject and those authorized by him; the certified relationship can be delegated to other individuals under the responsibility of the subject.

In some countries smartcards with *e*ID are issued by taxation authorities to tax payers; they are issued to the tax paying subject, and frequently they will be used by others on behalf of the identified subject. The purpose of the smartcards and their *e*ID is to ease the communication between the authorities and the tax payers, and not to assure that the identified subject himself is the actual user.

IDs and *e*IDs are issued to persons, but they can certify relation to biological individuals or to tasks connected to individuals; the biological ones are unique whereas the task relation can be delegated by the subject to several others. For re-use of ID and *e*ID it is very important to distinguish between the two; an *e*ID that has been issued for task oriented relations should never be re-used for other task oriented or strictly personal relations without the clear and explicit acceptance and understanding of the identified subject that the token is being used by the subject alone as a strictly personal one. An issuer that accepts re-use of tokens produced by him should ask the subject whether the token will be used by him alone or not;

5) *Online Certificate Service Protocol*

ISSN 0085-7130

the answer should be recorded, coded and included in the certificate.

A passport is a strictly personal ID, and an *e*-passport will be an *e*ID that is issued as according to standards for a travel document. This means that name, place and date of birth, some biometric data, passport number, dates of issue and expiry will be monitored, used and maintained. There are on-going discussions about the biometric data, their use and representation, and further development is expected. The passport number is a unique identifying attribute specifying the relation between the issuer and the passport holder: national passport authorities certify that the described holder is registered as a citizen of the issuing country.

Passports are used as proof of identity when crossing borders, using 'manual' bank services, when requesting public health services abroad, when entering contracts, etc. Several foreign authorities use passport numbers as an identifying attribute for citizens from other countries, and passports certify the relationship between national authorities and the citizens in a standardized way.

Hence, it could appear convenient to prepare an *e*ID based on standards developed for passports for more generic use, especially for the citizens' use of domestic and foreign public services. The re-use of identifying attributes between public services is challenging protection of privacy, and different ones are in use for passports, health services and others.

A general *e*ID could, however, hold a set of identifying attributes and ask the identified person to select one when accessing services requesting authentication. The subject might decide to delegate to others to use some of the attributes and keep some private; the complexity calls for standards to be developed. Further, an *e*ID certifying several identifying attributes calls for complex monitoring, responsibility and maintenance, the issuer's registers will be sensitive and attract phishing, and considered a potential threat to the protection of privacy.

Subject

There is no general and unique definition of an ID; the statement is, of course, valid for the electronic version, the *e*ID, as well. For strictly personal ones the bearer will use biometric evidence to prove that he is the rightful subject. This means that the issuer has registered biometric data as identifying attributes for verification. In these cases the issuer certifies that the biological subject has been registered.

Biometric data, like fingerprints, facial characteristics, and the iris, can be scanned, computed and transformed into strings of characters; however, there are no open standards for biometric data, and such strings made by equipment from different producers cannot be compared. The visual examination of picture, checking of fingerprints and iris patterns cannot yet be transformed into computerized validation of *e*ID in open systems; the validation has to be performed by its issuer only.

Several Asian countries omit the problem by using 'match on card biometry'. The character string produced by scanning the finger of the identified person is stored in the card's chip, the card has a fingerprint reader and its chip produces a positive signal when the card is touched by the subject's matching finger. The signal can be received and interpreted by other systems with necessary software; it can be put into use in open systems if the issuer can be verified as trustworthy. But again, the implementation of open standards that enable the automated verification of the issuer, is awaited.

The acronym *e*ID covers a technology that protects a set of data; when biometric data are used for authentication, the eID should be considered strictly personal. The acronym is also used to identify tasks, services or obligations connected to an identified subject, such an *e*ID is likely to be used by delegated persons.

An *e*ID might be protected by the most sophisticated technology and direct delivery of token, but the user, the identified subject, might have chosen to share the non-biometric access codes with friends, family members or aides in good faith. *A non-biometric eID should only be trusted for authentication of its subject by person, if he has confirmed that he will be the only person to use it, and the issuer of the token can be verified.*

Trusting an *e*ID in Open and International Communication

Electronic communication is replacing mail, fax and personal appearance; needs for security have been satisfied case by case, and many solutions are in use. Banks have been using user names and passwords, but the more advanced *e*ID is being implemented. In some countries public authorities have decided to issue smartcards with *e*ID to serve their communication with citizens.

The running of *e*ID systems with issuing, monitoring, validation and revoking is an expensive business; at least € 25 per *e*ID annually. For the banking industry it has paid off, as the customers have reduced the

ISSN 0085-7130

work of the banks' employees. The *e*IDs from banks are being used frequently and are thereby well monitored and maintained.

None of the current versions can be subject to validation in an open and international world; the issuer is not identified with standardized attributes and needs to be verified by a service that knows him. Further, the *e*ID from a bank will not yet contain biometric information, and the identified subject might have decided to share access codes with others. The certificate will contain identifying codes issued by the bank; and it not sure that they will help other service providers' authentication of the subject.

But, the banking industry is investing heavily in *e*ID systems that are needed to secure its relations with customers to avoid losses and build efficient financial services. The organization, skills and systems can be used to produce general *e*ID services for open and international communication when the market is willing to use resources on necessary standardization and deployment.

Electronic chips in citizen cards and passports have been mentioned. Some countries have been issuing citizen cards with *e*ID for some years, but even these have not been designed for use in open and international systems.

So far there are very few international services calling for the use of high quality standardized *e*ID, and neither user communities nor service providers have yet been willing to spend resources on development and implementation of general *e*ID systems. However, the number and the use of electronic services are growing rapidly, and they are increasingly replacing old manual processes.

Postal services are being replaced by their electronic successor, and it can be expected that *e*ID systems will be used to secure the delivery of electronic mail in the near future. Postal mail services are international, like the electronic ones of today. It will therefore be interesting to see how *e*ID with extended functionality, like digital signature and encryption, will be used to create more secure and personal global electronic mail services.

The volumes represented by electronic mail services indicate that there is a huge potential for the use of electronic ID technology in open and international communication.

An increasing number of individuals are becoming involved in international trade and business; this leads to more and more foreigners listed in different registers that call for maintenance across borders. Today each register is using its own or selected identifying attributes for its subjects, and when searching or up-dating a register its used attributes must be presented by the subject in order to get access. It can be imagined that users have *e*IDs with many subject attributes, so that the right one can be found and presented for each service, or users will be using several *e*IDs, like they do today.

A general ID for life has never been defined; the possibility to track a biological person through his life time will be needed in some few cases: for heirs of property, pensions and medical purposes; in most cases it is enough to be able to authenticate the same person through a contractual relationship: When a person opens a bank account, the bank needs to know that they serve the same person or his delegates until it is closed.

Passports and citizens' cards are valid for a limited period of time. However, the issuer's register will link a user to earlier or later versions, but not necessarily for the user's whole life span.

In most cases it is enough to ensure that a communication involves a contractual part or his duly appointed representatives, and *e*ID is an accepted technological solution for electronic certificates; it does not prove any biological identity.

The *e*ID technology can be misused, and therefore the use of open, international and processable standards to specify the information within *e*ID will ease monitoring and probably complicate exploitation:

- Identifying attributes for the issuer; the standard ISO 6523 could be taken into use to specify the business register identificator of the issuer as soon as codes for the information data element have been established, and when protocols for maintenance, exchange, interpretation and validation have been agreed.

- Identifying attributes for subject need to be standardized, and protocols must be agreed on.

- Codification of usage; will the *e*ID be used by the specified subject, or could it be used by others? Standardized and processable codes for qualified and non-qualified *e*IDs need to be established.

The *e*ID technology has been accepted, but an agreement has not been reached on how to use it. Banks in several countries have taken it into use to serve their customers; the implementations, however, are not meeting the requirements for open and international

ISSN 0085-7130

use. Public services in some countries are accepting authentication of *e*ID issued by themselves or their contractors, but the tokens in use do not contain information according to open standards; they mainly serve the issuer's purposes.

New passports are containing chips with ID information according to ICAO and ISO standards, and the information can be secured in an *e*ID file for open use. The passport authorities will not possess the competence and resources to run systems for issuing and maintening *e*ID, but they could of course contract a specialized company to do the job, including validation when used on the Internet.

Conclusions

Trusting an *e*ID in open and international communication can in future only be achieved by applying standardized data to build information inside an *e*ID. Trust requires the possibility to verify and validate information; the first step will be to implement standards that open for automated verification of issuer and validation of the *e*ID itself and its content. Validation of the content requires use of standards and protocols for implementation and use of data for issuer, for subject, for security level and fields of usage. The validation of issuer will require automated identification of and access to business registers. Further, the *e*ID issuers need automated access to originators of identifying attributes that they are certifying. The use of public business registers will be of great importance. Standards for passports are established, and *e*ID systems certifying passport numbers could become a useful access tool for maintenance of several registers and communication across borders. Such usage calls for open access to national passport registers, directly or through dedicated service providers, like ICAO.

It should be emphasised that the acronym *e*ID refers to technology, and not to an identifying functionality for a biological individual; it is important to distinguish between tokens used by a specified human being and those used to access tasks and services connected to a person. The latter is likely to be used by aides, relatives and delegated ones as well. Open and international use of *e*ID can only be achieved when required specifications, standards and protocols are in place; those who want to benefit from such an achievement should specify, resource and launch the needed standardization processes.

Sverre Bauck obtained his PhD in biophysics at the University of Oslo (1974) and was Postdoc at the National Institutes of Health in Washington DC, USA (1976–77). He has been working with digital systems and solutions for nearly forty years, including digital electronics, measurement technology, software development, solutions, security, teaching and standardization on a national and international level. From 1986 to 1988 he worked in the Norwegian Directorate of Customs and Excises with the development and implementation of the electronic declaration system TVINN. The following four years he worked as a representative of the EFTA countries in the Western European EDIFACT Board Secretariat in the European Commission coordinating the development of EDIFACT standards. He has since then been employed by Statskonsult and ErgoGroup working with analyses and solutions for public and private clients. In 2004 he joined the Brønnøysund Register Centre with a focus on the public sector's dialogue system Altinn and is now working to analyze directions for the development of the register's electronic services.

email: sverre.bauck@brreg.no

ISSN 0085-7130

Building a Federated Identity for Education: Feide

INGRID MELVE, ANDREAS ÅKRE SOLBERG

Ingrid Melve is Chief Technology Officer in UNINETT

Andreas Åkre Solberg is Scientist in UNINETT

Feide is Federated Electronic Identity Management for Norwegian education. A shared login service (Moria, the Feide Identity Provider) provides authentication and information profiles for use in a variety of web services. Moria was introduced in 2003 based on a proprietary protocol, but is now moved to SAML2.0 as the first educational federation internationally. Our reasons for introducing SAML as the interface includes integration support, multivendor environments and traffic exchange with other federations. International cross-federations are investigated at the policy and technical levels.

Identity management procedures are in place for educational institutions across Norway, and before adding end users to Feide each school, college and university go through a review of procedures and technology. Feide reduces the number of passwords to remember, moves authentication out of each service (thus reducing the password exposure), and provides reliable data about the end user (including roles) for use in personalization and authorization.

1 Introduction

Feide is the Federated Identity and Access Management solution for Norwegian education. Some of the goals are to lower the cost of user administration, lower the threshold for deployment of new services locally and nationally, raise the data quality for personal information, and focus on security and privacy. Feide is an important strategic tool for integration of IT-systems in the educational sector. Feide aims to simplify everyday life by giving users one single password to remember, to facilitate privacy enforcement, and to remove lock-in from vendors who control user information.

Technically, Feide consists of a distributed data storage at each host organization, a federated login service, and integration modules at each service provider. Contractual relationships govern both the information model, with an extensive set of attributes available for each end user, and the release of information to service providers.

2 Identity Management in Education

School owners, universities and colleges manage groups of Feide users: Username and password assignment, as well as handling of attribute values for each user affiliated with the organization. Feide itself stores neither authentication information nor user attributes, but assumes that the host organization will handle it.

The host organization is contractually obliged to fulfill a number of requirements to enable Feide to perform the authentication. Information about users must be stored in a secured system, and must be provided to Feide in a standard format. Otherwise, Feide could neither perform the authentication nor forward information about the users to service providers.

The information stores (LDAP directories) of the host organizations may, as a whole, be viewed as a distributed database, with Feide as the central controller. This choice, rather than the alternative of a single, centralized database with all information managed by Feide, is supported by several arguments:

- Local maintenance of user data is important. Updates should be done as close to the authoritative source of data as possible.
- Centralized data storage puts all the eggs in the same basket, exposing all user data. A centralized login service may expose user data if it suffers a break-in, but for a shorter period, and only for the users actively logging in at the time.
- Correct use of user attributes depends on well defined semantics. This is ensured through the Feide LDAP scheme, defining value sets, syntax and semantics for all attributes.
- Service providers should not have to relate directly to each individual host organization. They should be relieved of handling credentials such as passwords, but leave this to a centralized federated login facility.

Because authentication directly depends on information from host organizations, keeping the information as up-to-date as possible is essential. In other words, all modifications, additions and deletions in the authoritative sources must be reflected in the data made available to Feide, and which may be forwarded by Feide.

ISSN 0085-7130

2.1 Higher Education

Universities and university colleges have had access to Feide login since 2003, with 76 % of the 250,000 users participating in Feide by July 2007. Deployment has been dependent on campus identity management upgrades and investment. Local identity management is implemented technically by a combination of software for business logic and LDAP directories for storing information. Feide started preliminary work in 2000, investigating local identity management and technologies for sharing login information. Initial work was done on a PKI based solution, and it was decided to move on with a username/password based solution while PKI matured in the marketplace.

In 2003 the first five services were Feide enabled, using a proprietary SOAP-based protocol for authentication and attribute release. Late 2005 the decision was made to move to SAML2.0, and the SAML login went operational spring 2007 when there was 34 services to move from the old implementation. Services available with Feide login range from administrative applications, *e*-library, portals, wikis, mailing lists, electronic voting systems, software licenses and services for researchers, to smaller applications for a limited constituency. Higher education has a long standing tradition for collaboration around national scale IT services. Feide is used extensivly in national shared services. In August 2007 there were 37 applications available with federated access.

2.2 Schools and Feide

In 2006, it was decided to make Feide available to all Norwegian schools by 2009. The school sector covers approximately a million users, teachers and pupils, with an additional 1.2 million parents who should be able to communicate with the school solutions. Interoperability with the Norwegian public electronic identity portal, *e*ID, is important in order to facilitate parent login without requiring each individual school owner to maintain user information about the parents. Upper secondary schools are phasing in Feide this fall, with students in Østfold and Rogaland as the first users. Lower secondary ("ungdomsskole") and primary schools are in the pilot stage and will have Feide available on a large scale by 2009. The main work items are Feide-enabling service providers and putting good identity management practice into place at each school.

3 The Feide Information Model

The information flow in Feide runs from host organizations, through Feide, to service providers. Any information flow from service providers to host

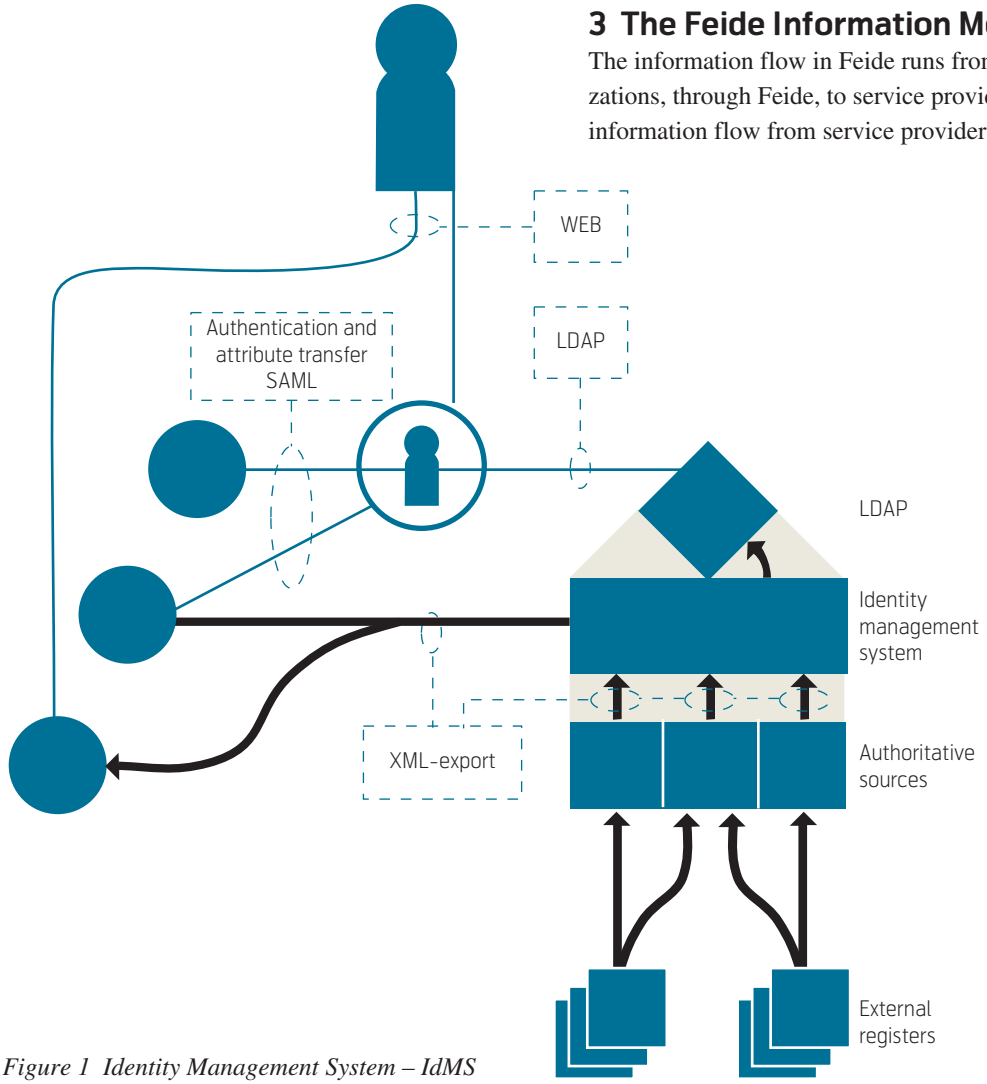


Figure 1 Identity Management System – IdMS

ISSN 0085-7130©Telenor ASA 2007

organizations bypasses Feide and is not defined in this architecture.

To ensure that information is structured, correct and up-to-date at any time, Feide requires each host organization to implement an automated information management system. An important component in this structure is an automatic *identity management system*, IdMS, as shown in Figure 1.

An IdMS handles information, retrieved from authoritative sources, regarding the organization's users. Examples are contact information, affiliation type (student, employee ...), authentication information etc.

Feide requires each host organization to provide a standard set of attribute values describing each of the organization's users, in an LDAP directory containing copies of the relevant parts of the information handled by the IdMS of the organization. Semantics and structure of the information are specified in the Feide LDAP scheme [norEdu*], which is based on international cooperation and current de-facto standards: the eduPerson/eduOrg-schemes [eduPerson, eduOrg].

The LDAP directory has the following properties:

- The LDAP directory is employed for user authentication. Current use is with passwords, tests have been made with one-time passwords and certificates.
- The LDAP directory contains descriptive attributes for all authenticated users.
- The information in the LDAP directory is, as far as possible, kept up-to-date and correct at all times from authoritative sources. The organization must define one source system as authoritative for each person attribute. All other occurrences of the attribute value are considered copies. If the correctness of the value is questioned, the value from the authoritative source takes precedence. Detailed requirements are indicated in guidelines specified in the Feide contracts.
- The login service retrieves user information from the LDAP directory of the user's host organization.

The following technical requirements must be satisfied for all LDAP directories:

- The communication between the directory and Feide is reliable and protected against eavesdropping.
- The directory is available through the standard LDAP protocol. The login service must be informed about the DN ("Distinguished Name") of the root of the Feide part of the directory tree, and must be given sufficient access rights to search for a user's DN based on his Feide name.
- The directory information is supplied from a campus identity management system (an IdMS) of the organization.

Feide defines attributes for home organizations and for each user. Of the 47 attributes defined in Feide for users, nine are mandatory. Two are of special interest, the *Feide name* and the *National Identity Number* (NIN, "fødselsnummer"). The Feide name is used as an identifier, whereas the NIN is used for internal consistency within each organization, leveraging off existing registration. Attributes are tagged with three levels of confidentiality, availability and consistence requirements. A service provider may only access those attributes it needs to be able to deliver its intended functionality, and reviews are done before releasing attributes tagged with the highest level of confidentiality to the service. In the login window, the end user may inspect which attributes are requested by each service, and may terminate the session without logging in.

The user attributes are stored in an *LDAP directory* of the host organization the user belongs to. When a user is authenticated, Feide retrieves information from the host organization's LDAP directory and forwards the selection of attributes that the service is entitled to according to the contract with Feide. This relieves the service from managing information about individual users: Rather than maintaining its own attribute store, the service asks Feide to provide the information on demand. For example, a library need not keep track of the email addresses of its users; when a user places a book reservation, the library is informed by Feide where to send a notification (i.e. the user's email address) when the book becomes available.

Becoming a host organization requires Feide approval. The following steps must be carried out:

- Install an IdMS and an LDAP directory capable of delivering user attributes to Feide according to the stated specifications and requirements. Both open source systems and commercial systems from various vendors are available.
- Clean up the data in the source systems, e.g. remove outdated and duplicate entries, and verify that essential attributes are semantically consistent.

ISSN 0085-7130

- Verify that procedures for management of personal data consistently ensure high data quality.

- Ensure that user password assignment procedures generate passwords of sufficient strength according to commonly accepted criteria. Ensure that all usernames are unique and are assigned according to rules which will be unchanged for some time. An IdMS may be of great help for these tasks.

- Verify that both authoritative data about each user, such as name and address, and generated data, such as username and encrypted password, are available in the LDAP directory.

- Apply to Feide to become a host organization, and enclose documentation showing that the above points are taken care of.

- Sign a contract with Feide. If the organization also plans to offer Feide services, this is covered by the same contract.

- Make the LDAP directory available to the Feide login service.

A service receiving information through Feide will always receive fully updated information. However, some services have a need for more information than what is available through Feide, and must locally maintain their own supplementary information directories. Usually, it is beneficial to use the local directory for supplementary attributes only, avoiding duplicate storing of information that is available through Feide. This ensures that the service does not rely on outdated information.

Feide fully controls which information is forwarded from host organization to service provider. Contracts with each individual service provider limit the user attributes made available: If a service provider cannot demonstrate a need for knowing the identity of the person requesting the service, no information about that person is revealed. Feide may still indicate e.g. that the user is an (unidentified) student. Information is revealed only when required to perform the service, and in agreement with the host organizations managing this information.

Feide stores no user attributes beyond the temporary storing necessary while a Feide session is active. The login service maintains a persistent directory of federation keys for each service a user is federated with.

4 Feide Architecture

Feide, through its login service Moria, provided in Feide, mediates information about Feide users to service providers.

Confirming that a user is the person he claims to be is called *authenticating* that Feide user. A person using Feide services is authenticated once by *logging in* through Moria at the start of a working session. Throughout the session, Feide will attest the identity of the user to the various service providers; Feide is a trusted *third party*.

Feide also communicates reliable information, *user attributes*, regarding authenticated Feide users. A service may therefore be relieved of the tasks of managing basic user data, and of keeping these data up to date. Authentication confirms the user's identity independent of the rights, the *authorization* that user is entitled to.

The service may determine the user's authorization based on user attributes communicated by Feide. The authorization may be directly given by the attribute values, e.g. the user's organizational affiliation or kind of affiliation (student, employee). The service may, if it knows the identity of the user, manage its own authorization information at an individual level.

Feide offers authentication and user information based on web protocols, and may be used by services offered across the web. The technical solutions employed in the current implementation are not adapted to systems with user interfaces based on other technologies.

4.1 Feide Requirements

The initial Feide requirements were:

- Minimal information release. Users control the release of information, home organizations set the policy for which information is available, and the service providers get updated information. Information transfer is encrypted to stop eavesdroppers. Feide should be a tool for enhancing privacy.

- Organization centric identity, scaling to the entire education population and to thousands of services.

- Distributed authentication; each organization has its own authentication point, but with central shared login service.

- Information model with release of attributes; affiliation (student, staff, faculty) is important.

ISSN 0085-7130©Telenor ASA 2007

- Support for multiple types of credentials; start with username/password. Single Sign On is not a goal.

- Security at a reasonable level.

One original requirement states that Single Sign-On (SSO) was not a goal. This was changed shortly before the launch of the first version of a shared login service in May 2003, when the service providers strongly advocated SSO. The reason for leaving SSO out in the first phase was the complex security issues that arise with SSO.

Requirements that have been added after the initial design:

- Single Sign On (SSO) and Single Log Out (SLO).

- Interoperability with MinSide, using the SAML2.0 protocol.

- Interoperability with international research and higher educational federations.

4.2 Federation

Making a Feide user known to a service provider is called to *federate* the user with the service. In the primary application environments of Feide, the user is first established as a Feide user through a host organization, and later federated with a service provider account. If the service has no need to recognize the Feide user from one session to another, a *one-time federation* may be automatically established at the start of the session, to be dissolved at the end of the session.

When desired or required, information about several users may be transferred in advance from a host organization to a service provider by *provisioning*: In a single operation, the host organization uploads information about e.g. all new students this semester, to the service. The upload is performed independently of the Feide services. When the user first connects to the service, Feide conveys sufficient information about the user to allow the service to identify the appropriate (pre-registered) account for federation with the user.

Provisioning is of particular interest when host organization and service provider are the same organization (i.e. local services), and the service requires information not defined in the Feide LDAP scheme. Standardization of information flow for education (PIFU) is proposed as Norwegian standard NS-4710.

4.3 Trust Management

The Feide network of trust is expressed through a set of agreements, contracts, implementations and guidelines. The term “network of trust” is an indication that a certain level of mutual confidence between the actors is required for the federation to work across organizational borders. The current implementation of Feide defines a single level of trust.

Feide users are managed by their host organizations. Feide requires that the users are held responsible for all their actions when accessing computer systems and services, and that agreed rules for acceptable use are enforced.

Requirement for protection of information and privacy is regulated by Norwegian law. When Feide authentication is used with host organizations or service providers outside of Norway, similar laws in other countries may apply. Feide has been developed to satisfy the requirements in the Norwegian Personal Data Act [POL], with particular attention to security and protection of information about individuals.

Employees, students and others belonging to an organization have a trust relationship to their host organization. Modifications of this relationship over time are managed by the host organization. Because Feide’s contact with users goes through host organizations, requirements and obligations to Feide will affect users indirectly only.

The trust relationship between users and services is an implicit relationship that users may choose to accept by logging in to Feide. The trust is based on the security mechanisms of Feide, the contract between the service provider and the host organization, Feide’s agreements with these two, and between the user and his host organization.

4.4 Federated Identity Management

Internet has become part of our everyday lives and people address dozens of Internet sites every day. Internet has developed from providing static text pages to providing dynamic, interactive and personalized content. To provide personalization, web sites must know the identity and characteristics of the user. Also, Internet web sites are no longer a one-way channel: Users may visit established web sites to publish and share information. Communication with web sites is often bound to a personal account. These accounts differ in degree of anonymity and security, but are usually protected by username and password.

Without a system for federated identity management the user may have to remember a dozen username/password combinations, or expose the same password

ISSN 0085-7130

to all the web sites. Federated identity management leaves the responsibility of authenticating a user to a separate entity, referred to as an Identity Provider (IdP). We will refer to the web site being accessed as the Service Provider (SP). Standards for federated identity management usually define protocols for a three-tier exchange of security assertions between the SP, the IdP and the user. The user is here represented with a web browser, sometimes referred to as an HTTP user agent.

The *Security Assertion Markup Language* (SAML) version 1.0 is a standard which defines a protocol between the user, the SP and the IdP. SAML was adopted by the OASIS working group in November 2002. In May 2003, OASIS published a revised SAML, version 1.1. This update contained minor changes outlined in [saml1diff].

Even though a common standard for federated identity management would be great, further development of SAML 1.1 grew in separate directions:

In 2003 Microsoft and IBM together published WS-Federation [understandingws], which makes use of components from SAML 1.1 but adds features such as support for communication between web services.

Liberty Alliance [liberty] was formed in 2001 by 30 organizations aiming to standardize protocols for federated identity management. They saw limitations in the SAML 1.1 standard, and introduced several new features in their identity framework ID-FF [idff12specs]. One of the most requested features from SAML 1.1 was the *Authentication Request*: In SAML 1.1, the user starts out at the IdP for authentication, before proceeding to the SP. With the ID-FF introduction of the Authentication Request, the user goes directly to the SP, and is redirected to the IdP if the SP requires authentication. Another concept missing in SAML 1.1 is logout. ID-FF introduced *Single Logout* (SLO), which allows users to log out from all services within an active session by a single click.

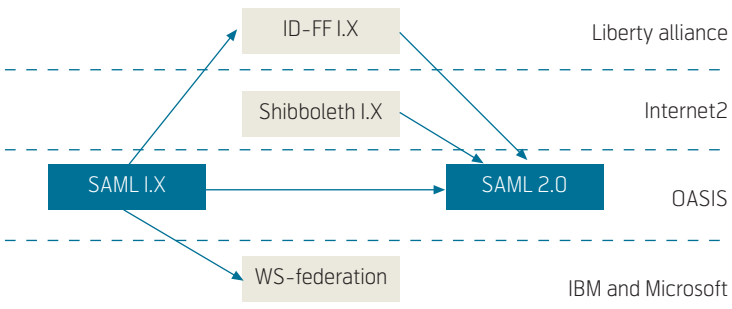


Figure 2 Identity and Federation Standards

Liberty Alliance also introduced *federated identifiers*. Federated identifiers are opaque identifiers, unique to each SP. One of their features, relating to user privacy, is that due to the lack of a global identifier, consolidating data across different SPs about a given user is not feasible.

Another offspring from SAML 1.1 is *Shibboleth* from Internet2, both a protocol extension to SAML 1.1 and a software product for identity management, implementing SP and IdP functionality. The most significant extension is the addition of the Authentication Request. The Authentication Request in Shibboleth is much simpler and not compatible with the request defined in ID-FF.

These standards developed independently and revisions were published. To curb diversification, Liberty Alliance, Shibboleth and OASIS agreed to develop a common standard, SAML 2.0, and to discontinue both ID-FF and the Shibboleth protocol. No new updates of the ID-FF standard are going to be published, and the Shibboleth 2.0 software package will natively support SAML 2.0.

SAML 2.0 adds an authentication request message and includes important features from ID-FF, such as federated identifiers and single logout.

SAML 2.0 was standardized [saml2] through the OASIS standardization body and was published March 2005. Feide evaluated login solutions in 2005 and concluded that SAML2.0 met the requirements for integration support, multi vendor environments and traffic exchange with other federations; as well as the requirements for security, privacy protection, collaboration and support for our information model.

5 A Feide Login Scenario

The following steps are the ones recognized by the user performing a successful login to a service using Feide. The syntax and semantics of the protocol messages sent between the steps will be outlined later in the article.

1 The user attempts to open a web page for the service he wants to activate.

2 The service makes an authentication request to Feide, and Feide displays a login form to the user. The user fills in his Feide username, organization and password in the login form and returns it to Feide. Feide verifies the username/password combination towards the LDAP of the selected organization.

ISSN 0085-7130©Telenor ASA 2007

3 If the user is successfully authenticated at Feide, the user is sent back to the service with a signed security assertion with the attributes that are configured to be released for that SP.

4 The service validates the security assertion sent by Feide, and determines whether the authenticated user should have access to the service. Next, the service generates contents, and sends it back to the service. The user now has an active authenticated session at the service and can access other pages at this service without going through Feide.

If a user has already specified his Feide username and password when activating another service, the user has an SSO (Single Sign-On) session at Feide, and step 2 is skipped. The user will not notice the redirection to Feide and will experience being automatically logged into the service.

This login scenario is the same with our profile of SAML2.0 and with the proprietary interface (Moria2) developed earlier in Feide.

5.1 The SAML 2.0 Standard Suite

The SAML 2.0 standard is a set of building blocks. Looking at Figure 4, at the bottom we have the *security assertions*, which are the core information elements in SAML 2.0. A security assertion is a block of one or more statements regarding authentication, attributes or authorization. Feide uses authentication and attribute statements only, because authorization is performed locally at the service and not transported via Feide.

In the communication between the SP and the IdP, assertions are enclosed in various *protocol messages* such as requests, queries and responses. Feide uses the Authentication Request and the Authentication Response.

Security assertions and protocol messages are XML documents. SAML 2.0 specifies the format and usage of these messages in the form text and XML schemas [xmlschema]. The standard defines alternate ways of transporting these messages between the SP and the IdP, both directly and via the user's web browser, specified in the SAML 2.0 *bindings* [saml2binding].

SAML 2.0 also specifies different *profiles*, i.e. sets of protocol messages that can be exchanged to fit into a useful scenario. For example, the WebSSO (Web Single Sign-On) profile specifies how an SP can issue an Authentication Request and how the IdP should respond with an Authentication Response to obtain Web Single Sign-On functionality. The *attribute* profile is another, defining how the SP can request additional attributes from the IdP after authentication.

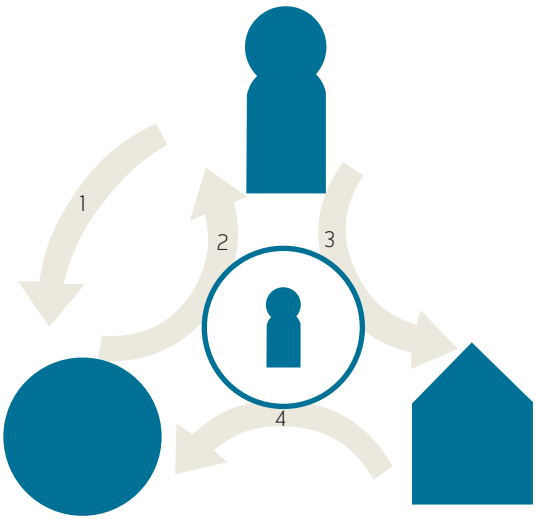


Figure 3 Feide login scenario

5.2 Web Sessions

To fully understand SAML 2.0 it is important to understand sessions on the web. At the introduction of HTTP [http], the web was completely stateless: A web site could not associate HTTP requests from a user with prior requests from the same user, except by looking at the IP address. To approach a stateful web, HTTP introduced *cookies* in [rfc2965].

Using cookies, the web site can send a **`Set-Cookie`** header in the HTTP response to the user's browser.

```
Set-Cookie: UserID=janedoe; path=/
```

The cookie is an attribute name-value pair bound to the domain name of the web site. When the web site

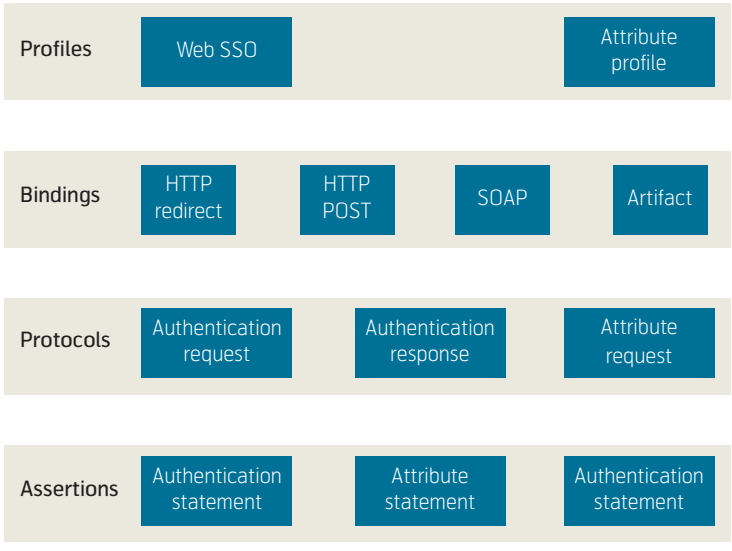


Figure 4 SAML 2.0 standard

ISSN 0085-7130

sends a valid cookie to the browser, the browser will return the same cookie back to the web site for every subsequent request. That way the web site can keep track of the user and create sessions with authentication information.

```
Cookie: UserID=janedoe
```

Rather than sending voluminous session data to the browser, they may be managed at the server side, associated with a session identifier. By sending the session identifier to the browser as a cookie, the web site can keep track of the user session, and retrieve and store all the necessary data between requests. Here is an example of setting a session ID as a cookie:

```
Set-Cookie: SESSIONID=
5ba808fa4795c150fadad6880c77; path=/
```

This way, session data is never exposed to the user. The protocol exchange is also more efficient, as the session ID usually is more compact than the data itself.

In SAML 2.0, local sessions are usually created at the service provider after receiving a valid *security assertion* from the identity provider. The SP will remember the user (keep the session alive) until explicitly terminated by a logout, or on session time out after a configured session lifetime.

5.3 SAML 2.0 Message Flow

To understand the basics of SAML 2.0, we will look into the message flow between the user, the SP and the IdP in a Web Single Sign-On scenario.

Figure 5 shows the message flow in a scenario where a user is trying to access a web site (SP 1) and is authenticated via the IdP (Feide). Next, the user tries to access another web site (SP 2) and is automatically logged in using the SAML 2.0 Web Single Sign-On profile. Dotted lines indicate messages that are part of the SAML 2.0 protocol.

We will go through the steps in detail:

Step 1) The user accesses a web site, SP 1: His web browser sends an HTTP GET command [http].

Step 2) As the web site contains protected content, it requires user authentication. The service delegates the actual authentication to an IdP, with which it has a trust relationship. Because the service has no prior established authenticated session with the user, it will issue an *SAML 2.0 Authentication Request*.

Step 3) The SAML 2.0 authentication request is sent to the IdP via the user, transported by HTTP, using one of several alternate ways of performing SAML 2.0 message redirection. This is further explained in section 5.8, *SAML 2.0 Bindings*.

Step 4) The IdP recognizes no prior authenticated session with the user and requests the user to authenticate. Note that SAML 2.0 protocol does *not* define the actual authentication method; this is up to the IdP (and in the password-based login in Feide is delegated to the local LDAP directory of the user's home organization). In the scenario above, the IdP sends an XHTML form to the user as an HTTP response to the authentication request. In Feide, users authenticate themselves by username and password, but the IdP could as well require authentication, e.g. by an X.509 client certificate.

Step 5) The user's browser displays the web page with the login form, and the user submits the username and password (often referred to as *credentials*). The Feide IdP also requests the user to select her educational institution from a list of recognized institutions.

Step 6) The IdP then validates the credentials. The Feide IdP connects to the LDAP owned by the educational institution selected by the user, and retrieves attributes for this user. If validation is successful, the IdP then establishes a local authenticated session with the user. Next time she enters the IdP, she is remembered and need not re-enter the credentials – that is what Single Sign-On is all about.

Step 7) The IdP issues an *SAML 2.0 Authentication Response*; a signed message asserting that the user is authenticated, optionally including some attributes about the user. The authentication response is sent via the user to the service (SP 1).

Step 8) SP 1 parses the response from the IdP. Because of the already established trust relation between the IdP and SP 1, it trusts the asserted identity included in the response, verified by the digital signature.

Step 9) After some surfing, the user moves on to another service, SP 2, by following a web link or by typing in a new URL. The browser sends an HTTP GET request to the SP 2 web site.

Step 10) Similar to step 2: SP 2 requires authentication, but has not yet established a local session with the user. SP 2 issues a SAML 2.0 Authentication Request.

ISSN 0085-7130©Telenor ASA 2007

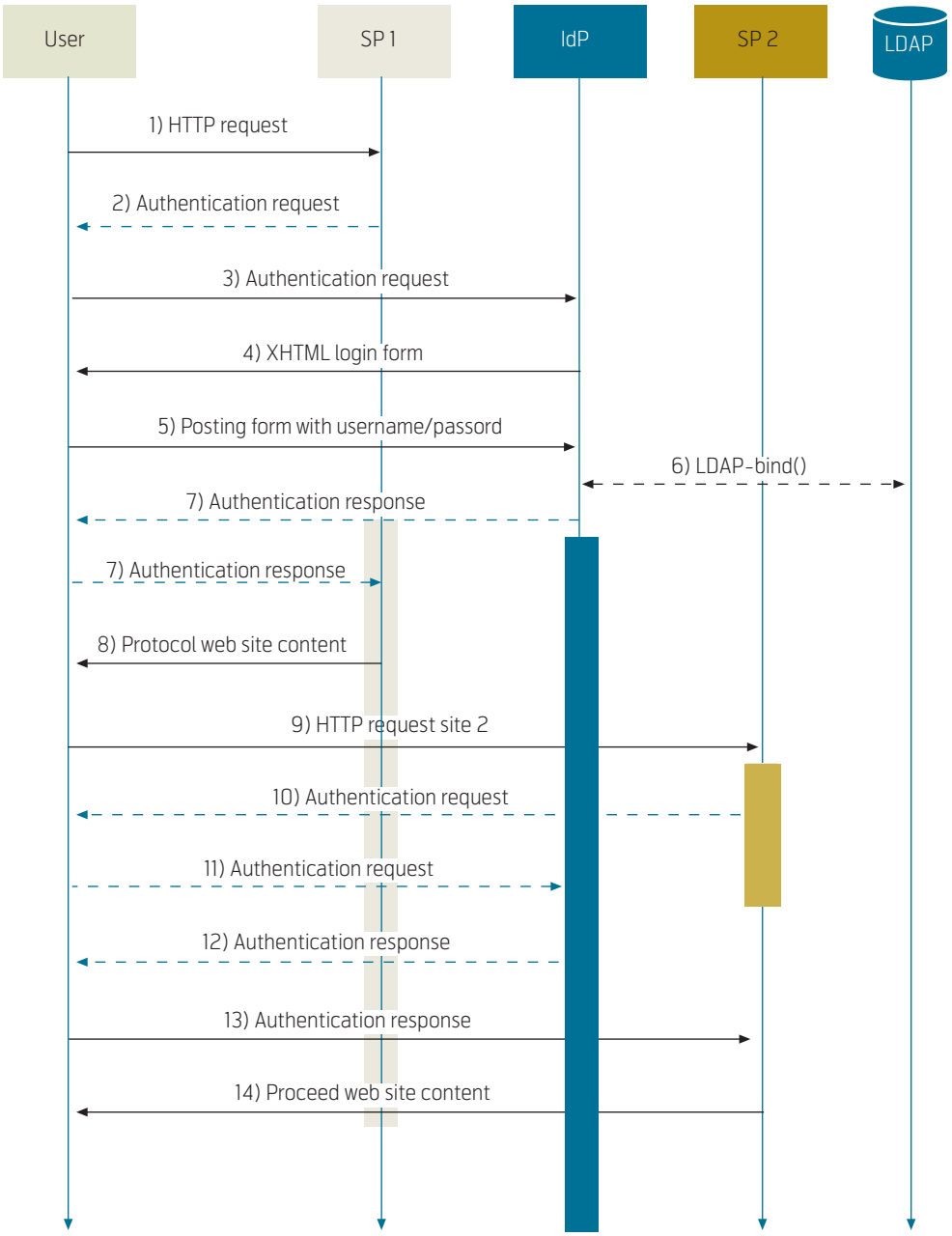


Figure 5 SAML 2.0 message flow

Step 11) The request is sent to the IdP via the user by HTTP redirection.

Step 12) This time the IdP recognizes the established session with the user and knows that this user is already authenticated. Instead of requesting authentication from the user again, the IdP immediately issues an Authentication Response.

Step 13) This response is sent to SP 2 via the user by HTTP redirection.

Step 14) SP 2 validates the asserted identity, and if the user's identity entitles the user to access the service content, the service delivers the content to the user.

Note that the IdP only issues assertions about authentication and attributes. Access control or *authorization* is left to the SP, but the SP may use the attributes received from the IdP to make its decision about authorization.

Which attributes are sent from the IdP to the SP is predefined in the IdP configuration. According to Shibboleth terminology this configuration is called *Attribute Release Policy*. The user's privacy is important to Feide, so the service will not have access to other attributes than those strictly needed. In some cases, the identity of the authenticated user is hidden to the service, and only a few anonymous attributes are revealed. For example, a web page presenting content intended for students only will not be

ISSN 0085-7130

informed about the user's identity, but will receive the attribute **affiliation**, which should have a value of **student** if access to the page is to be granted.

5.4 The Authentication Request

When an SP receives a request from a user who is not yet associated with a local authenticated session (identified by cookies), the SP may issue an authentication request.

An example of an SAML 2.0 Authentication Request is shown in Figure 6. In the request, we see two XML namespaces [xmlnamespace], both part of the SAML 2.0 specification: One refers to the assertion part, the other to the protocol part of SAML 2.0.

The authentication request element is represented by the **AuthnRequest** element. It contains an ID which is later used in the response to map the response to this particular request. The ID also protects against replay attacks. The element also contains information about the SAML version and the time the request was issued.

The **ProtocolBinding** attribute indicates to the IdP the preferred binding of the SP for the authentication response.

All SPs and IdPs in SAML 2.0 have an identifier called the *Entity ID*. The **Issuer** element of the request contains the Entity ID of the SP. The IdP uses the Entity ID to look up trust information and metadata for the particular SP.

In SAML 2.0 terminology the identifier representing the user's identity is called *Name Identifier* (or NameID). SAML 2.0 supports multiple identifier kinds; two predefined kinds are **transient** and **persistent**. The former works like the opaque temporary handle in Shibboleth, a new one is generated for each session on each SP. This makes sure that no information about the user's true identity is provided at the maximum privacy level. Required data about the user can be provided as attributes. The latter, **persistent**, defines a permanent handle that is the same each time a user logs in to a given service, but different for different services. This type of identifier is compatible with the federated identifiers introduced by Liberty Alliance in ID-FF. There are also other NameID formats, such as email-address, and SAML even lets you define your own if required.

SAML 2.0 has a concept called *authentication context classes* for defining different levels of authentication. Assume e.g. an IdP that supports authentication with both password and with software-PKI and that software-PKI is considered more secure. Some services may request a minimum authentication class of password, while others may require that the user is using software-PKI. In Figure 6, the request specifies that the authentication context must be exactly the password class.

5.5 SAML 2.0 Assertions

When an IdP receives an authentication request, and the user has been authenticated, the IdP creates *assertions* about the user to be sent to the SP.

An assertion contains one or more statements. SAML defines three different statement types:

- Authentication statements
- Attribute statements
- Authorization decision statements.

When an SP requests authentication, Feide generates one authentication statement and one attribute statement. The assertion element has the attributes as seen in Figure 7.

The **ID** attribute of the assertion is used to refer to the correct assertion for the contained XML signature [xmlsig]. The assertion also contains version information and the time the assertion was issued.

The **Issuer** element is similar to the one in the request, and identifies the IdP by its Entity ID.

```
1  <samlp:AuthnRequest
2     xmlns:samlp="urn:oasis:names:tc:SAML:2.0:protocol"
3     xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion"
4     ID="_4e498f11858f139ddb491f2f1303658b4c4cc6e4b6" Version="2.0"
5     IssueInstant="2007-08-17T10:51:31Z"
6     ProtocolBinding="urn:oasis:names:tc:SAML:2.0:bindings:HTTP-POST">
7     <saml:Issuer>https://testsp.feide.no</saml:Issuer>
8     <samlp:NameIDPolicy AllowCreate="true"
9        Format="urn:oasis:names:tc:SAML:2.0:nameid-format:transient"/>
10    <samlp:RequestedAuthnContext Comparison="exact">
11       <saml:AuthnContextClassRef
12          >urn:oasis:names:tc:SAML:2.0:ac:classes:PasswordProtectedTransport</saml:AuthnContextClassRef>
13    </samlp:RequestedAuthnContext>
14 </samlp:AuthnRequest>
```

Figure 6 SAML 2.0 Authentication Request

ISSN 0085-7130©Telenor ASA 2007

```
<saml:Assertion Version="2.0" ID="s2010959d600bde4eb554901f23737cca324b984fe"
   IssueInstant="2007-08-17T11:49:50Z">

   <saml:Issuer>sam.feide.no</saml:Issuer>

   <Signature xmlns="http://www.w3.org/2000/09/xmldsig#">[...]</Signature>

```

*Figure 7 The **Assertion** element*

```
<saml:Subject>
   <saml:NameID Format="urn:oasis:names:tc:SAML:2.0:nameid-format:transient"
      >5/CI82rUJ6o53pR9Mo78BKrvGGmx</saml:NameID>
   <saml:SubjectConfirmation Method="urn:oasis:names:tc:SAML:2.0:cm:bearer">
      <saml:SubjectConfirmationData NotOnOrAfter="2007-08-17T19:49:50Z"
         InResponseTo="_4e498f11858f139ddb491f2f1303658b4c4cc6e4b6"
         Recipient="https://sptest.feide.no/saml2/AssertionConsumerService"/>
   </saml:SubjectConfirmation>
</saml:Subject>
```

*Figure 8 The **Subject** of an **Assertion***

```
<saml:Conditions
   NotBefore="2007-08-17T11:39:50Z"
   NotOnOrAfter="2007-08-17T19:49:50Z">
   <saml:AudienceRestriction>
      <saml:Audience>https://sptest.feide.no</saml:Audience>
   </saml:AudienceRestriction>
</saml:Conditions>
```

*Figure 9 The **Conditions** of an **Assertion***

The **`Signature`** digitally signs the assertion. SAML 2.0 allows two different approaches to signatures: You can sign each assertion individually (as shown in Figure 7), or you can sign the entire Authentication Response.

The **`Subject`** element of the assertion identifies the name identifier of the authenticated user. In Figure 8, a transient NameID is given. The **`SubjectConfirmation`** element tells the SP how it can know that it is communicating with the user that this assertion represents. The **`bearer`** confirmation method is used in Web Single-Sign-On, meaning that since the message is sent via the user, the user that presents the assertion is the one the assertion refers to. The **`SubjectConfirmationData`** maps the subject to the ID of the authentication request and includes the URL to which the assertion will be sent.

The **`Conditions`** element adds restrictions to the validity of the assertion. In Figure 9, two timestamps indicate a time frame when the assertion is valid, and **`AudienceRestriction`** identifies the Entity ID of the SP. An SP will only accept an assertion where its own Entity ID is included in the audience.

Following the elements above, the assertion can contain one or more statements. In the response to an Authentication Request, the IdP should add an Authentication Statement, Figure 10.

The **`AuthnInstant`** attribute says exactly when the IdP accepted the user's credentials. The SessionIndex attribute, a new feature in SAML 2.0, allows the SP and IdP to exchange SAML messages regarding a user directly, rather than by redirection via the user's browser. The IdP can identify the correct session through the **`SessionIndex`** without the need for a cookie sent by the browser. An example might be for the IdP to send logout requests to the SP with SOAP [SOAP].

There are two common ways of issuing attributes from the IdP to the SP: With *attribute push*, an attribute statement is included in the authentication response. The alternative is to let the SP request

```
<saml:AuthnStatement AuthnInstant="2007-08-17T11:49:50Z"
   SessionIndex="s22c0d60769d7eb12475d3210935a07daa1710d501">
   <saml:AuthnContext>
      <saml:AuthnContextClassRef
         >urn:oasis:names:tc:SAML:2.0:ac:classes:PasswordProtectedTransport</saml:AuthnContextClassRef>
   </saml:AuthnContext>
</saml:AuthnStatement>
```

Figure 10 The Authentication Statement

ISSN 0085-7130

```
<saml:AttributeStatement>
   <saml:Attribute Name="givenName">
      <saml:AttributeValue>Andreas</saml:AttributeValue>
   </saml:Attribute>
   <saml:Attribute Name="eduPersonPrincipalName">
      <saml:AttributeValue>andreas@uninett.no</saml:AttributeValue>
   </saml:Attribute>
   <saml:Attribute Name="eduPersonAffiliation">
      <saml:AttributeValue>employee</saml:AttributeValue>
   </saml:Attribute>
</saml:AttributeStatement>
```

Figure 11 The Attribute Statement

attributes separately using the SAML 2.0 *attribute profile*, and a separate request / response is used to retrieve attributes.

The example in Figure 11 illustrates attribute push: An attribute statement is included in the assertion. Three attributes are included: **givenName** – user's name, **eduPersonPrincipalName** – user's global unique Feide name and **eduPersonAffiliation** – the user's relation to the host organization. Attribute names and semantics are agreed upon in advance between the SP and the IdP. Feide uses a specification common to the Nordic and European educational communities, *NorEdu* [noredu].

The complete assertion given in the example above will look as shown in Figure 12.

5.6 The Authentication Response

The IdP sends the assertion back to the SP in response to the authentication request in an *Authentication Response* message, see Figure 13.

The **ID** attribute assigns a unique ID to the response. **InResponseTo** associates it with a unique request.

5.7 Single Log Out

Analogous to the Web Single Sign-On profile, there is also a profile for *Single Log-Out* (SLO).

SAML 2.0 defines two protocol elements for logout, the **LogoutRequest** and the **LogoutResponse**. A SAML entity (SP or IdP) can send a **LogoutRequest** to terminate the user's session at the entity in which the request is sent to, see Figure 14.

A **LogoutRequest** includes the **Issuer** element we have seen earlier, identifying the sender of the

```
<saml:Assertion Version="2.0" ID="s2010959d600bde4eb554901f23737cca324b984fe"
   IssueInstant="2007-08-17T11:49:50Z">
   <saml:Issuer>sam.feide.no</saml:Issuer>
   <Signature xmlns="http://www.w3.org/2000/09/xmldsig#">[...]</Signature>
   <saml:Subject>
      <saml:NameID Format="urn:oasis:names:tc:SAML:2.0:nameid-format:transient"
         >5/CI82rUJ6o53pR9Mo7IBKrvGGmx</saml:NameID>
      <saml:SubjectConfirmation Method="urn:oasis:names:tc:SAML:2.0:cm:bearer">
         <saml:SubjectConfirmationData NotOnOrAfter="2007-08-17T19:49:50Z"
            InResponseTo="_4e498f11858f139ddb491f2f1303658b4c4cc6e4b6"
            Recipient="https://sptest.feide.no/saml2/AssertionConsumerService"/>
      </saml:SubjectConfirmation>
   </saml:Subject>
   <saml:Conditions
      NotBefore="2007-08-17T11:39:50Z"
      NotOnOrAfter="2007-08-17T19:49:50Z">
      <saml:AudienceRestriction>
         <saml:Audience>https://sptest.feide.no</saml:Audience>
      </saml:AudienceRestriction>
   </saml:Conditions>
   <saml:AuthnStatement AuthnInstant="2007-08-17T11:49:50Z"
      SessionIndex="s22c0d60769d7eb12475d3210935a07daa1710d501">
      <saml:AuthnContext>
         <saml:AuthnContextClassRef
            >urn:oasis:names:tc:SAML:2.0:ac:classes:PasswordProtectedTransport</saml:AuthnContextClassRef>
      </saml:AuthnContext>
   </saml:AuthnStatement>
   <saml:AttributeStatement>
      <saml:Attribute Name="givenName"> [2 lines]
      <saml:Attribute Name="eduPersonPrincipalName">
         <saml:AttributeValue>andreas@uninett.no</saml:AttributeValue>
      </saml:Attribute>
      <saml:Attribute Name="eduPersonAffiliation">
         <saml:AttributeValue>employee</saml:AttributeValue>
      </saml:Attribute>
   </saml:AttributeStatement>
</saml:Assertion>
```

Figure 12 The complete assertion

ISSN 0085-7130©Telenor ASA 2007

```
<samlp:Response
   xmlns:samlp="urn:oasis:names:tc:SAML:2.0:protocol"
   xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion"
   ID="s2d4c93af63b9962dff5cdb8a0ff5f487b98655864"
   InResponseTo="_4e498f11858f139ddb491f2f103658b4c4cc6e4b6" Version="2.0"
   IssueInstant="2007-08-17T11:49:50Z"
   Destination="https://sptest.feide.no/saml2/AssertionConsumerService">

   <saml:Issuer xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion">sam.feide.no</saml:Issuer>

   <samlp:Status xmlns:samlp="urn:oasis:names:tc:SAML:2.0:protocol">
      <samlp:StatusCode xmlns:samlp="urn:oasis:names:tc:SAML:2.0:protocol"
         Value="urn:oasis:names:tc:SAMML:2.0:status:Success"> </samlp:StatusCode>
   </samlp:Status>

   <saml:Assertion Version="2.0" ID="s2010959d600bde4eb554901f23737cca324b984fe" [38 lines]

</samlp:Response>
```

Figure 13 The Authentication Response

```
<samlp:LogoutRequest
   xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion"
   xmlns:samlp="urn:oasis:names:tc:SAML:2.0:protocol"
   ID="c0add0bb3c1ca362b59ba40ef8e06df2ea0bab7d89" Version="2.0"
   IssueInstant="2007-08-20T06:54:49Z">
   <saml:Issuer>https://sptest.feide.no</saml:Issuer>
   <saml:NameID Format="urn:oasis:names:tc:SAML:2.0:nameid-format:transient"
      >5/CI82rUJ6o53pR9Mo7IBKrvGGmx</saml:NameID>
   <samlp:SessionIndex xmlns:samlp="urn:oasis:names:tc:SAML:2.0:protocol"
      >s22c0d60769d7eb12475d3210935a07daa1710d501</samlp:SessionIndex>
</samlp:LogoutRequest>
```

Figure 14 The Logout Request

protocol message. **`NameID`** specifies the user to be logged out, **`SessionIndex`** the session. This allows the receiver to distinguish between multiple simultaneous sessions initiated by the same user.

When a SAML entity receives a logout request, it terminates its own session and issues logout requests to all other parties that it has received from or sent assertions to. Figure 15 shows how the SP first issues a logout request to the IdP, and the IdP then issues a logout request to all other parties. At last, the IdP sends back a logout response to the SP to tell that the logout operation was successful.

5.8 SAML 2.0 Bindings

We have looked at message syntax, content and protocol flow. The SAML 2.0 *bindings* specifications [SAMLbinding] define how these XML documents are sent on the wire. A binding describes a way of transferring a SAML 2.0 message from one entity to another.

An important property of a binding is whether the message is sent *back-* or *front-channel*. Front-channel implies that a message issued by a SAML entity is sent via the user's browser to the other SAML entity. Back-channel message exchange is performed directly between SAML entities.

Front-channel communication is asynchronous, in the sense that the user's browser is diverted from the website, and the requester awaits a response on a predefined response consumer endpoint. Both entities have access to the local session cookie sent by the user to identify the session. The SAML entities do not communicate directly in a front-channel message exchange, but via the user's browser. To counter user tampering with the message, front channel bindings are often combined with some sort of signing mechanism, either XML signatures in the message itself, or a binding specific mechanism.

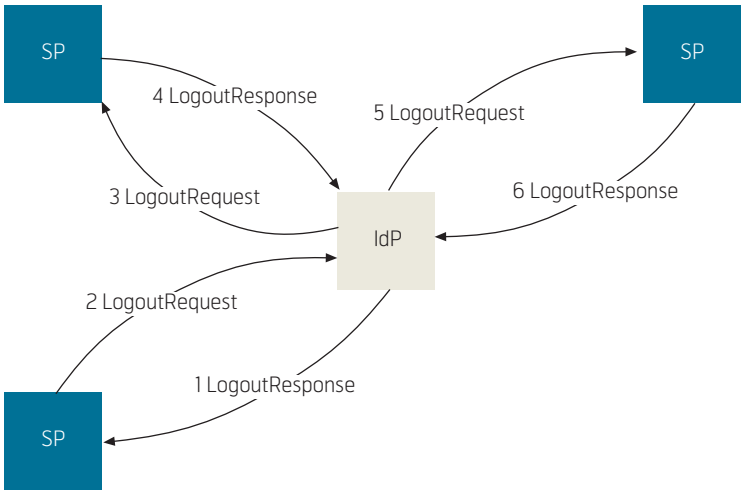


Figure 15 Single Log-Out

ISSN 0085-7130

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
      "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<body onload="document.forms[0].submit<>">

<form action="https://sam.feide.no/SingleSignOn"
method="post">
   <input type="hidden"
      name="SAMLRequest"
      value="PHNhbWxw0kxvAskxhwW[snipp]" />
</form>

</body>
</html>
```

Figure 16 The HTTP POST binding

The *HTTP Redirect* binding is based on the redirect functionality of the HTTP protocol [http]. As an example: The user requests some information from the SP, but the SP requires authentication by the IdP. Before returning any readable content to the user, the SP encodes an authentication request conforming to the HTTP Redirect specifications and returns this to the user. The redirect causes the user's browser to send a new request to some endpoint (i.e. a specific URL) at the IdP. This endpoint decodes the request, and then transfers the user back to the SP with the response message, using either HTTP Redirect, or some other front-channel binding.

The HTTP Redirect binding embeds the entire message in the URL, so it is not suited for transport of large SAML documents, such as authentication responses with many attributes.

An alternative binding, very similar to the HTTP Redirect binding, does not use normal redirection. Rather, it sends to the user's browser an HTML page with a **`FORM`** [html] that is automatically submitted to the URL specified in the **`FORM`** definition, by using Javascript code. The binding is called *HTTP POST binding*. The SAML message itself is base64 encoded and is included in a hidden HTML form **`input`** element (Figure 16).

A third alternative, *HTTP Artifact*, is suitable for transferring larger SAML messages from one entity to another. This binding uses the HTTP redirect functionality the same way as the HTTP Redirect binding, but instead of encoding the SAML message in the URL, a unique identifier, an *artifact*, represents the message. The artifact is included in the URL and sent to the receiver.

The artifact conveys no information by itself. The receiver must obtain the SAML message from the artifact via a separate SOAP call directly from the artifact recipient to the artifact issuer.

Back-channel communication is messages exchanged directly between SAML entities, without going via the user's web browser. The HTTP bindings specification [HTTPbinding] defines one back-channel binding called the *SOAP binding*. Two important aspects of SOAP binding:

1 The receiver will not receive cookies from the user's browser, hence cannot map the user to an existing session. Instead, the **`NameID`** and **`SessionIndex`** attributes must be used to lookup the appropriate session from a session storage.

2 The receiver has no control over the user's browser, which will keep its focus at the issuer of the request, the SP. So, it makes no sense for an SP to send an authentication request over SOAP – until the user is authenticated there is no way the IdP could ask the user for the credentials.

The HTTP Artifact binding is considered front-channel, but the artifact resolution part of this binding follows the rules of the SOAP binding.

As the Authentication Request is usually small, a HTTP Redirect binding is often used for this request. For the Authentication Response, HTTP POST and HTTP Artifact are both widely used.

For Logout Requests and responses, HTTP Redirect or SOAP are most common.

5.9 SAML 2.0 Metadata

SPs and IdPs are associated in a predefined trust relationship. Groups of SPs and IdPs trusting each other are sometimes referred to as *federations*. Liberty Alliance uses the term *Circles of Trust* (CoT) [idff12].

Each SAML entity (SP or IdP) is configured with a list of all trusted entities within the same federation. SAML 2.0 specifies an XML format for exchanging such trust information, as well as necessary information about SAML entities, such as which protocols, profiles and bindings each entity uses. Figure 17

ISSN 0085-7130©Telenor ASA 2007

```
<EntityDescriptor entityID="sptest.feide.no" xmlns="urn:oasis:names:tc:SAML:2.0:metadata">
   <SPSSODescriptor
      AuthnRequestsSigned="false"
      WantAssertionsSigned="false"
      protocolSupportEnumeration="urn:oasis:names:tc:SAML:2.0:protocol>

      <SingleLogoutService
         Binding="urn:oasis:names:tc:SAML:2.0:bindings:HTTP-Redirect"
         Location="https://sptest.feide.no/SAML2/LogoutService"/>

      <NameIDFormat>urn:oasis:names:tc:SAML:2.0:nameid-format:transient</NameIDFormat>
      <NameIDFormat>urn:oasis:names:tc:SAML:2.0:nameid-format:persistent</NameIDFormat>

      <AssertionConsumerService index="0" isDefault="true"
         Binding="urn:oasis:names:tc:SAML:2.0:bindings:HTTP-POST"
         Location="https://sptest.feide.no/SAML2/AssertionConsumer"/>
   </SPSSODescriptor>
</EntityDescriptor>
```

Figure 17 SAML 2.0 Metadata for an SP (Service Provider)

shows an example of SAML 2.0 metadata for a SP. The metadata in Figure 17 defines a **SingleLogoutService** endpoint that receives Logout Requests sent using the HTTP Redirect binding, and an **AssertionConsumerService** that receives Authentication Responses sent using the HTTP POST binding.

Similar, the document in Figure 18 shows the metadata for the Feide IdP. The document contains the X.509 certificate that Feide is using to sign assertions, as well as a HTTP-REDIRECT endpoint for Single Sign-On and one for logout.

The Single Sign-On endpoint (identified by the **SingleSignOnService** element) accepts incoming

```
<EntityDescriptor xmlns:xsi="https://www.w3.org/2001/XMLSchema-instance"
xmlns="urn:oasis:names:tc:SAML:2.0:metadata"
entityID="sam.feide.no">
   <IDPSSODescriptor
      WantAuthnRequestsSigned="false"
      protocolSupportEnumeration="urn:oasis:names:tc:SAML:2.0:protocol">

      <KeyDescriptor use="signing">
         <ds:KeyInfo xmlns:ds="http://www.w3.org/2000/09/xmldsig*">
            <ds:X509Data>
               <ds:X509Certificate>MIICIDCCAYkCBEXArTMwDQYJKoZIhvcNAQEFBQAwVzELMAkGA1UEBhMCVVMxDDAKBgNVBAoTA1N1
bjEnMCUGA1UECxMeU1VOIEphdmEgU31zdGVtIEFjY2VzcyBNYW5hZ2VyMREwDwYDVQQDEwhzdW4t
dW5peDAeFw0wNzAxMzExNDUyMzVaFw0wNzA1MDExNDUyMzVaMFcxCzAJBgNVBAYTAlVTMQwwCgYD
VQQKEwNTdW4xJzAlBgNVBAsTHlNVTiBKYXZhIFN5c3RlbSBBY2nLc3MgTWFuYWdlcjERMA8GA1UE
AxMIc3VuLXVuaXgwgZ8wDQYJKoZlhvcNAQEBBQADgY0AMIGJAoGBAIwp8/BWf8oRZv/llnoQAaQc
Rz7BbZSrKU+x/+sG0s/a/8NFsIZZZK3BDfSZtm9Te0gzhUEDb/2y0wyxLDeXzrG04lwQtfWz9oSW
Q70gsy9mVCEwJACSseIL7SgzDU0qWrrn/YelBujhu2RIY6dyWxe9lYp2h07/UmViATMQJEO/RgMB
AAEwDQYJKoZIhvcNAQEFBQADgYERacGGXl1ApPdULnBwdiMKdURA/d31uKiLfJzS4LfHpoh9M1Xq
54OROEPN07qVcMFZR1Vtl37csY4EUdpwSjGboNQscG4lFkw2/lvldDmZWPkid/2rDZR7vxdZv/RK
Wp2xWqtDlsEgMY7s+h0UbfW1wX6W9hIAMmpAFei9hWgPmiQ=</ds:X509Certificate>
            </ds:X509Data>
         </ds:KeyInfo>
   </KeyDescriptor>


   <!-- Logout endpoint -->
   <SingleLogoutService
      Binding="urn:oasis:names:tc:SAML:2.0:bindings:HTTP-Redirect"
      Location="https://sam.feide.no/amserver/IDPSloRedirect/metaAlias/idp"
      ResponseLocation="https://sam.feide.no/amserver/IDPSloRedirect/metaAlias/idp"
      index="0"
      isDefault="true"
      />

   <!-- Supported Name Identifier Formats -->
   <NameIDFormat>urn:oasis:names:tc:SAML:2.0:nameid-format:persistent</NameIDFormat>
   <NameIDFormat>urn:oasis:names:tc:SAML:2.0:nameid-format:transient</NameIDFormat>

   <!-- AuthenticationRequest Consumer endpoint -->
   <SingleSignOnService
      Binding="urn:oasis:names:tc:SAML:2.0:bindings:HTTP-Redirect"
      Location="https://sam.feide.no/amserver/SSORedirect/metaAlias/idp"
      index="0"
      isDefault="true"
      />

   </IDPSSODescriptor>
</EntityDescriptor>
```

Figure 18 SAML 2.0 Metadata for an IdP (Identity Provider)

ISSN 0085-7130

SAML 2.0 authentication requests, and the Single Logout service (identified by the **`SingleLogoutService`** element) accepts incoming logout requests.

6 Cross-Federation

Some users are not affiliated with any Feide organization, but are known in other federations offering Single Sign-On and similar services. These users may be given access to Feide services by cross federating. One example is parents of elementary school pupils; the parents may be granted access to Feide services by logging in to the Norwegian public authentication service "MinSide".

Service providers not belonging to Feide may also recognize Feide authentication for users of their own services, such as Norwegian researchers gaining access to the supercomputing facilities of the Finnish universities through Feide authentication.

Feide cooperates with several federations to allow a user authenticated by one federation to use the services offered by another federation without having to go through a second login procedure. Cross-federation requires the cooperating federations to trust each other's authentication procedures. When a service requests authentication of a user from another federation, and attribute information for this user, Feide forwards the request to the federation responsible for authenticating the user. A Feide user identifies himself by his Feide name. In the user interface, the user specifies his local username and selects his organizational affiliation from a drop down list. Internally, these are combined to one string: <UserName>@<HostOrganization>. The second part indicates where the Feide login service should go for authentication and information attributes about the user. A host organization manages a standard set of attributes for each user.

6.1 Cross-Federation Policy

The Feide federation may cross-federate with other federations through agreements that set the following conditions:

- The direction of federation: Whether externally authenticated users shall be allowed access to Feide services, and/or Feide authenticated user shall be allowed access to services in the other federation;
- Legally responsible persons representing each federation;
- Which technical standards the cross federation is based on;
- Contacts for administrative and technical issues;
- Guidelines for attribute exchange: Which attributes may be exchanged (this depends on the direction of federation) and how attributes in other federations are mapped to Feide attributes.

Feide will only cross federate with other federations that authenticate users at a confidence level comparable to Feide's own procedures.

Feide is testing cross-federation with the Norwegian government *e*ID (MinSide), on a Nordic scale (Kalmar union), in Europe (eduGAIN) and other more experimental solutions like provisioning for user centric identity. The interconnection with *e*ID is straightforward SAML2 interconnecting Circles of Trust, and is not presented here.

6.2 Nordic Interconnections: Kalmar Union

The Nordic research and education community is organized into national federations, where HAKA (Finland) and Feide have the most extensive participation from their constituencies (around 73 % of Finnish and Norwegian users mid-2007), SWAMI in Sweden and DK-AAI in Denmark have less universities participating, and Iceland has a different infrastructure. Testing of Nordic interconnections has been successful [Linden2006], connecting Shibboleth 1.3 with the Feide SAML2.0 infrastructure. Information models share similarities, with SWAMI and Feide using the same schemas, and HAKA using extensions to eduPerson. A feasibility study [Tveter2007] indicates that the remaining policy work entails minor adjustments of local contractual framework; the rest is covered in a Kalmar union framework [Kalmar] compliant with the EU privacy regulations [94/46/EC].

6.3 European Research Infrastructure: eduGAIN

European work on interconnecting existing authentication and authorization infrastructure has led to eduGAIN, a design [Lopez2005] which is backward compatible with a number of European authentication and authorization infrastructures in higher education and research (PAPI in Spain, Shibboleth 1.3 in various countries, Feide in Norway). Our work has focused on leveraging the SAML2 capabilities with existing solutions, providing bridges between installed base infrastructure to ensure ubiquitous access to research and educational material across Europe.

There is on-going work implementing universal sign-on for web access (SAML/eduGAIN) and Wi-Fi access (801.X/eduroam). eduroam is an established infrastructure for Wi-Fi access. eduroam stands for

ISSN 0085-7130©Telenor ASA 2007

Education Roaming and is a federation of institutions that offers Internet connectivity to the users of all other member institutions. This is done by accepting the credentials from the user's home institution through a RADIUS hierarchy. eduroam started as a European cooperation but now spans beyond the European borders.

6.4 Provisioning for User Centric Identity

A special case of cross-federation is when the Feide infrastructure is used to provision a user centric account, for example openID [openID], for each user. Experiments with the technology show that it is easy to use an organization centric solution (Feide) to provision user centric identities (openID). Other user centric solutions, such as CardSpace have not been investigated fully, but preliminary investigation in the Shibboleth community shows challenges with attribute release. The policy part of setting up such provisioning solutions is more unclear at this stage, and needs further investigation of levels of authentication, privacy and security.

7 Future Work

Feide has established a working federation for the majority of users in higher education in Norway. Schools are starting to add users, and we see the benefit in having an integrated user space, with a standardized infrastructure for Single Sign-On and information release. Integration of IT services is a growing field, and identity management is a necessary, but not sufficient condition for integration. Rolling out Feide to all users in education is the major focus of the next years, and encourage service developers to reap the benefits from the identity management infrastructure.

Service provider integration will become easier as the field matures, both in understanding of issues and in software available. A service provider may establish a single contract with Feide and add SAML2 capabilities to his solutions, and thereby have an easy way to reach all educational users in Norway.

Further work on cross-federation includes establishing operational cross-federations for services demanded by international users; this requires technical testing and reaching agreements on cross-federation policies. There are challenges in crossing borders outside the European Union sphere (Norway subscribes to regulations through the EEZ agreement), for instance with universities in the USA. In the European context eduGAIN is expected to get traction within the next year and provide a basis for operational experience and investigation.

8 References

[norEdu*] Melve, I et al. *norEdu* Object Class Specification*. February 2007. http://feide.no/dokumenter/norEdu-current.pdf

[eduPerson] Hazelton, K. *EDUCAUSE/Internet2, eduPerson specification*. April 2006. http://www.nmi-edit.org/eduPerson/internet2-mace-dir-eduperson-200604.html

[eduOrg] Hazelton, K. *EDUCAUSE/Internet2, eduOrg specification*. October 2002. http://www.nmi-edit.org/eduOrg/internet2-mace-dir-eduOrg-200210.pdf

[POL] *Act of 14 April 2000 No. 31 relating to the processing of personal data (Personal Data Act)*. http://www.datatilsynet.no/upload/Dokumenter/regelverk/lov_forskrift/lov-20000414-031-eng.pdf

[94/46/EC] EP95 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data. *Official Journal*, L 281, 23 Nov 1995, 0031-0050.

[Tveter2007] Tveter, W M, Melve, I, Linden, M. *Towards interconnecting the Nordic Identity Federations, Selected papers of Terena Networking Conference 2007*, May 2007. http://www.terena.org/publications/tnc2007-proceedings/

[Linden2006] Linden, M, Solberg, A. *Nordic Middleware Federation, NORDUnet2006 conference*, September 2006. http://www.nordu.net/conference 2006/presentations/We14.pdf

[Sheckler2007] Sheckler, V (ed) et al. *Liberty Alliance Contractual Framework Outline for Circles of Trust*. September 24, 2007 [online] – URL: http://www.project-liberty.org/liberty/ content/download/2962/19808/file/Liberty%20Legal%20 Frameworks.pdf

[Kalmar] *Kalmar Union Policy*. September 3, 2007 [online] – URL: http://feide.no/kalmar/

[Lopez2005] Lopez, D R et al. *GÉANT2 Authorisation and Authentication Infrastructure (AAI) Architecture*, first edition July 2005, second edition April 2007. http://www.geant2.net/upload/pdf/GN2-07-024-DJ5-2-2-2-GEANT2_AAI_Architecture_And_Design.pdf

ISSN 0085-7130©Telenor ASA 2007

[saml1diff] *Differences between OASIS Security 2 Assertion Markup Language (SAML) 3 V1.1 and V1.0*. September 3, 2007 [online] – URL: http://www.oasis-open.org/ committees/ download.php/3412/sstc-saml-diff-1.1-draft-01.pdf

[SAML2] *SAML V2.0 OASIS Standard specification set.* September 3, 2007 [online] – URL: http://www.oasis-open.org/committees/ tc_home.php?wg_abbrev=security#samlv20

[understandingws] *Understanding WS-Federation*. September 3, 2007 [online] – URL: http://download.boulder.ibm.com/ibmdl/pub/software/dw/specs/ws-fed/WS-FederationSpec05282007.pdf

[liberty] *Liberty Alliance Homepage*. September 3, 2007 [online] – URL: http://www.projectliberty.org/

[idff12] *ID-FF specifications*. September 3, 2007 [online] – URL: http://www.projectliberty.org/ resource_center/specifications/liberty_alliance_id_ff_1_2_specifications

[rfc2965] IETF. *HTTP State Management Mechanism*. October 2000. (RFC 2965) [online] – URL: http://www.ietf.org/rfc/rfc2965.txt

[SOAP] *SOAP Messaging Framework W3C Recommendation*. September 3, 2007 [online] – URL: http://www.w3.org/TR/soap12-part1/

[SAMLBinding] *Bindings for the OASIS Security Assertion Markup Language (SAML) V1.2*. September 3, 2007 [online] – URL: http://docs.oasis-open.org/security/saml/v2.0/saml-bindings-2.0-os.pdf

[xmlschema] *W3C XML Schema*. September 3, 2007 [online] – URL: http://www.w3.org/XML/Schema

[xmlnamespace] *Namespaces in XML 1.0*. September 3, 2007 [online] – URL: http://www.w3.org/TR/REC-xml-names/

[XMLsig] *XML-Signature Syntax and Processing*. September 3, 2007 [online] – URL: http://www.w3.org/TR/xmldsig-core/

[html] *HTML 4.01 Specification*. September 3, 2007 [online] – URL: http://www.w3.org/TR/html401/

[http] IETF. *Hypertext Transfer Protocol HTTP/1.1*. November 19, 2006. (RFC2616bis) [online] – URL: http://www.w3.org/Protocols/HTTP/1.1/rfc2616bis/ draft-lafon-rfc2616bis-02.txt

Ingrid Melve is Chief Technology Officer in UNINETT since 2006. She is the manager of Feide, identity management for education, and was the main architect of the design of Feide in 2000. From 1998 she was Manager of Applications and Middleware for UNINETT, the Norwegian research network, where she has been working since 1994 in information services. She holds an MSc in Telecommunications from the Norwegian Institute of Technology (NTH), Trondheim.

email: ingrid.melve@uninett.no

Andreas Åkre Solberg has been a Researcher in UNINETT since 2004. He is involved in the research activities on authentication and authorization infrastructures organized through the pan-European research network GÉANT2. He has been part of the Feide project since the beginning of 2006. Prior to that, he worked with passive network monitoring in the department of research and development in UNINETT. Andreas graduated from the department of telemathics at the Norwegian University of Science and Technology (NTNU) in Trondheim in 2004 with an MSc in communication technology specializing in teletraffic engineering.

email: andreas.solberg@uninett.no

ISSN 0085-7130©Telenor ASA 2007

Identity Federation in a Multi Circle-of-Trust Constellation

DAO VAN TRAN, PÅL LØKSTAD, DO VAN THANH

Dao Van Tran is Research Scientist in Telenor R&I

The efficient management of user identities has become an essential function in *e*Business and *e*Government applications. But at present user identities on the Internet are fragmented across various identity providers: *e*Business services, portals, employers, public on-line services, etc. The management of multiple login/password combinations to access *e*Services is neither efficient for the professional (regarding functionality, cost and security), nor user-friendly and trustful for the end-user. This paper presents a Federated network identity management imlementation based on the Liberty Alliance that allows Singe-Sign-On (SSO) in a pan-European multi Circle of Trust environment. Bot the technical challenges and the business opportunities are presented.

Pål Løkstad is Adviser in Telenor R&I

Do Van Thanh is Senior Research Scientist in Telenor R&I

1 Introduction

An efficient management of user identities has become an essential requirement in *e*Business and *e*Government applications. But at present, user identities on the Internet are fragmented across various identity providers: *e*Business services, portals, employers, public on-line services, etc. The management of multiple login/password combinations to access *e*Services is neither efficient for the professional (regarding functionality, cost and security), nor user-friendly and trustful for the end-user (see Figure 1).

Federated network identity concepts that allow Single Sign-On (SSO) are proposed as a solution to the current shortcomings and as new business enablers. The Liberty Alliance [1] has elaborated a federated Identity Management (IDM) model based on open architectures and standards as opposed to proprietary solutions. Whilst the Liberty specification work is progressing quite well and quite fast, no complete evaluation has been made to test this concept. To evaluate the Liberty specifications the Fidelity Project (Federated Identity Management based on Liberty) was initiated. The goal of the Fidelity Project was to implement a federated pan-European IDM system based on the Liberty concept, and to evaluate its technical viability and performance and capability to meet business, end-user and security/privacy requirements. In the Fidelity Project a consortium of leading European telcos, industry and research organizations was established. Four Circles of Trust (CoTs) in four different countries were also implemented according to Liberty specifications. Their goal was to demonstrate interoperability, showing that local identity federations can interact at pan-European level, enabling exchange of identity and authentication of citizens between service and identity providers, whilst the usage and validity duration of identity data remains totally under the user's control and acceptance. The Fidelity Project also evaluated technical solutions for the implementation of appropriate elements in the fixed network and in the smart card of the mobile network (SIM). The

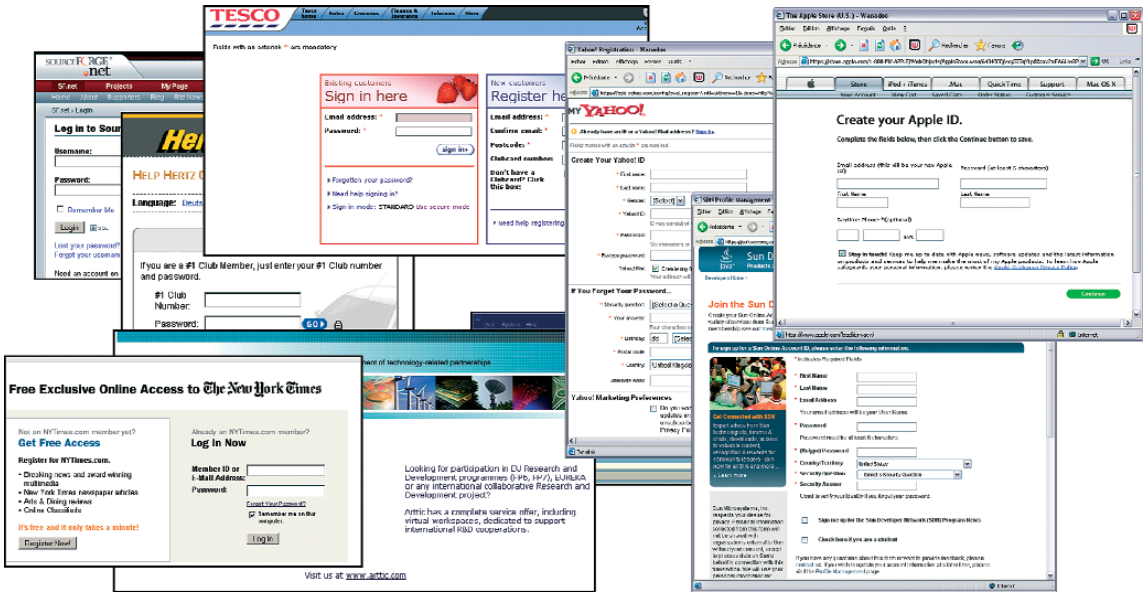


Figure 1 At present, user identities on the Internet are fragmented across various identity providers

ISSN 0085-7130

proof-of-concept tests and demonstrations included mobile, fixed and Internet scenarios. Added value services based on users' attributes, such as presence and geo-location, or other personal identity attributes, enhanced the demonstrations. The project results were analyzed and made available with recommendations and considerations about a totally new range of services particularly suited for telcos, on behalf of *e*Service providers: identity management, personal ID and attribute providers, identity/ attribute roaming in Inter-CoT context, and the negotiation of user controlled security/data levels in electronic transactions. The Fidelity consortium consists of the following partners: France Telecom, Telenor, Telia Sonera, Amena, Ericsson, Oslo University College, Linus, InetsSecure, Moviquity, Italtel and Gemalto.

2 Business Scenarios

From a business point of view, the Fidelity Project has identified different scenarios which can be divided into the five following categories:

- *Inside a company domain:* Large enterprises reorganization process with merging of different business units, each one of them with an already existing identity management system deployed.

- *Service Provider (SP) and Enterprises:* Typical B2B scenario where a SP has relationships with many large companies.

- *Alliances between operators:* Various nationally based operators form an alliance where each of them offers its SP to the alliance.

- *Between an operator and its trusted partners:* Traditional model of a mobile operator and its traditional service and content providers, with a relatively high level of trust due to the long term relationship between parties and regulation under "paper contract".

- *Between operators and Internet partners:* The Identity Provider (IDP) assumes the risk by providing identity management services to small content providers without contract, so if a content provider does not live up to the IDP's expectations it is expulsed from the CoT.

The foreseen economic savings can be quite substantial (in relative terms) depending on the business scenario. The general conclusion is that the bulk of the savings will probably come in the area of *customer management* with anticipated savings of up to 50 %. The second area where savings are foreseen is in the *cost of implementing a new software* application and hardware acquisitions. One should bear in mind that savings are defined as spending less money/time on the same activity, or spending the same amount of money/time on more activities.

The Fidelity Project set out to test commercial scenarios in order to evaluate the interoperability of the used Liberty implementations. As part of the project a number of use cases where defined which would satisfy the main objective of the project: testing the interoperability between various CoTs. These use cases are all but one use cases for the end-user and contain various scenarios like: fixed Internet HTTP scenario, mobile (Internet) environment, non-HTTP service. Also considering the implementation of concrete use cases in the project, one can draw additional conclusions as identified below.

The model as specified in Figure 2 shows a typical Intra-CoT scenario. The IDP (i.e. telecommunication operator) offers the services under its own brand. One Fidelity scenario which corresponds to this use is "Purchasing a Game". This means that the operator is not only capable but must provide a consistent user interface for the user consent service, i.e. when a service requires user data which is located in the CoT, for instance data from the Personal Profile. This can be done to a certain degree using today's Liberty implementations on the market.

It does however require a lot of integration work. For a commercial environment additional software is still required to provide the IDP/DS (Discovery Service) with tools for charging the SP/CP (Content Provider) for the services offered, like for instance high security authentication.

Going one step further we see the scheme of an IDP with Internet Partners. This is depicted in Figure 3. This scenario is similar to the previous one (i.e. also an Intra-CoT scenario) but where the operator provides generic services to the Service and Content Providers. The operator will provide trust to the service/content providers and may market authentication methods with a high level of security. The Fidelity

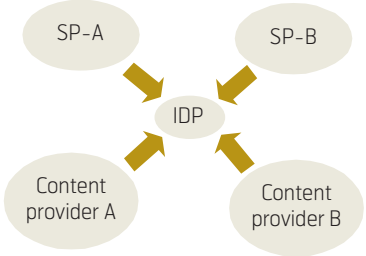


Figure 2 Intra-CoT scenario

ISSN 0085-7130©Telenor ASA 2007

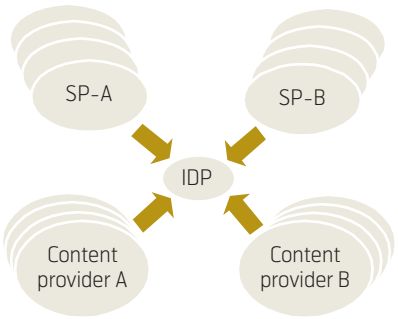


Figure 3 Intra-CoT scenario: IDP with Internet Partners

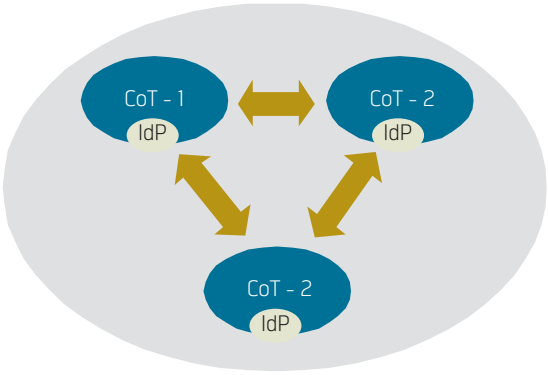


Figure 4 Inter-CoT scenario

examples are for instance the scenarios Book a Hotel and Search a Restaurant. These are very simple use cases, and it shows that this model is possible, requires less work for the operator (the user consent does not need to be homogeneous), but marketing is the key issue. As the operator is selling services to other companies, either directly or as a broker, it also needs additional software that provide the IDP/DS with tools for charging the SP/CP for the services offered.

The real Fidelity scenario is shown in Figure 4. This is the alliances of CoTs under a common name. It has been demonstrated that this can be set up, although it requires a lot of Liberty know-how, a lot of integration work and even more marketing work. A commercial relationship exists between the CoTs, and such tools are required to make sure that one CoT can charge the other CoT for services rendered. In the beginning of this relationship, when the "Liberty traffic" between the Telcos involved is low there is no need for this functionality. The major problem outstanding with this model is a marketing issue, questions like how to educate the end-user so he/she understands and trusts the Liberty technology, and how to market a Circle of Trust alliance to the end-user.

3 Fidelity Project's Test Bed

The Fidelity Project has established an Inter-CoT as shown in Figure 5, this Inter-CoT contains four CoTs: Norway CoT, Finland CoT, France CoT and Spain CoT.

The Fidelity Inter-CoT's test bed consists of the following components:

Norway CoT:
- SP/AT. Hotel Room Booking, Call a Contact, Personal Profile, Contact Book;
- IDP: Telenor IDP (with Discovery service) used IDP/IDM software from SUN.

France CoT:
- SP/AT: Hotel Room Booking, Attribute Registration, Wallet, Personal Profile, Student Exchange, etc.;
- IDP: French IDP, used IDP/IDM software from France Telecom R&D, IDMP.

Spain CoT:
- SP/AT: Hotel Room Booking, Where Restaurant, Personal Profile, Wallet, etc.;
- IDP: Spanish IDP, used IDP/IDM software from Ericsson.

Finland CoT:
- SP/AT: Hotel Room Booking, Download a Game, Personal Profile, Wallet, Discovery Service, etc.;
- IDP: Finland IDP, used IDP/IDM software from Trustgenix.

3.1 Hotel Room Booking Test Scenario

Below is a description of one of the test scenarios implemented and tested in the Fidelity Project. Most of the tests used the One Time Identify method as described in section 4.1.

Booking a hotel room is a quite simple and common task for travelers. If a customer wants to have accommodation for one or several nights, he/she goes to or calls the hotel reception in order to inquire whether the hotel has an appropriate free room and services with the price he/she is willing to pay. Using the hotel's reservation system, the receptionist checks the status of free rooms. To book a room, the receptionist asks the customer's name, address, phone number, preferences (smoking/non-smoking, top floor/second floor etc.) and typically credit card number. The receptionist puts this information into the reservation system and books the room.

The Internet provides an excellent means to perform the transaction described above using web-based online self-service. It is possible to use an online

ISSN 0085-7130

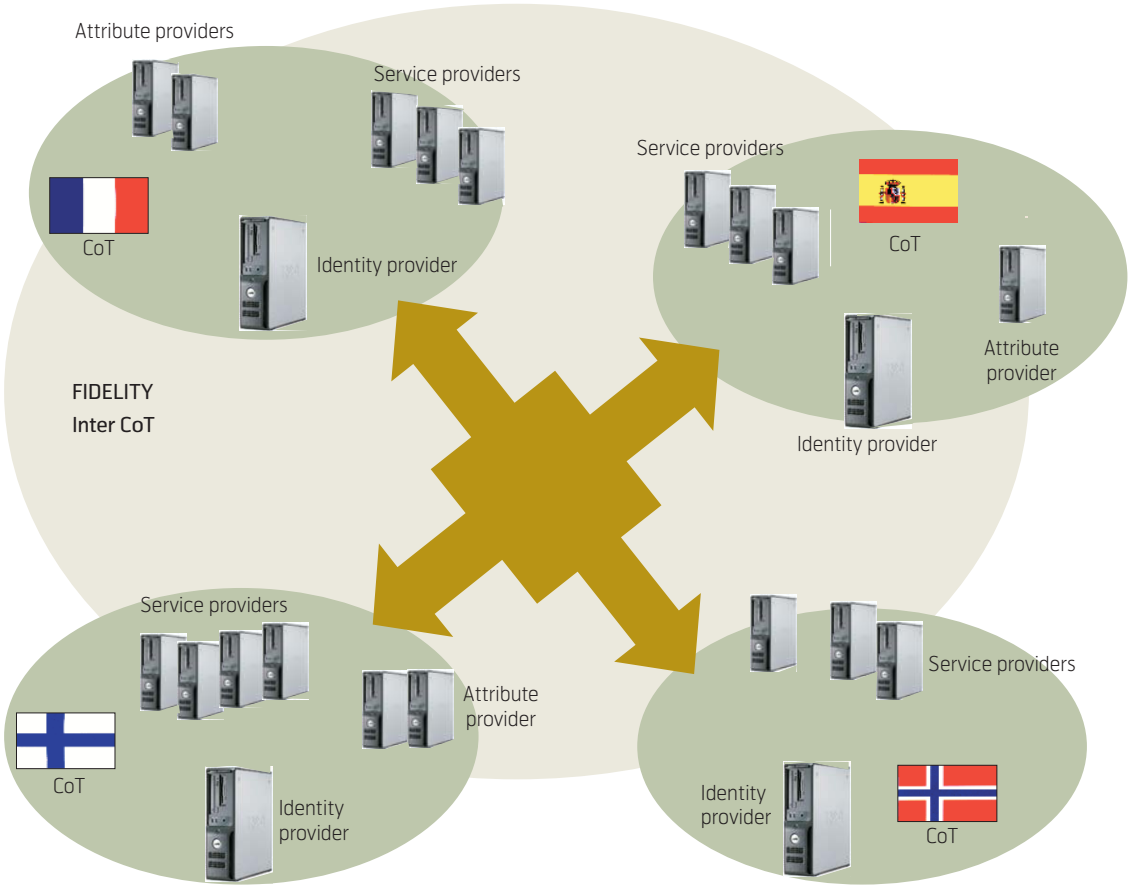


Figure 5 Inter-CoT test bed

hotel reservation service without registration and login. However, registered users will typically get a better service and it is easier for the user to perform reservation using information stored in the CRM (Customer Relationship Management) system of the hotel or hotel chain.

It is obvious that an online hotel reservation service fits well in an intra-CoT and especially an Inter-CoT eco-system. Using IDP's customer base, identity federation, one-click registration, Single Sign-On and identity-based services, hotels will get potentially more customers and they will be able to provide simplicity and trust to their customers' test scenarios.

The Fidelity Project tested this service, both on Intra-CoTs and Inter-CoTs. Every country tested the hotel booking service within its Intra-CoT to confirm that everything was working before testing in the Inter-CoT environment.

In this test scenario the user books a room on the online hotel reservation service which automatically retrieves the needed personal information (name, address, etc.) and the user's preferences (e.g. smoker/non-smoker, etc.), and following some prerequisites:

- Business agreements have been signed between the online hotel reservation service provider and the IDP telecommunication operator in another CoT, such that this IDP may authenticate the service provider's users.

- Business agreements have been signed between the two telecommunications operators, such that their IDPs trust each other and attribute exchange is possible.

The test scenario is described in detail below. The different points in the description are shown in Figure 6.

An end-user from Norway CoT who only has an account on his home IDP in Norway wants to log in to an SP – a hotel booking in another country/CoT, in this case Spain. He prefers to use One Time Identifier to do this, the Single Sign-On process will go as described below:

1 The user accesses the online hotel reservation service (SP) in Spain CoT with his/her desktop/laptop or his/her mobile and books a room for his/her travel. In order to complete the reservation, the online hotel reservation service needs some addi-

ISSN 0085-7130©Telenor ASA 2007

tional information such as user name, address or preferences. The online hotel reservation service proposes that the user automatically retrieve this information. The user approves this proposition.

2 As the user is not authenticated on the online hotel reservation service, the hotel reservation system will ask him which CoT he belongs to and he is redirected to the Visited IDP (Spain CoT) for authentication purposes. Meanwhile, the Visited IDP does not know this user and redirects him/her to his/her Home IDP (Norway CoT).

3 If the user has no authenticated session on his/her Home IDP, a user authentication is performed.

4 When user authentication is succeeded, the user is redirected to the online hotel reservation service placed in the Visited CoT (Spain CoT). Due to a commercial agreement between the two CoTs, the user is now authenticated on the online hotel reservation service.

5 The online hotel reservation service acts as a Web Service Consumer (WSC) and locates a Personal Profile Web Service Provider (WSP) for the user thanks to the Discovery Service (DS). The WSC requests the needed attributes of the user from the Personal Profile WSP (Norway CoT).

6 According to the user's policy, the WSP may need to get the user's consent to provide information to the WSC. If required, the WSP informs the user about the request from the online hotel reservation service for his/her personal and preferences information. The user gives his explicit consent.

7 The WSP returns the user's requested attributes to the WSC. Thus, the online hotel reservation service is able to provide to the user a pre-completed form (with name, address, phone numbers …) to confirm the room reservation and take into account the user's preferences (for example, smoker/non-smoker, …) to select the most appropriate room.

8 The online hotel reservation service returns a pre-completed form for room reservation with all the user's personal information and preferences.

9 The user modifies some information if needed, and validates the online reservation by clicking a button. The online hotel reservation service returns a confirmation message. A confirmation message can also be sent to the user by e-mail or SMS.

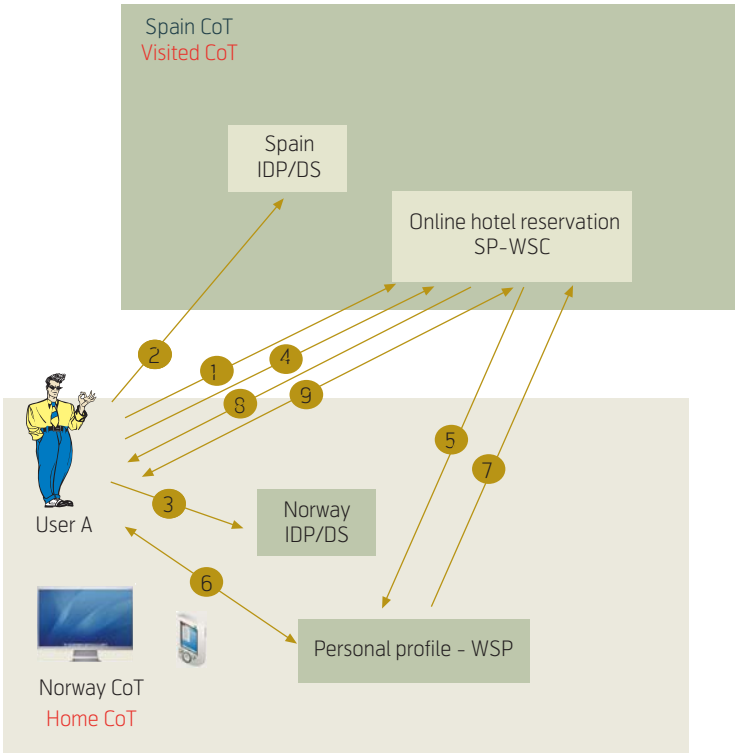


Figure 6 Inter-CoT test scenario: Hotel room booking

4 Inter-COT Single Sign-On

This section will give the results of the Single Sign-On (SSO) tests that have been done by Fidelity, focusing on every SSO functional discussions. SSO is described in the Liberty Alliance IDentity Federation Framework (ID-FF) specifications [2].

4.1 Federated Identifier / One Time Identifier

Liberty components (SP, IDP) identify a principal by sharing an identifier (or alias). The Liberty specification defines two methods for the management of the identifier. When a Service Provider asks his IDP for the authentication of a user, it specifies the type of identifier (i.e. alias) it wants to share with the IDP to identify the principal: Federated ID or One Time ID. The section below discusses the advantages and disadvantages of each method in an Inter-CoT environment, and then presents the functional results of the tests.

4.1.1 Federated Identifier

The prerequisite is that the user has created a local account on the Service Provider and has an account on the Identity Provider.

Federation is an organization formed by merging several groups or parties. In Federated Identity management, the various identities of the user are linked together. A federated identifier (alias) must be generated during a federation process and must be stored/

ISSN 0085-7130

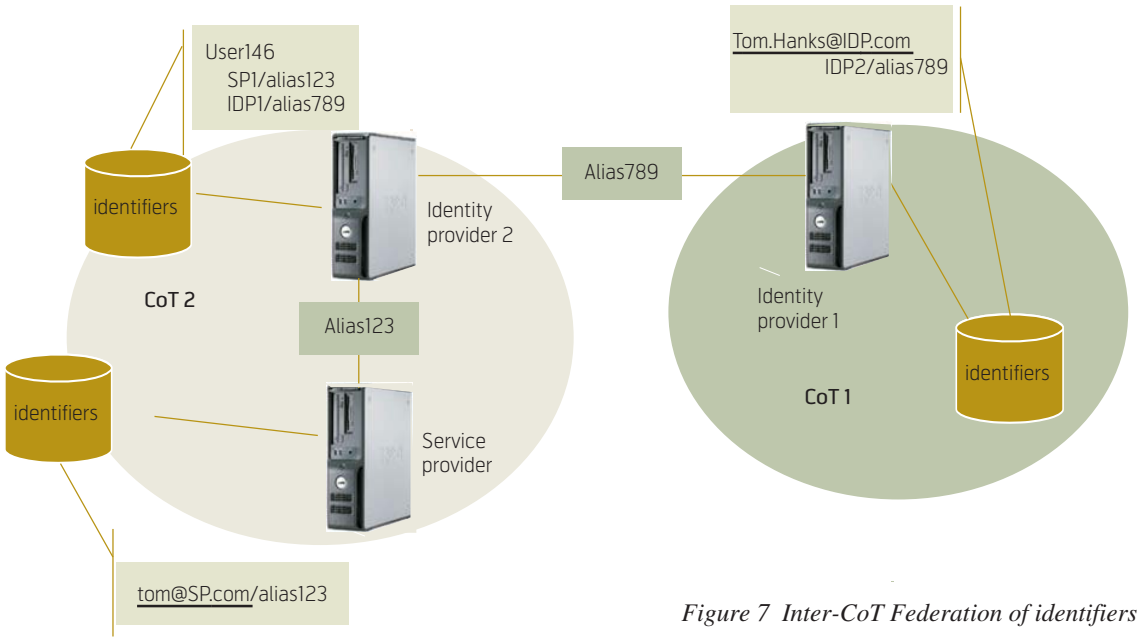


Figure 7 Inter-CoT Federation of identifiers

used by the two entities until a federation termination is requested.

In an Inter-CoT environment, the Visited Identity Provider (IDP 2) acts as a proxy IDP (as a SP of CoT 1), and a technical account has to be created on it. This technical account has to be federated with both the visited Service Provider account and the Home Identity Provider (IDP 1) account. Figure 7 shows Inter-CoT federation of identifiers.

Advantages:

- Service Provider and Identity Provider accounts are separated. If the user terminates his IDP account, he does not lose his Service Provider account.
- Service Provider has a complete account on the user. It can use the user's personal information even when the user is not logged in (e.g. for commercial purposes).
- User can log in locally as he used to do before federation.
- SP can use all the authentication methods offered by its IDP.

Disadvantages:

- User must create a local account.
- User must explicitly ask for the federation of his Service Provider account and his Identity Provider account.
- Local account information might not be as accurate or reliable as the information linked to the Identity Provider account.
- Service Provider must maintain a user database.
- If the account to federate is hosted by a Service Provider belonging to another Circle of Trust, an account has to be created on the visited IDP, and federated with both Service Provider account and Identity Provider account.

4.1.2 Federated Identifier with Technical Account

Just like Federated Identifier, federation with technical account merges the various identities of the user by defining a common alias between each identity hosted on a Service Provider and the identity hosted on the Identity Provider.

In this case, the account created on the Service Provider (SP1) is neither associated with credentials nor personal information; we call it technical account.

In an Inter-CoT environment, a technical account has to be created on the Visited Identity provider (IDP2). This Technical account has to be federated with both the visited Service provider technical account and the Home Identity provider account. Figure 8 shows Inter-CoT federation with technical account.

Advantages:

- User has a "virtual" local account on the SP. The SP can use the user's personal information even when the user is not logged in (e.g. for commercial purposes).
- Service Provider can recognize the user every time he logs in and store technical information about him.

ISSN 0085-7130©Telenor ASA 2007

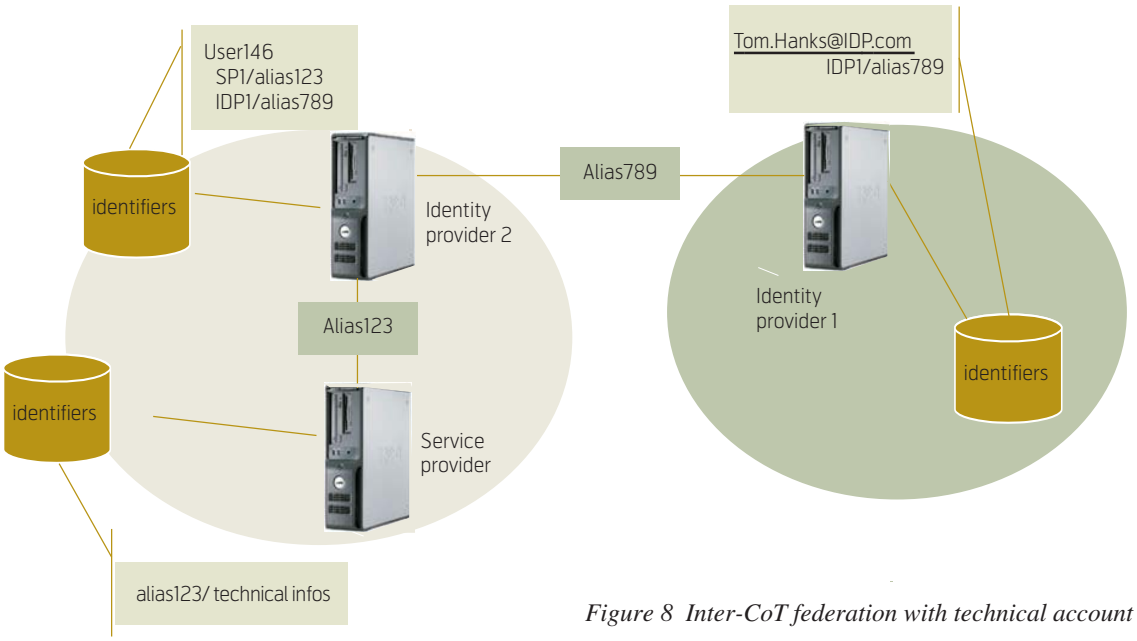


Figure 8 Inter-CoT federation with technical account

- User does not have to create a local account to log in to the Service Provider (it is made automatically).
- Personal attributes, hosted by Liberty components and used by the Service Provider are reliable.
- SP can use all the authentication methods offered by its IDP.

Disadvantages:

- If the Service Provider belongs to another Circle of Trust as the Identity Provider, an account has to be created on the visited IDP and federated with both Service Provider account and Identity Provider account.
- The Service Provider can only give access to users who have accounts on the Identity Provider of the CoT.
- If the user terminates his account on the IDP, he cannot login the Service Provider any more.
- Service Provider must maintain a user database.
- Service provider and Visited Identity Provider user databases might host lots of ghost accounts (created for unique visitors).

4.1.3 One Time Identifier

In the One Time identifier management process, the Service Provider does not store any local account for the user. Any user who is authenticated on the Identity Provider will be authorized by the Service Provider. A new one-time identifier must be generated at each time when a new authenticated session is created for the user and this one-time identifier may be used by the two entities until a single-logout is requested or until the end of the session.

The user can only log in to the Service Provider using his Identity Provider account. If he comes back on the SP and tries to log in again, a different One Time Alias will be given by the IDP to the SP, so that the SP cannot recognize the user. Once he is logged on the IDP, no other authentication will be asked of him. The One Time Alias allows the Service Provider to retrieve personal attributes of the user.

In an Inter-CoT Architecture the Visited Identity Provider (IDP 2) asks the Home Identity Provider (IDP 1) for a One Time Identifier. It will create another One Time Identifier for the Service Provider. Figure 9 shows Inter-CoT use of One Time identifier.

Advantages:

- User does not have to create a local account to log in to the Service Provider.
- It is a simple process to access a Service Provider and take advantage of the attributes sharing functionalities, especially in an Inter-CoT environment.
- The Service Provider does not know anything about the user except that he has been successfully authenticated: security and transparency from the user's point of view.
- The Service Provider does not have to manage any User Database.
- SP can use all the authentication methods offered by its IDP.

ISSN 0085-7130

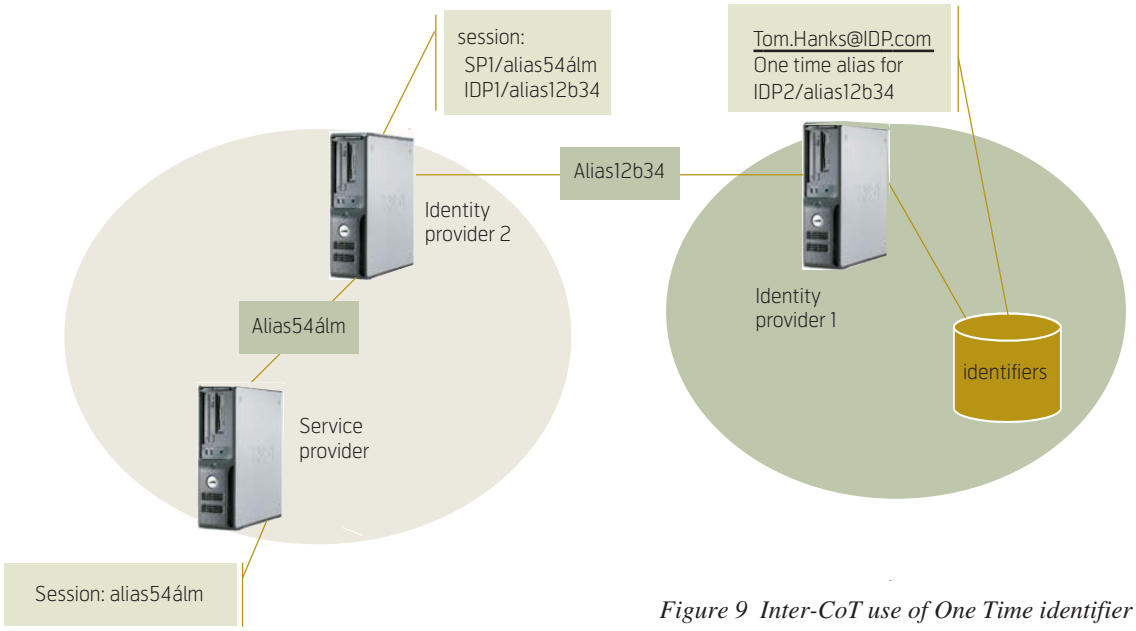


Figure 9 Inter-CoT use of One Time identifier

Disadvantages:

- The Service Provider can only give access to users who have accounts on the Identity Provider.
- The Service Provider cannot access user information when he is not logged in.
- The Service Provider is not able to recognize the user if he returns.
- If the user terminates his account on the Identity Provider, he cannot log in to the Service Provider any more.

One Time Identity management is the simplest way of taking advantage of Liberty functionalities (Single Sign-On and Attribute Sharing) and seems usually more relevant in most Inter-CoT usages.

The federation of a local account with the Identity Provider account is a more complicated method which requires to be clearly explained for end-user's acceptation. This method seems usually more adapted to most Intra-CoT usages.

4.2 IDP Introduction

Identity Provider Introduction is defined as the mechanism by which a provider discovers which identity providers a Principal may be using. In an Inter-CoT environment, the IDP introduction must be performed by the V-IDP:

- To determine if the Principal is a local (Intra-CoT) or roaming (Inter-CoT) user for the IDP;
- To resolve Principal's H-IDP in case of a roaming user.

The documents [5] and [8] issued by Liberty Alliance describes an "Identity Provider Introduction profile" based on "Common Domain Cookie" (CDC), where cookies are used to share information about user's IDP(s). This CDC mechanism does not seem suitable for an Inter-CoT environment, which aims to have "loosely coupled" relationships between CoTs.

Thus, the Fidelity Project has identified, studied and evaluated other IDP Introduction methods which seem to be more adapted to the Inter-CoT context:

- *User-interaction-based resolution*. Principal participates in his/her H-IDP resolution. This method relies on asking the Principal for his/her H-IDP.
- *User-agent-based resolution*. Information about Principal (e.g. MSISDN, HTTP header, IP address, H-IDP information on smart card, etc.) is available through user agent and can be used to automatically resolve Principal's H-IDP by the V-IDP.
- *Authentication-based resolution*. Authentication method supports resolution of H-IDP. This is applicable when the V-IDP authenticates the Principal via a shared/roaming authentication method and the method provides V-IDP with H-IDP information. V-IDP subsequently notifies H-IDP about Principal authentication.

Figure 10 shows an example of user-interaction-based resolution.

Figure 11 shows an example of IDP authentication (left side) and IDP Introduction page (right side).

ISSN 0085-7130©Telenor ASA 2007

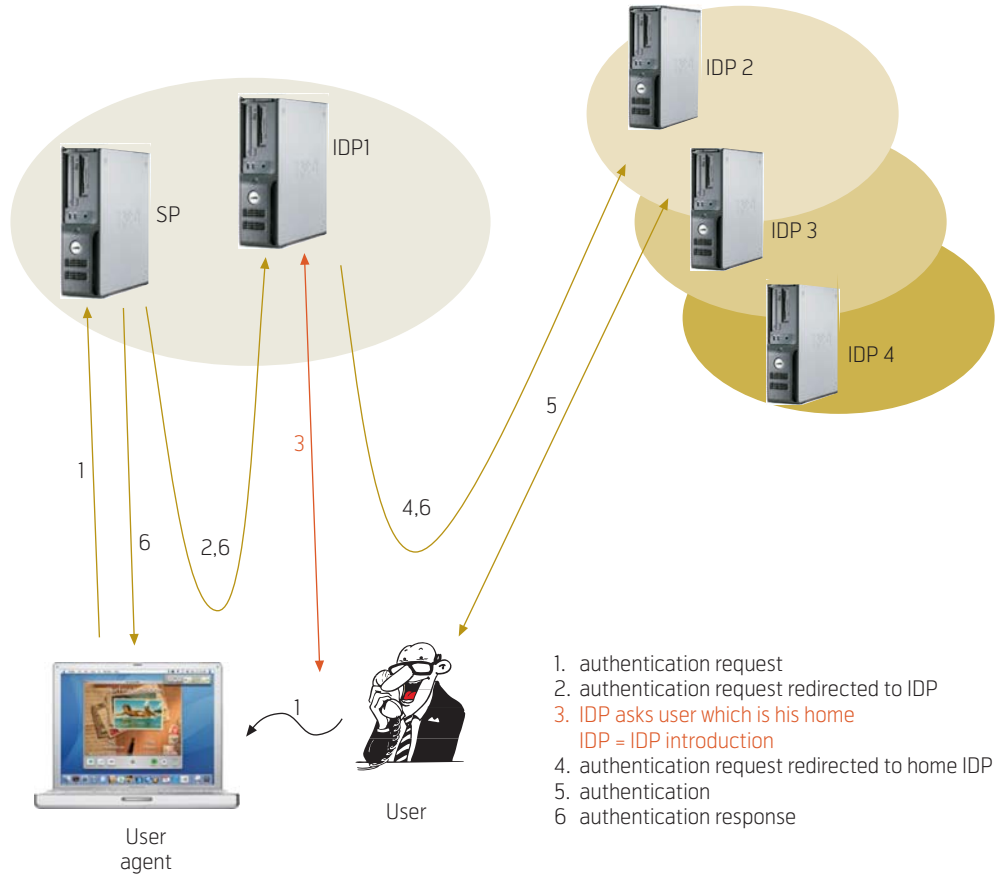


Figure 10 IDP introduction with user interaction

Figure 11 Fidelity IDP introduction with user interaction

ISSN 0085-7130

Fidelity experimentations focused on some user-agent-based resolution mechanisms (browser's preferred language, IP address, H-IDP given in smart card, etc.) and have raised feedbacks/limitations for each of these methods. One of the main conclusions is that user-agent-based resolution cannot be considered as a fully deterministic introduction method. Therefore, V-IDP should always implement a user-agent-based resolution method which combines different resolution mechanisms in order to obtain the most accurate result.

Transparent (or automatic) introduction methods (e.g. User-agent-based resolution or Authentication-based resolution), which do not require any user interaction, should be preferred since they are more user-friendly. However the end-user should always be able to correct or force "manually" the IDP resolution, and for this reason a user-interaction-based resolution should always be implemented to complete transparent/automatic introduction methods.

In order to enhance end-user experience, the result of the IDP user-interaction-based resolution may be maintained in a session/persistent cookie.

4.3 Authentication Context

Liberty Authentication Contexts, specified in the document "Liberty Authentication Context Specification" [7], are used to characterize the level/quality of the authentication the IDP provides to the SP.

An SP may include an Authentication Context into the authentication request to choose an authentication quality depending on its requirements or to ask an IDP to re-authenticate the Principal using a different authentication scheme. An IDP may provide in the authentication response an Authentication Context which informs the SP about the chosen authentication method.

Liberty specifications [7] define two ways of providing Authentication Context information: Authentication Context classes and Authentication Context statements.

In Inter-CoT context, some CoTs may use Authentication Context Class to characterize the level/quality of the authentication, whereas other CoTs may use Authentication Context Statements.

- How to manage Inter-CoT exchanges between two CoTs which do not use the same way of providing Authentication Context information?

- Who is in charge of doing the Authentication Context mapping: H-IDP and/or V-IDP?

Thus, it has been noted that the conformance of Liberty specifications is not enough to ensure interoperability of Authentication Context exchanges in Inter-CoT environments. The Fidelity Project had completed Liberty specifications by a set of technical and business recommendations to be followed by all H-IDP and V-IDP of the Fidelity CoT.

In order to enable Authentication Context (AC) exchanges between two CoTs which may potentially use different ACs, it seems essential to define a mapping table between AC from the first CoT and AC from the second one. To this end, three authentication levels have been defined for the Fidelity implementation and all authentication methods used have been classified into these three levels. Of course, this simple classification should not be considered as a global statement for all Inter-CoT Liberty deployments, which require a specific mapping table for each IDP-IDP relation.

As the V-SP knows only AC used in its own CoT, the H-IDP may qualify end-user's authentication with an AC not known by the V-SP. Thus, the V-IDP should not return the V-SP an authentication response with an unknown AC (for V-SP).

4.4 Metadata Exchanges

Liberty specifications [6] define format and protocol for the out-of-band exchanges of metadata information between entities. This metadata format aims to ensure the interoperability between Liberty components, which is a critical point in an Inter-CoT environment built with multi-vendor technology products. The feedback of the Fidelity Project about Liberty metadata exchanges:

- Some Liberty technology products do not implement this metadata format and/or do not provide any import/export capability.

- Several interoperability issues have been noticed during the integration stage between technology products: format errors, incorrect elements, missing elements, disordered elements.

- A slight mismatch has been identified between XSD schema and written specifications about the status (mandatory or optional) of the SOAP endpoint element: "Liberty Metadata Description and Discovery Specification" [6] and the related XML Schema ("Liberty-metadata-v1.1.xsd").

ISSN 0085-7130©Telenor ASA 2007

5 Inter-CoT Attribute Sharing

This section will give the result of Inter-CoT attribute sharing tests involving the user's own attributes. Attribute sharing is described in the Liberty Alliance IDentity Web Services Framework (ID-WSF) specifications [3].

5.1 PKI Trust Model

While in IDentity Federation Framework (ID-FF) scenarios the trust between communicating parties is always based on bilateral metadata exchange, establishing trust relationships within ID-WSF scenarios in Inter-CoT environments is a more complicated issue. Indeed, although Liberty specifications state that the Discovery Service (DS) or a Web Service Provider (WSP) has to know the Web Service Consumer (WSC) and load its metadata, it does not seem to be suitable to large CoTs (with numerous SP) and for Inter-CoT, where metadata exchanges may be considered as inconvenient or tedious. Furthermore, prior to a metadata exchange, a business agreement is typically established, but in Inter-CoT the V-SP and the H-IDP do not have a direct business agreement.

Due to this constraint, an innovative Public-Key Infrastructure (PKI) based trust model, using a hierarchical structure of Certificate Authorities (CAs), was demonstrated in Fidelity experimentation in order to avoid the provisioning problems related to metadata exchange and to establish a trust relationship based on certificates installed on all entities.

For each CoT, Fidelity established an intermediary CA which issued all the end-entity certificates of a particular CoT. The CA certificates of the four CoT-specific CAs were signed by one common root CA (TeliaSonera). The root certificate and/or the intermediate CA certificates for each four CoTs were deployed in all SPs (WSCs) and WSPs (e.g. DS). The ID-WSF entities do need to trust the CA of the opposing CoT's ID-WSF entity which requires deployment of the certificate chain of the CA that has issued the certificate of the ID-WSF counterpart.

The requirement for distributing CA certificates stems from the reason that a recipient of an ID-WSF message must be able to recognize an end-entity certificate embedded in the message to be signed by a trusted CA in order to perform a *certification path validation procedure* on which the PKI trust model is based functionality-wise.

Figure 12 depicts an exchange and distribution of CA certificate chains between two CoTs. This has to take place prior to any ID-WSF messaging in Inter-CoT environment.

Figure 13 shows how the PKI trust model works in action by demonstrating the case where an SP sends a service request to a WSP in a partner CoT.

5.2 ID-SIS Services

Liberty specifications define a set of IDentity Services Interface Specifications (ID-SIS) [4] services which favor the interoperability of attribute sharing between Liberty compliant components. Due to these specifications, WSC knows the list of available users' attributes exposed by a service and may query them without any previous exchanges.

Although most technology vendors implement ID-SIS Personal Profile, it has been noted that few of them implement one of the other Liberty ID-SIS services. Furthermore, the implementation of other ID-

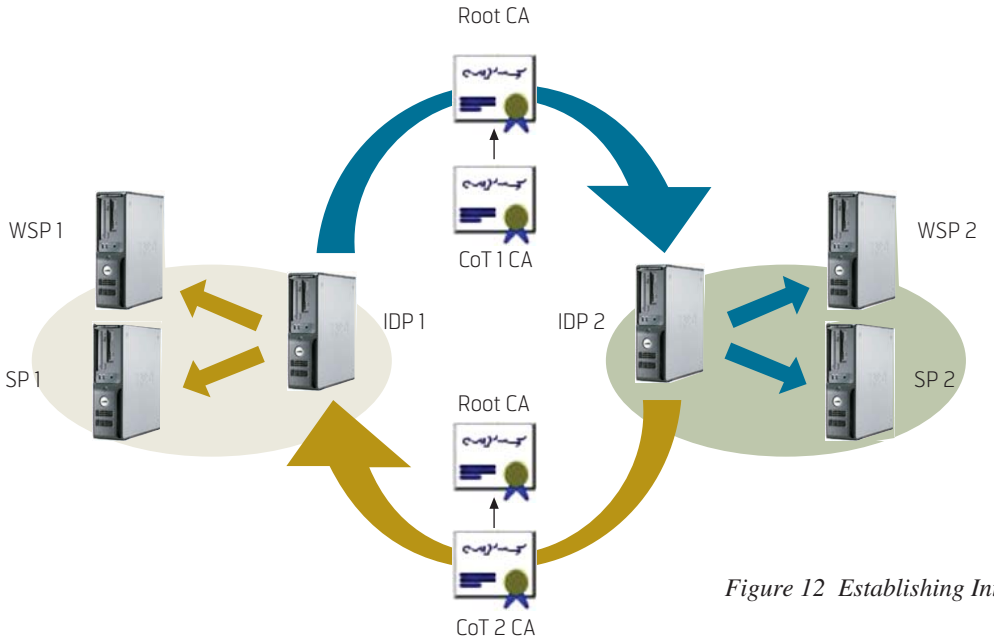


Figure 12 Establishing Inter-CoT trust relationship

ISSN 0085-7130

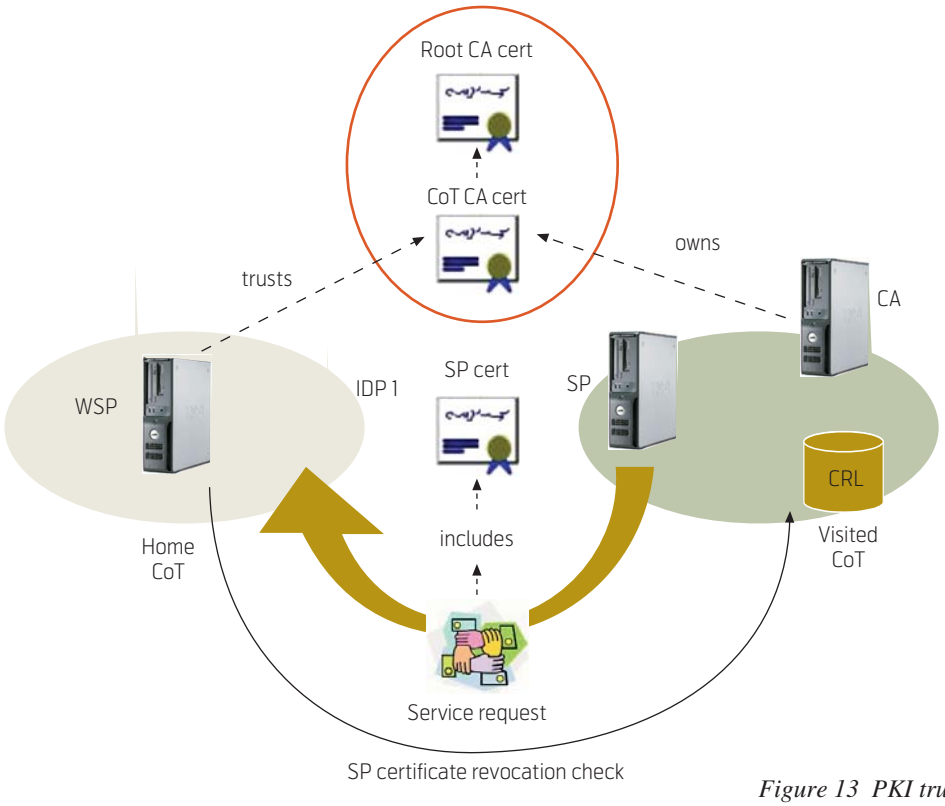


Figure 13 PKI trust model in action

SIS requires either product modification or non-negligible integration efforts, and then could not be qualified as plug&play.

This collection of ID-SIS is not always broad enough to cover all business needs and the definition/usage of new attributes may be required. The Fidelity Project has identified and implemented two distinct approaches to tackle this issue:

- *Complementing existing profiles with specific attributes*. Existing profiles defined by Liberty ID-SIS specifications may be extended (XML extensibility) with required business specific attributes. This approach seems to be suitable for small complements, for instance users' preference attributes may be added to Liberty ID-PP (Personal Profile).

- *Creating specific profiles*. This approach seems to be well adjusted for new data fields, but may lead to interoperability issues between WSC and WSP.

However, it must be noted that both approaches reduce the level of interoperability between Liberty components (WSC/WSP) which have to exchange information about the extension of the service or new services before sharing these new attributes.

According to scenarios, the Fidelity Project has used these two approaches:

- Fidelity extensions for ID-PP have been defined to include user's preferences (hotel room, game, food).

- Two new specific profiles have been defined:
  - A Liberty-compliant Wallet for the Fidelity Project. The Wallet service is a repository of data, including a set of payment means (banking cards, invoices, etc.) and a set of addresses.

  - A Liberty-compliant Calendar for the Fidelity Project. The Calendar service has been specified according to the format of xcal *vevent* structure (based on IETF draft document).

Although ID-SIS Personal Profile extension capabilities are defined in Liberty specifications, this mechanism does not ensure interoperability between implementations. Indeed, the description of attribute extensions is not detailed enough and may involve different interpretations by technology vendors, as noted during the Fidelity experimentation.

6 Security and Privacy

6.1 Privacy Considerations

Privacy is one of the major concerns of the users of *e*Services, mainly in an international scenario. Being aware of this, the Fidelity Project has analysed the resulting detailed requirements and the capability of Liberty Alliance protocols and services to provide the adequate security mechanisms to guarantee the ex-

ISSN 0085-7130©Telenor ASA 2007

pected levels of privacy, depending on the business scenario and the type and quality of the data being processed in the service provision.

6.1.1 European Privacy Requirements

Privacy requirements are relevant all over the world, but in Europe we may find some special constraints, which may introduce additional requirements, coming mainly from the fact that there is some specific legislation on this subject, and also from the special concern of citizens to keep their freedom to preserve their privacy in their relationships with service providers.

These two viewpoints, legal and societal, are analysed, in addition to a third one, whereby the privacy requirements are seen from the enterprise viewpoint, stating what should be taken into consideration by service providers to keep their liability over the actions performed by users and service providers on the right level.

The following privacy requirements have been identified by the Fidelity Project:

- *Legal framework*. EU Directives, relevant to CoT operations *(Personal Data Privacy, eSignature, eCommerce, Telecommunications)*, determine the minimum mandatory constraints about service provision as a set of basic Best Practices that regulate the interaction between the actors, of the information and of the services.

- *Personal Data types*. Personal should be classified into one of the five following categories according to its level of privacy: Private, Financial, Personal, National and Shared. This level of privacy determines the appropriate security level (basic, medium or high) as described by EU directives.

- *Principal's consent*. According to the types of data (and related levels of privacy), it requires *explicit consent* (private data), *unambiguous/tacit consent* (financial and personal data), or *no consent* at all (national and shared data).

- *Principal's rights*. User must be granted access, cancellation, opposition and rectification on data storage and processing. User rights concerning stored attributes require setting up of adequate incident prevention and recovery mechanisms, as well as control of the operations performed with personal data by other service providers.

- *Security Policy adoption*. All security mechanisms that have to be applied to adequately protect the data, have to be described and adopted following a Security Policy. The Security policy must be defined, applied, controlled and evaluated regularly.

- *Data transfer control*. User's consent and authorisation of the Control Authority (public organisation at each member state) is required for data transfer to third countries, when there are no bilateral agreements or risks for rights/freedom of the Principal.

- *Roles and liability of Inter-CoT functional components*. Functional components may play one of two possible roles: Controller (collection, storage and custody of personal data) or Processor (processes personal data on behalf of the controller). Business obligations between controller and processor must be regulated by a contract which must state at least: scope of the functions of each one of them when processing the data; liability in front of the Principal (applied privacy mechanisms, personal data rights management, processing liability transfer to third parties), that each one is obliged in front of the other to fulfil totally the legal obligations.

6.1.2 Liberty Privacy Capabilities

Privacy and security are key concepts in the Liberty specifications. On a global scale, the goal of the Liberty privacy policy framework is to enable the exchange of attribute information under end-user control and knowledge of the requestor's privacy policy. As a consequence, the Liberty specifications provide a list of mechanisms into their frameworks enhancing the privacy of the Principal and implementing good privacy practices.

The Liberty ID-FF 1.2 framework defines or enables some privacy-enhancing mechanisms which allow Principal's choice and control:

- *Federation:* User anonymity may be granted through the usage of opaque federated identifier and unique identifiers for each SP-IDP or IDP/IDP pair.

- *Single Sign-On (SSO):* Confidentiality, integrity and authenticity of information ensured thanks to technical means, exchange of a minimum set of authentication information, pseudonymous and anonymous access capabilities.

- *Single Logout (SLO):* SLO handled by IDP only (no direct links between SP).

- *Identity Federation Termination Notification:* Pseudo-random federated name identifiers used with the IDP to allow keeping distinct local identi-

ISSN 0085-7130

fiers within one SP, preserving anonymity once the federation has been terminated.

- *Cookies:* Common Domain Cookies must not have any personally identifiable information or authentication information.

Liberty ID-WSF 1.1 provides the framework to build interoperable identity services, permission based attribute sharing, identity service description and discovery, and the associated security profiles. To accomplish that, Liberty-enabled entities must build a trust relationship, based upon two key concepts: authentication and authorization. Whereas the ID-FF framework dealt with authentication issues, authorization issues are covered by ID-WSF. Best practices and available mechanisms are the following:

- *Secured communications:* Security mechanisms of Liberty ID-WSF specifications help to ensure confidentiality and integrity.

- *Usage Directives:* Enhance privacy by allowing requesters to designate the use they intend for requested data and providers to designate the permitted uses of released data.

- *Resource Offering:* Inform the WSC about privacy constraints (usage directives); ResourceID is sent encrypted using an encryption key that, for privacy reasons, must exhibit nonce-like characteristics.

- *User consent:* To collect attribute values or to obtain permission to share the data with a WSC; Principals may have expressed *explicit* consent, claimed through a Consent header message, or *implicit* consent, through unlimited authorisation to share the data with a specific type of WSC.

- *Attribute Storage:* A Principal may have several identities and may store his/her attributes in various and independent trusted APs, whose access is restricted and checked.

- *Anonymous Service Requests:* Opaque handle provided by trusted third parties allow keeping secret the Principal's name identifier for the WSC and any subsequent intermediaries (creation and consumption of encrypted identifiers).

It must be underlined that Usage Directive mechanisms really enhance user privacy and his control over his attributes, thanks to the definition of Privacy Policies. However, as the policies are defined in each CoT-specific policy language, this Usage Directive mechanism may, in practice, be hard to implement in an Inter-CoT context, which aims at a loosely coupled and plug&play trust relationships. An automated process like Service Level Agreement (SLA) may be a good solution to achieve automatic privacy policy agreements between CoTs.

The Liberty ID-SIS framework provides a common data structure with a list of attributes that must be interchanged between the actors of a Liberty empowered service provision environment. This list improves Principal's privacy, since only the necessary and requested information is disclosed.

6.1.3 Privacy Context

Although privacy issues for personal information are well assured within the Liberty Alliance specifications, the nature of such information cannot be clearly specified, implying that all data is considered with the same level of privacy. This is not compliant with European privacy laws and recommendations for such information.

To fulfil these requirements, the Fidelity Project proposes to take advantage of the already defined Liberty Authentication Context as a basis for a special *Privacy Context*. These "Privacy Contexts" may define three different levels of privacy (high, medium and basic), so information could be treated, from a privacy point of view, in a different way depending on its importance or its nature:

- *High*. Private data relative to religion, ideology, trade unions membership, sexual life, race and health.

- *Medium*. Data relative to administrative and penal infractions, economic solvency, fiscal information and users' personality profiles.

- *Basic*. All the rest of personal data.

Guidelines should be established to find the best way to integrate this privacy context into the actual Liberty specifications. This Privacy Context should be taken into account by the different Liberty entities involved in the negotiations of the security mechanisms and policies used to retrieve, store and manage the personal data, as well as its persistency, through Business and Service Level Agreements.

6.2 Inter-CoT Threats and Security Evaluation

The Fidelity Project has described some scenarios to test the security of the Inter-CoT data transfers, the testing tools and processes to be set up in order to check the adequacy of the security mechanisms implemented in the interactions between the CoT and the external world. Especially, the Fidelity Pro-

ISSN 0085-7130©Telenor ASA 2007

ject focused on the categories of threats, defined by Liberty Alliance, which are more relevant to account takeover in Inter-CoT scenarios, where someone gains access to the victim's existing accounts: user identity theft and user session hijacking.

The complexity and variety of the tests to be performed in order to guarantee the adequate implementation and installation of the Liberty components, and the fact that these tests should be performed each time a new component is incorporated in the CoT, suggest the convenience to well document and automate these tests in order to save time and guarantee adequate and uniform quality.

6.3 Protection Profile

The Fidelity Project aimed to identify the specific security threats that system integrators of Liberty Alliance protocols may need to avoid when implementing a multi-national scenario, with Inter-CoT communications. Thus, the Fidelity Project has described the Target of Evaluation (ToE), its potential threats and the security objectives to be addressed, providing the guidelines to build a protection profile of a Circle of Trust (CoT), according to the ISO-15408 (Common Criteria). These guidelines address exclusively the security aspects covered by the Liberty Alliance protocols, and the specific use cases implemented in the Fidelity Project. This restriction leaves for further work the identification and description of the security aspects common to many Information Systems management practices, or to the data communication within a CoT, where specific network security controls may be applied, outside the scope of the Liberty Alliance protocols.

The ToE is basically a CoT, with all the typical components: Identity (IDP), Attribute (AP) and Service Providers (SP), Discovery (DS) and Web Services (WS). All of them interact directly or indirectly with end-users, other IDPs and other WSs. The formal description of the ToE follow the guidelines of a medium level Protection Profile in ISO 15408 (Common Criteria), for security evaluation:

- Its specific and common threats;
- The security objectives to be achieved;
- The potential common and specific attacks to CoT components, coming from agents not belonging to the CoT;
- The functional requirements on the components and their interactions;
- The capabilities of Liberty Alliance security mechanisms to prevent those attacks and to achieve the security requirements.

The aim of the security analysis of an Inter-CoT communication is that Liberty Alliance Protocols provide adequate security mechanisms to preserve users' identity credentials and attributes, and that testing of its adequate use is feasible with limited effort and affordable tools, following the ISO-15408 Common Criteria methodology.

7 Conclusion

In summary, the Fidelity Project has achieved the following:

- *From a technical point of view:* Fidelity successfully deployed a "real-life" infrastructure, using heterogeneous products to implement all of the Liberty functionalities.
- *From an economic point of view:* Fidelity, through the deployment of close to market use cases, proposed a business model for Identity Management in which a Telco company is Identity Provider.
- *From a European regulation point of view:* Fidelity provided recommendations regarding the European laws protecting the user's privacy.

The Fidelity Project provided the clue that there is no technical barrier setting up a community of Circles of Trust. The effort that all the partners put into dissemination activities has persuaded our business units and partners that Identity Management is a key feature of Internet *e*Business and *e*Government coming out very soon. If Identity management solutions are numerous, Fidelity partners are convinced that the Liberty approach is the most relevant and this exploitation plan showed that some of the partners have already began to use the results of Fidelity Project for R&D and commercial purposes.

Resources

1 *Liberty Alliance*. September 21, 2007 [online] – URL: http://www.projectliberty.org

2 *Liberty Alliance ID-FF v1.2 specifications*. September 21, 2007 [online] – URL: http://www.projectliberty.org/resource_center/specifications/liberty_alliance_id_ff_1_2_specifications

3 *Liberty Alliance ID-WSF v1.1 specifications*. September 21, 2007 [online] – URL: http://www.

ISSN 0085-7130©Telenor ASA 2007

projectliberty.org/resource_center/specifications/liberty_alliance_id_wsf_1_1_specifications

4 *Liberty Alliance ID-SIS v1.0 specifications*. September 21, 2007 [online] – URL: http://www.projectliberty.org/resource_center/specifications/liberty_alliance_id_sis_1_0_specifications

5 *Liberty ID-FF Bindings and Profiles Specification v1.2*. September 21, 2007 [online] – URL: http://www.projectliberty.org/liberty/content/download/319/2369/file/draft-liberty-idff-bindingsprofiles-1.2-errata-v2.0.pdf

6 *Liberty Metadata Description and Discovery Specification v1.1.* September 21, 2007 [online] – URL: www.projectliberty.org/liberty/content/download/1224/7973/file/liberty-metadata-v1.1.pdf

7 *Liberty Authentication Context Specification v1.3*. September 21, 2007 [online] – URL: http://www.projectliberty.org/liberty/content/download/1208/7924/file/liberty-authentication-context-v1.3.pdf

8 *Liberty ID-FF Architecture Overview v1.2*. September 21, 2007 [online] – URL: http://www.projectliberty.org/liberty/content/download/318/2366/file/draft-liberty-idff-arch-overview-1.2-errata-v1.0.pdf

Dao Van Tran obtained his MSc in Informatics from the University of Oslo. He began his work in Televerket/ Telenor in Drammen Tele District in 1986, and in 1996 he joined Telenor R&I. He is now involved in the Eureka project Mobicome, which focuses on IMS in a fixed mobile convergence.

email: dao-van.tran@telenor.com

Pål Løkstad obtained his MSc in Applied Physics from the University of Tromsø (UiTø) in 1995. He worked as assistant lecturer at UiTø until he started work as a developer in Moesarc Technology in 1996. In 1998 he joined Telenor R&I. In Telenor his main research activities have been within SIP, IMS, Web Services, Service Composition, Mobile Terminals and development of demonstrators.

email: pal.lokstad@telenor.com

For a presentation of Do Van Thanh, please turn to page 2.

ISSN 0085-7130©Telenor ASA 2007

Identity Management in Telecommunication Business

DO VAN THANH, IVAR JØRSTAD, DO VAN THUAN, NICOLAY BANG

Do Van Thanh is Senior Research Scientist in Telenor R&I

Ivar Jørstad is CEO of Ubisafe AS

Do Van Thuan is Lead Scientist in Linus AS

Nicolay Bang is Manager of Linus AS

The number of electronic identities, i.e. user names and passwords that each person has, is increasing everyday and the situation will soon be unmanageable. The Liberty Alliance proposes a solution based on federated network identity, which alleviates the user's burden of account administration and offers single-sign-on. However, without a well balanced and sound business model, the Identity Federation can never be a reality. This paper explores in detail the business scenarios, which are both attractive and realistic to the telecom operators. The paper concludes with some recommendations about deployment and suggestions for future work.

1 Introduction

The success of the Internet and especially the World Wide Web is mostly due to the ability to convey information abundantly, easily and directly. It is simple for anyone both to publish and to access information, which can be presented in varied forms like text, picture, logo, animation, video, etc. The goal of these information services is to reach as many people as possible. Neither selection nor filtering is required. However, for more advanced services it is often necessary to have partial or total knowledge about the user in order to deliver the required service. Examples of such services are flight ticket booking, hotel reservation, car renting, music shopping, user group participation, etc. These services are usually offered by different service providers, and the user ends up with a large number of accounts and passwords. There is no indication that this trend will stop, and the situation will soon be unbearable for the user. To remedy the situation, the Liberty Alliance Project, an alliance of more than 150 companies, non-profit and government organizations from around the globe, is committed to developing an open standard for the management of federated network identities that supports all current and emerging network devices. But, no matter how technically good the federated network identity could be, its success relies on compelling business. In this paper, business scenarios of interest for telcos will be presented and examined carefully. The paper starts with a brief introduction of the Federated Network Identity concept.

Identity management has been used intensively in telecommunications since its start. Indeed, in order to establish a call from one person to another, the identity of the callee must be known.

2 The Federated Network Identity Concept

2.1 Overview

As stated in [1], [2], [3], *network identity* refers to the global set of attributes that are contained in an individual's various accounts with different service providers. Currently the user's network identities are like isolated islands and the user is responsible for remembering numerous usernames and passwords for each of these identity islands. The user will typically either try to always use the same password or to record the passwords somewhere. Either way, the result is a drop in the level of security.

The most logical solution to the problem caused by the isolated network identity is to build bridges that interconnect them and allow information flows between them. This is precisely what *"Federation"* does.

Federation refers to the technologies that make identity and entitlements portable across autonomous policy domains. Consequently, the *Federated Network Identity* is a portable identity.

The establishment of federated relationships between service providers will hence allow the users to move seamlessly from one service provider to another. However, if every service provider has to make alliance with each of the other service providers it will be time consuming and require tremendous efforts. For n service providers, it requires $n(n\text{-}1)/2$ established relationships.

To circumvent this problem, the Liberty Alliance proposes a new role called *Identity Provider*. The Identity Provider assumes the management of the user's Federated Network Identity and the user authentication.

ISSN 0085-7130

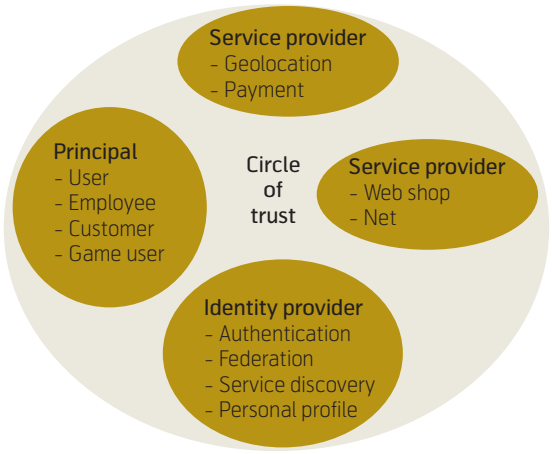


Figure 1 A Liberty Alliance Circle of Trust

A *Circle of Trust* (CoT) is a group of service providers and identity providers that have business relationships based on Liberty architecture and operational agreements and with whom users can transact business in a secure and apparently seamless environment.

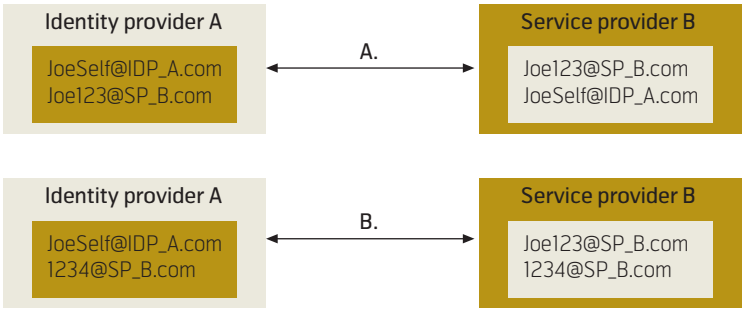


Figure 2 Identity Federation schemes

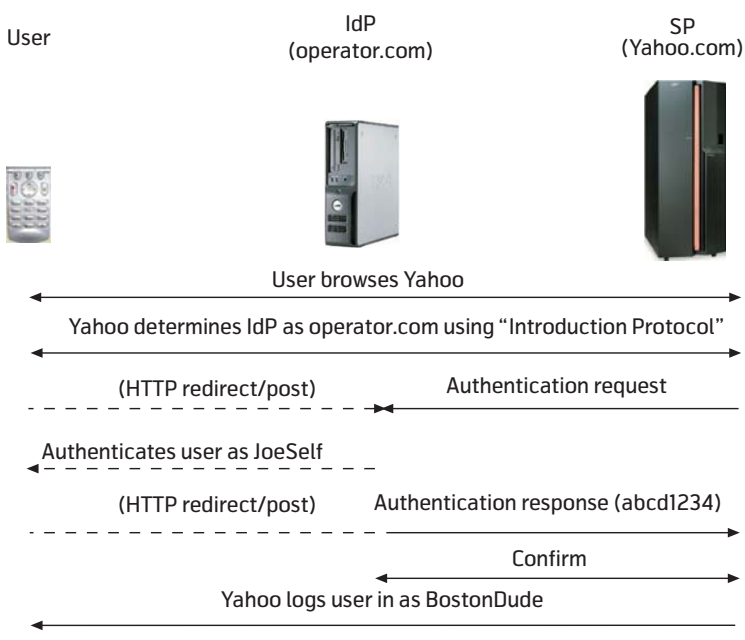


Figure 3 Single-Sign-On

Figure 1 shows a Circle of Trust. The *Principal* is the user, employer, customer, game user, etc. whose Network Federated Identity is managed by the Identity Provider.

2.2 Federation

To federate network identities, one can simply define direct associations between them. This can be done as shown in Figure 2 part A, where both the Identity Provider and Service Provider store the user name that the user has at the other party, but the anonymity and privacy of the user are seriously compromised. To solve this issue as shown in Figure 2 Part B, an opaque handle "1234" is used by both parties to address to the same user JoeSelf without having to know the user name at the other party.

2.3 Single-Sign-On User Experience

Once federation is done, the user can enjoy Single Sign On. As shown in Figure 3, Joe has logged in at his IDP as JoeSelf. When visiting Yahoo.com, he is "automatically" logged in as BostonDude.

3 The Business Scenarios of Interest for Telcos

The following business scenarios are identified:

- Telco as Identity Provider (IDP)
- Internal enterprise
- Inter-CoT
- Mobile-fixed collaboration.

Each scenario will successively be studied in more detail.

3.1 Telco as Identity Provider

3.1.1 Short Description

To be an Identity Provider is a natural role for telcos as they already possess and manage large numbers of customer identification and authentication data in regards to their own systems and services:

- Telcos are already *service providers* for their own value added services.
- In many cases telcos also act as *identity issuers*, either for their own services or increasingly to 3rd parties.
- Telcos hold a large amount of information on their customers, which enables them to act as an *attribute provider* also.
- *Directory services* could also belong to the service portfolio of telcos.

ISSN 0085-7130©Telenor ASA 2007

By managing all these roles, it is possible to shorten and simplify the value chain and thus reduce costs as fewer middlemen are taking part in the business.

The revenue model is as follows:

End user fees: might consist of a monthly fee and some extra fees if the end user subscribes to value added services or attributes, such as positioning information. These extra fees might be packaged to the monthly fee or be charged per transaction. The most important issue is that the end user is aware of the pricing model and the pricing is transparent. Especially this must be noticed in roaming usage when visiting a foreign service provider. The value of the fees will be related to the added value that the service will provide and the will of the user to pay these fees for the service.

Service/Content Provider fees: the pricing could be based on several options:

- CoT fee for the membership in a certain CoT, charged periodically (for example monthly) and covering the registration, maintenance and directory services;
- Transaction fee for each end user identification;
- Roaming fee for the possibility to enable foreign usage;
- Registration fee for each end user registration to SP/CP's service;
- Possible promotional charges if applicable.

SP/CP fees could be easily justified by the savings they get when they do not need to make interfaces and systems to several identity providers as the situation is today. Also the contracting is easier with one single party.

3.1.2 Participating Actors

The actors identified in this business scenario are:

- The Identity Provider, which is the telcos;
- The Service/Content Provider;
- The end-user, who is the telecom subscriber.

3.1.3 Value Proposition

Value to the Telco

In a time of deregulation and fierce competition, the revenues of the telcos are diminishing every day. By becoming an IDP, a telco will obtain a new source of revenues, and at the same time customer loyalty will be improved considerably thanks to the improvement of the services offered to the user in terms of quality, user-friendliness, cost and variety.

Value to the Service Provider

By joining a Circle of Trust the Service Provider will reach a larger customer base without having to put efforts and resources into advertisements or recruitment campaigns. Costs can be saved by sharing development efforts with other organizations. Risks and liabilities are mitigated. Investments in authentication infrastructure and services can be minimized within the Circles of Trust. Integration effort is minimized and time-to-market is reduced. New services can be introduced; for example, third-party billing, based on events, can be supported.

Value to the End User

For the end user who is the telecom subscriber, the most important value is to have access to a larger, more diversified and more geographically distributed services. Next is naturally the quality of the services that can be personalised and adapted to the context of the user. Last but not least, the burden of having to manage multiple accounts and password will be considerably alleviated.

3.2 Internal Enterprise

3.2.1 Short Description

Medium and large enterprises usually consist of several business units or departments that are operating more or less autonomously. There is a big demand for information sharing between these units at the same time as autonomy prevails. In addition to these two conflicting requirements, the operation and management must also be easy, flexible and dynamic.

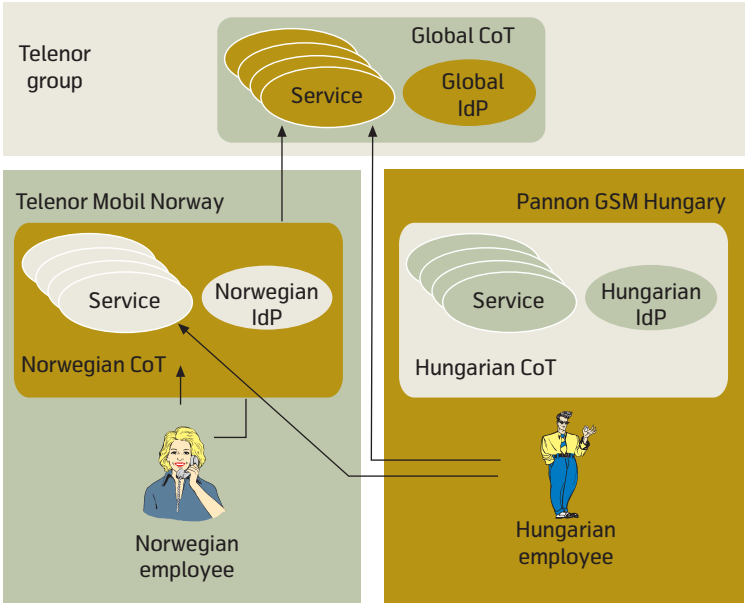


Figure 4 Internal Enterprise Business Model

ISSN 0085-7130

To elucidate on the Internal Enterprise scenario let us now take the example of Telenor. Telenor is the largest provider of telecommunications services in Norway, and has substantial international mobile operations in 12 countries in Europe and Asia.

It is desirable that employees in one Telenor company can have controlled access to the information and services from the other companies. To simplify, without losing generality, let us consider the collaboration of only two Telenor companies: Telenor Mobil Norway and Pannon GSM Hungary (see Figure 4).

The following basic high level requirements are identified:

1 An employee of Telenor Mobil Norway should be able to access the defined services at Pannon GSM Hungary.

2 An employee of Pannon GSM Hungary should be able to access the defined services at Telenor Mobil Norway.

3 The services made available to the partners' employees are defined by the agreement between the two partners.

4 Each partner should be able to add, remove or alter the access rights of its employees at any time and the modification should come into effect immediately.

As shown in Figure 2 every employee of the Telenor Group should have access to the global services in addition to the services offered by its companies. In addition, two Telenor companies may also set up agreement that allows certain types of their partners' employees to have access rights to a defined set of services. This set of services could be modified depending on the agreement between the companies. The administration of the employee, i.e. add, remove or modify the access right, is still fully assumed by the employer company.

Three Circles of Trust (CoT) are established for the Telenor Group, Telenor Mobil Norway and Pannon GSM Hungary respectively. Each CoT contains an IDP (Identity Provider) who acts as an "Authentication authority" and as many SPs (Service Provider).

3.2.2 Participating Actors

The actors identified in this business scenario are:
- The enterprise business units (Telenor company)
- The enterprise employee
- The partner's employee.

3.2.3 Value Proposition

The Internal Enterprise business model brings value not only to the Telenor employees who are the end-users but also more importantly, to the Telenor companies and partners.

Value to the Enterprise Employee

As shown in Figure 3, a Hungarian employee will have access to both the defined services at the global CoT and the Norwegian CoT. The user will experience the following:

1 The Hungarian employee browses on the web and visits the Telenor Mobil Norway portal. He/she clicks to access a Norwegian internal service.

2 This internal service needs to authenticate the user and sends an authentication request to the Norwegian IDP.

3 The Norwegian IDP detects that the user originates from the "trusted" Hungarian CoT. It turns to an SP and requests the Hungarian IDP to do proxy authentication. The result of this authentication is a federation key provided by the Hungarian IDP to the Norwegian IDP.

4 The user is granted access to the Norwegian internal service.

Seen from the Hungarian employee, the value of this business scenario lies in the simplicity and user-friendliness. He/she does not have to remember several login names and passwords and the access to the Norwegian internal services is seamless or identical to the access to the Hungarian internal services.

Value to the Enterprise Business Units

The values to the Telenor companies are as follows:

Through agreement, they can decide which internal services should be made available to which types of employees from the other companies. Controlled information sharing will strengthen the expertise of the employees, which reduces training cost at the same time as it increases efficiency.

Each company starts or terminates their employment as it suits them. This helps to sustain autonomy which is necessary for the adaptation to the local cultural, economic and judicial conditions of each country.

The employees' access right can be modified dynamically and the modifications are brought into effect immediately. This is very important since prolonged access to a former employee is a common major security breach.

ISSN 0085-7130©Telenor ASA 2007

Each CoT should not maintain authentication information related to the user (login and password) for each employee that can access each CoT, but the user trusts the authentication of only one CoT (his home CoT).

3.3 InterCoT Business Model

3.3.1 Short Description

In this business model, a CoT is connected to other CoTs to allow users and service providers in one CoT to interact with users and service providers in other CoTs. This configuration is quite similar to GSM roaming [4] allowing travelling users to access services at visiting locations. IDPs in different CoTs have an agreement and the complexities are hidden to both the SPs and the users.

Identity federation between Inter-CoTs enables the end user to

1 Use the user's preferred identity for authentication,
2 Federate user's identity and attributes.

For the visitor service provider Inter-CoT offers:

1 Authentication of the user,
2 Getting the user's assertion,
3 Fulfilling the Inter-CoT business model requirements.

3.3.2 Participating Actors

The following actors can be identified from this business model:

- End user
- Home IDP
- Visitor IDP
- Visitor Service Provider
- Home Attribute Provider.

3.3.3 Value Proposition

Value Proposition to the User

Reach more services without having to establish new accounts.

Value Proposition for SPs

- Larger available user market;
- A Liberty Circle-of-Trust for operators lets Service Providers offer their services to all possible users (regardless of their choice of operator) through a single interface;
- Lower Cost-of-Business;
- A standard developer platform for using operator services (i.e. payment, messaging applications, etc.) lowers the integration effort and accelerates time-to-market;
- The service provider can rely on an identity provider and 'outsource' the overhead of user authentication. This allows the SP to focus on the value it provides users;
- New Levels of Customer Care.

Value Proposition for IDPs

- Reach more service providers;
- New authentication revenue.

3.3.4 Market Segment

The Inter-CoT business model can be applied in many ways and the target customers can be end-users, service providers, enterprises and governments.

3.3.5 Value Chain Structure

To realise the Inter-CoT business model, several CoTs must be established and an architecture must be elaborated. If a hierarchical architecture is chosen the positions of the different CoTs in relation to each other has to be defined. In a flat structure, all the CoTs are at the same level and federated together. The roles of the IDPs are symmetric since they are both home IDP and visitor IDP.

3.4 Mobile-Fixed Collaboration

3.4.1 Short Description

To increase loyalty and to reduce churn, it is quite crucial for telecom operators to be able to offer total service solutions, which include mobile, fixed and broadband services. However, not all telecom operators are complete operators but can be specialised operators. Even for complete operators, it may be difficult to offer total service solutions due to the differences in terms of technologies, organisation, policies, etc.

The requirements are as follows:

- It should be possible to offer the users a total service offering that includes mobile, fixed and broadband services operated by different telecom operators or units.
- It should be possible to offer the bridging for users who already have separate subscriptions at each operator such that they look like a unified subscription.
- It should be possible for the users to seamlessly access services operated by the federated operators without having to authenticate themselves.

The Liberty Alliance specifications propose a simple but feasible way to enable the telecom operators to

ISSN 0085-7130

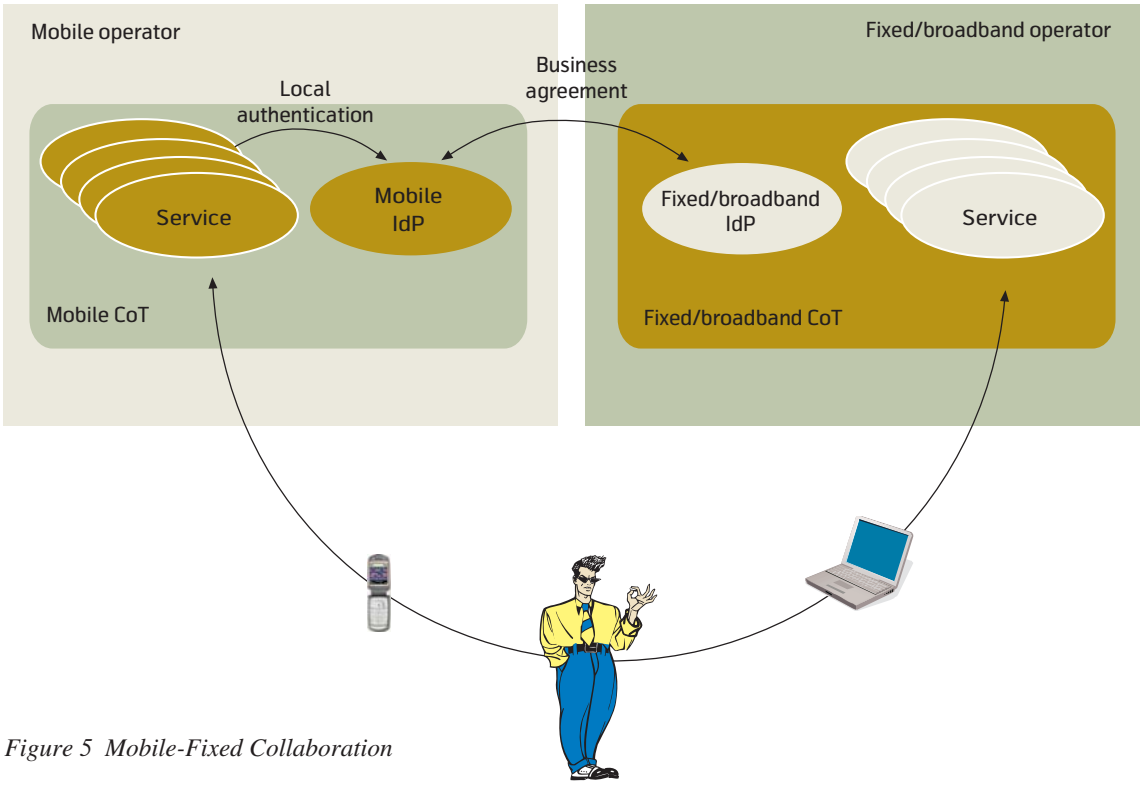


Figure 5 Mobile-Fixed Collaboration

offer total service solutions. Each complementary telecom operator or complementary business of an operator like mobile, fixed or broadband operates as an Identity Provider and together with their Service Providers and Content Providers form their Circle of Trust.

These CoTs are bridged through business agreements. The following requirements are fulfilled:

- The users having subscriptions to multiple CoTs prior to the business agreement shall have the possibilities to federate their accounts and benefit single-sign-on.

- The users having subscriptions to one or some CoTs both before and after the business agreement shall also have access to services offered by other CoTs in the business agreement without having to register to these CoTs or to register to the SP/CP in these CoTs.

Figure 5 shows an example of federation between mobile and fixed/broadband CoTs. A user having subscription from one of the two operators will have access to the services offered by both operators. The service range that a user can enjoy is extended to the sum of the service ranges.

3.4.2 Participating Actors

The actors identified in this business scenario are:

- The Mobile operator
- The Mobile user
- The Mobile Service Provider/Content Provider
- The Fixed/Broadband operator
- The Fixed/Broadband user
- The Fixed/Broadband Service Provider/Content Provider.

3.4.3 Value Proposition

Value to the End Users

The end users will benefit from the following:

- To have access to a considerably larger range of services offered by all the operators;

- To have much simpler account management, i.e. not having to remember several user names and passwords;

- To have improved protection in terms of privacy intrusion and attacks.

Value to the Mobile/Fixed/Broadband Operator

Both the Mobile operator and the Fixed/Broadband Operator will benefit from the following:

- To obtain more loyal customers and reduce churn from the extended range of services that they are able to offer;

- To strengthen the new role of Identity Providers that they assume;

- To get additional revenues due to the visit of partners' subscribers.

ISSN 0085-7130©Telenor ASA 2007

Value to the Mobile/Fixed/Broadband Service Provider/Content Provider

For the Service Providers/Content Providers federated in either the Mobile CoT or the Fixed/Broadband CoT, the benefits are as follows:

- To reach a substantially larger customer base due to the union of the customer bases;
- To avoid having agreements with multiple Identity Providers that can be both time and cost consuming.

Conclusion

According to the study done in this paper, the identity management solution proposed by the Liberty Alliance based on network identity federation has several sound and compelling business models for the telcos. However, for the telcos, identity management is a new business that needs time to mature. One deployment suggestion is to start with the Internal Enterprise business model. When sufficient knowledge and experiences are acquired, the telco can take the role of IDP. The next step can be to establish collaboration with other IDPs, national or international. It is worth noting that for the Inter-CoT business model it may be necessary to have an international standardised agreement as the GSM roaming agreement.

References

1 Liberty Alliance Project. *Liberty ID-FF Architecture Overview, ver. 1.2-errata-vi.0*. September 12, 2004.

2 Liberty Alliance Project. *Identity Systems and Liberty Specification Version 1.1, Interoperability*. Technical Whitepaper, February 14, 2003.

3 Liberty Alliance Project. *Introduction to the Liberty Alliance Identity architecture, Revision 1.0*. March, 2003.

4 The GSM World. *Roaming*. August 31, 2007 [online] – URL: http://www.gsmworld.com/roaming/index.shtml

For a presentation of Do Van Thanh, Ivar Jørstad, and Do Van Thuan, please turn to page 2, 10, and 18, respectively.

Nicolay Bang has a degree in Physics, CandScient from the University of Oslo. He first joined ABB where he worked on Emergency Shutdown Systems for the Oil and Gas industry. He is the manager and one of the founders of Linus AS, a Norwegian consulting and systems house specialising in products for mobile communications and custom development and system integration for telcos. Nicolay has participated in the ADPO and Fidelity projects and is currently involved in MOBICOME, another EUREKA project focusing on Fixed Mobile Convergence and IMS. In addition to project administration activities, his contributions are related to business processes and value chains.

email: n.bang@linus.no

ISSN 0085-7130©Telenor ASA 2007

Strong Authentication for Internet Applications with the GSM SIM

IVAR JØRSTAD, DO VAN THUAN, TORE JØNVIK, DO VAN THANH

Ivar Jørstad is CEO of Ubisafe AS

Do Van Thuan is Lead Scientist in Linus AS

Tore Jønvik is Associate Professor at Oslo University College

Do Van Thanh is Senior Research Scientist in Telenor R&I

This paper presents an innovative service called SIM Strong Authentication that extends the usage of GSM SIM authentication to Internet Web services. The goal of this proof-of-concept is to demonstrate the possibility of implementing innovative services in a heterogeneous environment using the Liberty Alliance Federation Standard. Telenor, Gemalto, Linus and Oslo University College have implemented a proof-of-concept prototype in Oslo. The architecture is based on a multi-vendor environment where Sun provides the Identity Provider, Lucent Technologies the Radius server and Ulticom the SS7 MAP Authentication Gateway connecting the prototype to the Telenor mobile network.

A typical user flow for such a service would be the case of a user browsing on the World Wide Web from home, a customer premise, an Internet café, etc. When trying to access a protected resource such as Webmail, company portal, or bank account, he logs on to the requested secured site simply by placing his mobile phone close by and communicating with his PC via Bluetooth, or using a SIM card-equipped dongle, card reader, or 2G/3G PC card. Alternatively, the user can choose to perform the SIM strong authentication using the SMS service on his mobile phone.

This service is available anywhere and can support any Internet service. It is ideal for services like Internet Banking, *e*Administration or enterprise internal web pages. The SIM strong authentication is both user-friendly and cost efficient, with a low deployment threshold. The technology is also capable of supporting other Smart-Card based identity services such as USIM (UMTS), certificate based schemes (e.g. TLS) and One Time Password schemes (OTP).

Introduction

The popularity of the World Wide Web continues to grow due to the abundance of information, services, commerce, and recreation that people enjoy from Internet based resources. However, in order to have access to the most useful information and services while keeping an acceptable level of security, users must remember more and more usernames and passwords. The number of username and password pairs continues to increase and will soon be a nightmare to users. Furthermore, the use of passwords as a means of authentication is not strong enough for services that require added security, like e-commerce, online banking, government portals, corporate Intranet access, IP telephony, etc. Stronger authentications are required but, unfortunately, are usually both costly and not particularly user-friendly.

Telenor, Linus, Oslo University College and Gemalto, in collaboration with Sun, Lucent Technologies and Ulticom, have designed and implemented a strong authentication service that is both cost-efficient and user-friendly. The idea is to extend the usage of the current SIM authentication used in GSM to Web services. Indeed, this is a step further than earlier work that uses SIM authentication for WLAN (Wi-Fi – EAP-SIM). The idea of making the mobile phone and its SIM a universal authentication token is compelling, since the mobile phone is so common nowadays, and the GSM network is currently the largest mobile network and is ubiquitous in much of the world.

This paper presents the Telenor SIM Strong Authentication Service. It starts by summarising the state-of-the-art solutions for strong authentication and their limitations. An overview of our SIM Strong Authentication Service followed by a scenario showing how our SIM Strong Authentication Service works will be depicted. The value brought to users and service providers will be identified. The business opportunities for the mobile operators are also analysed.

Finally, the paper will explain how the Liberty Alliance Framework can be used to leverage this SIM-based strong authentication solution in a heterogeneous, multi-vendor environment that bridges Internet-based services and the GSM network.

Limitations of State-of-the-Art Solutions

Passwords

As mentioned earlier, the most common authentication scheme today is based on passwords. It is both weak and not user-friendly due to its plurality. There are many issues connected to user password management, but from a security point of view, there are three main issues:

ISSN 0085-7130

- *User-friendliness:* It is always possible to propose systems with high security, but if they are not sufficiently simple and friendly, the user will find a way to bypass them.

- *Phishing (stealing a user's password by tricking them into giving their credentials away to the wrong party)*: Keep asking gently for a password from a user, and at some point he will give it away. The best known methods for phishing user passwords are either to reproduce an almost identical login page to the one the user is used to, or to pretend to be from customer service and request a password for some special operation. The main rule of phishing is "if you can lock a user for a reason" then he will be ready to give you all the passwords he knows to unlock the situation "current one, old one, one from another site ..."

- *Brain limit:* Typical users will only remember three to five logins/passwords. They will either reuse the same credential all over, creating a potential risk of correlation in-between service providers, or will stick the most secure one on a "Post-It" somewhere on a very well hidden place such as "under his keyboard."

To tackle the latter problem and other identity related issues, the Liberty Alliance [1] has promoted the concept of federated network identity that enables users to seamlessly jump from one service provider to another using Single-Sign-On, while warranting user privacy, an adequate level of authentication for the requested service and provider independence. However, while Liberty specifies how a service provider requests a given level of authentication, it does not normalize how the CoT authentication authority (i.e. Identity Provider) negotiates credentials with, or on behalf of the principal. The problem of weak authentication then remains unsolved, leaving room for user password Web phishing and Post-It leaking.

Stronger Authentication Schemes

There exist today several strong authentication alternatives that require the user to present at least two factors; i.e. something that you know (PIN, code or password), combined with something that you have (a smart card or an authentication token), or sometimes something that characterizes you (biometrics). The smart card or authentication token may carry One-Time-Password (OTP) or Public Key Infrastructure (PKI). These solutions bring sufficient protection both to users and service providers but, unfortunately, they all suffer from significant drawbacks:

- *Costly infrastructure:* Strong-authentication solutions require specialized security hardware (such as tokens and Smart Cards), dedicated software and IT server infrastructure. In addition, there is a cost related to the administration of the keys and certificates.

- *Lack of interoperability:* Strong-authentication solutions are quite often proprietary and do not operate with each other.

- *Poor structure:* They do not provide well-defined interfaces that allow integration with new applications or services.

- *Lack of scalability:* Most current solutions are stand-alone and it is very difficult to extend them to be a global solution that can be used by every user, everywhere and anytime.

- *Cost of deployment:* Not only do special devices have to be given to each user, but each service provider needs to be customized to support the specific API and handshake protocols specific to the chosen device.

Because of the cost of deployment, this solution has been mostly limited to protect access gates to a secure zone (typically a VPN for an enterprise).

Dynamic Passwords

One alternative addressing some of the mentioned issues is to provide users with dynamic passwords they can use to log in. The users do not have to remember them, and there is no risk of compromised passwords since they are used only once. All that the users need are mobile phones that are capable of receiving the passwords as SMS messages from the service provider. This solution is, however, not very user-friendly since the users have to type in the password. In addition, a system for generating dynamic passwords is also needed and may be costly.

Because of the lack of user friendliness, this solution cannot be used for day-to-day operation and is mostly limited to exceptional operations such as connecting to the Internet from a hotspot at an airport, hotel, petrol station, etc. However, despite the lack of user friendliness, such solutions are currently also used by many Internet banking services.

Our SIM Strong Authentication Service

A SIM-based strong authentication service that extends SIM card GSM authentication to Web services is proposed as a remedy to the situations described above. It can briefly be described as follows:

ISSN 0085-7130

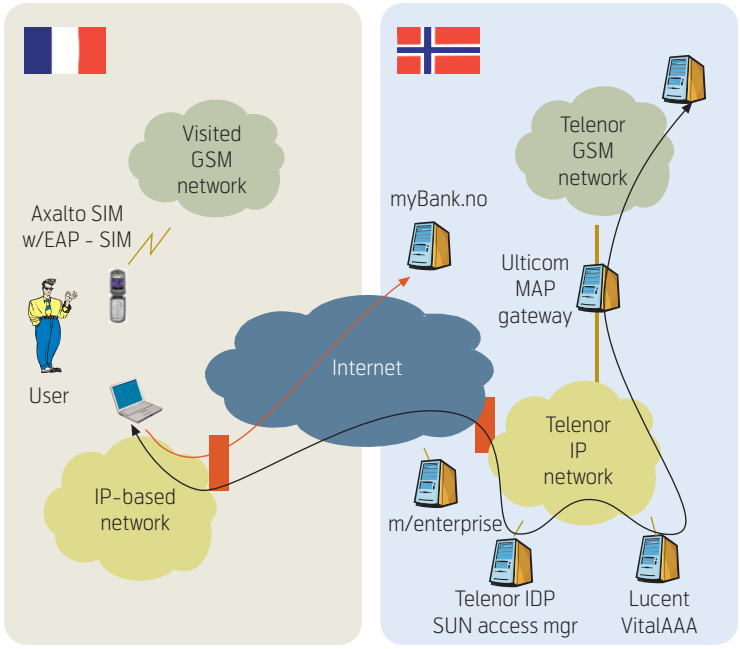


Figure 1 Overall architecture of the SIM strong authentication service

- A user with a valid Telenor mobile subscription having one of the following:
  - A mobile phone with a SIM and Bluetooth placed close to a Bluetooth-enabled PC
  - A dongle (with a SIM) mounted on the PC
  - A mobile phone with a SIM and SMS service
  - A Smart Card reader (with a SIM) installed in the PC
  - A GPRS/3G PC card (with a SIM) installed in the PC;
- May quite easily and securely log on to:
  - An Internet bank
  - A corporate intranet
  - A commerce web shop
  - An Enterprise website
  - An *e*Government application;
- At any time and anywhere in the world;
- The authentication is done by the Telenor Identity Provider (IDP) server based on Sun Access Manager in collaboration with a Lucent Technologies Vital*AAA* server that communicates with the Telenor Home Location Register (HLR) via an Ulticom Signalware SS7/IP MAP Authentication Gateway.

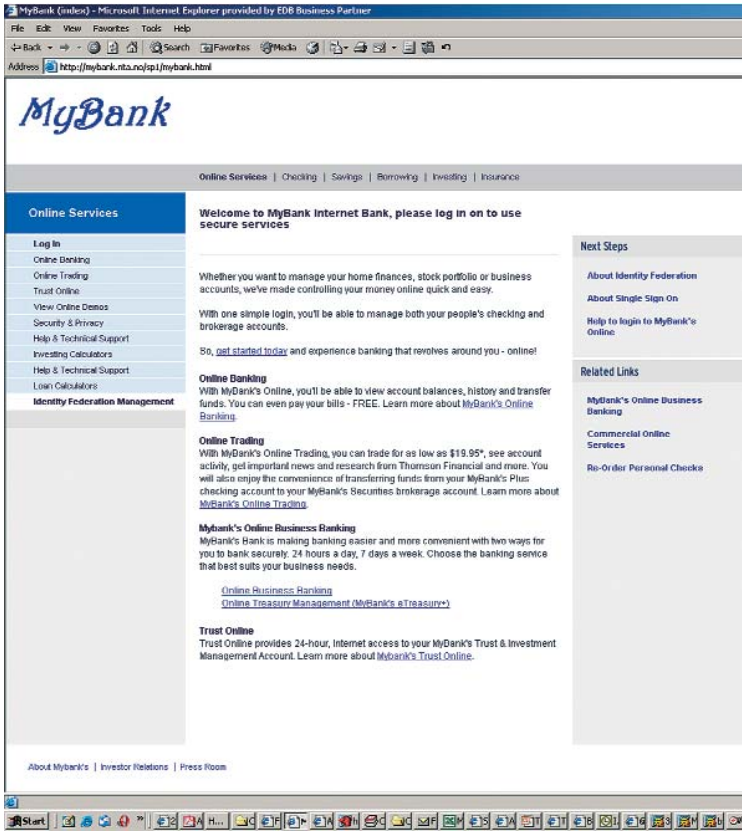


Figure 2 myBank website

The overall architecture of the SIM Strong Authentication Service proof-of-concept implementation is shown in Figure 1.

The advantages of the SIM Strong Authentication Service can be summarised as follows:

- Removes the user's burden of remembering passwords;
- Provides an authentication service that is both strong and easy to use;
- Allows rapid deployment due to the high penetration of mobile phones;
- Reuses existing GSM authentication structures (SIM card and HLR);
- Allows the integration of all services and applications;
- Uses open standards and supports interoperability with other systems;
- Provides scalability and supports a large number of users and service providers.

How Does the SIM Strong Authentication Service Work?

To illustrate how the SIM Strong Authentication Service works, let us consider the scenario of Kari, a user travelling abroad who attempts to log on to her Internet bank. Kari has a mobile subscription with Telenor and her mobile phone is equipped with a Gemalto SIM which supports the EAP-SIM protocol

ISSN 0085-7130©Telenor ASA 2007

provided by Telenor. Her bank myBank has a business agreement with Telenor concerning the usage of the SIM authentication service for its customers.

1. Kari connects her laptop to the Internet and visits the myBank website as shown in Figure 2.

2. When she attempts to log in, she is redirected to the Telenor Identity Provider website as shown in Figure 3.

Kari will now be presented with two options:
- Login using the SIM card;
- Login using SMS.

Login Using the SIM Card

3. Kari clicks on the "Logon with SIM Card" button. She is then asked to select one of the three authentication options (Figure 4):

- Using the SIM card in the card reader;
- Using the USB dongle or integrating the SIM card;
- Using the cellular phone with Bluetooth.

When ready, Kari clicks on "SIM logon" and the authentication begins.

4. A mutual authentication using EAP-SIM [2] is performed between the Telenor network and the SIM card (a detailed description of the authentication process is included later in this paper). Depending on the security settings Kari has established for her SIM card, she may be asked to enter her EAP-SIM card application PIN code to allow the mutual authentication to be performed.

5. After the successful authentication, the Telenor IDP redirects the browser back to myBank where Kari is now logged in and a Welcome page is displayed. Kari can carry out all her transactions.

6. After a while, Kari goes to her enterprise Intranet. This time she is automatically logged in since she has already been authenticated and that authentication is still valid.

Login Using SMS

3. Kari clicks on the "Logon with EAP-SMS" button and is presented with a web page asking for a session ID (Figure 5).

4. On Kari's mobile phone, a request to select SIM-Strong application pops up (Figure 6).

5. Kari selects the application and it displays the Session ID on the mobile phone screen (Figure 7).

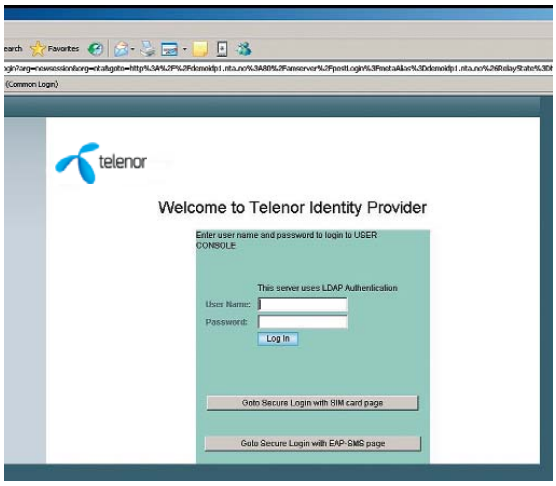


Figure 3 Telenor identity provider website

6. Kari types this Session ID on the web page and clicks on the "Send" button to initiate the authentication with the AAA server (Figure 8).

7. Kari is requested to click on the OK button of her mobile phone to initiate the authentication (Figure 9).

8. Kari is notified that the connected server is "secured" and has been authenticated (Figure 10).

9. Kari is now notified that her authentication is successful. The Telenor IDP redirects the browser back to myBank where Kari is now logged in and a Welcome page is displayed. Kari can carry out all her transactions (Figure 11).

EAP-SIM

EAP-SIM is a recognized EAP (Extensible Authentication Protocol) Type and is defined in an IETF draft

Figure 4 Selection of authentication token

ISSN 0085-7130

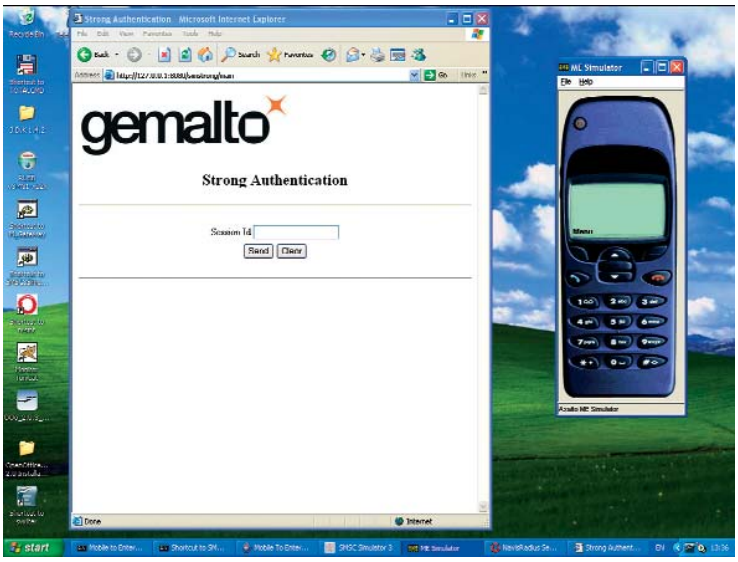


Figure 5 A session ID is requested

Figure 6 SIMstrong request pop-up on mobile phone

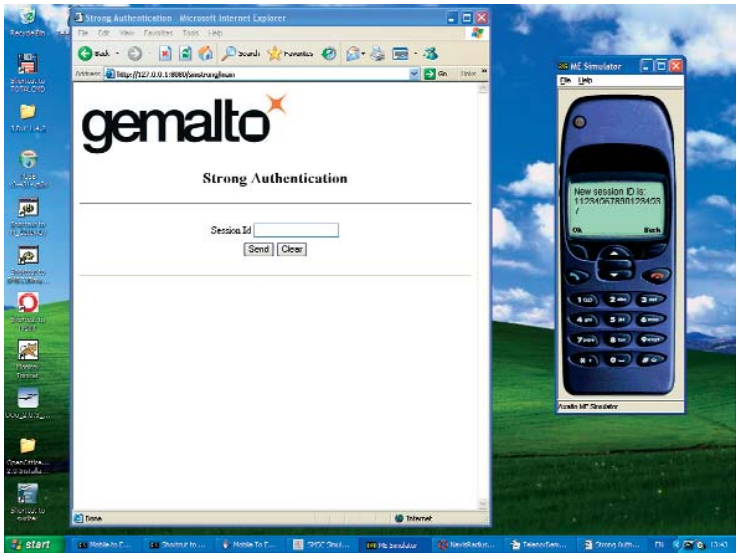


Figure 7 Session ID is displayed on the mobile phone

(draft-haverinen-pppext-eap-sim-16.txt). The EAP-SIM peer interface between the terminal and SIM is standardized by:

- ETSI in TS 102.310, and
- "WLAN Smart Card Consortium" in "WLAN-SIM-V11.pdf".

EAP-SIM specifies an Extensible Authentication Protocol (EAP) mechanism, called an EAP *Type*, for authentication and session key distribution using the GSM Subscriber Identity Module (SIM).

GSM authentication is based on a challenge-response mechanism. The A3/A8 authentication algorithms that run on the SIM can be given a 128-bit random number (RAND) as a challenge. The algorithm takes the RAND and a secret key K_i stored on the SIM as input and produces a 32-bit response (SRES) and a 64-bit long key K_c as output.

EAP SIM mechanisms specify enhancements to GSM authentication and key agreement whereby multiple authentication triplets can be combined to create authentication responses and encryption keys of greater strength than the individual GSM triplets. The mechanism also includes network authentication, user anonymity support and a fast re-authentication procedure.

Authentication Example

Figure 5 shows an example of EAP-SIM full authentication. Authentication is started with a request for client identification. The software process on the client platform that performs the EAP-SIM negotiation is called the *supplicant*. The supplicant's response includes either the user's International Mobile Subscriber Identity (IMSI) or a temporary identity (pseudonym). From this point on, the Authenticator only plays the role of a relay agent, shuttling messages back and forth between the supplicant and the AAA server.

Next, the supplicant receives an EAP Request of type SIM/Start from the Authenticator and replies with the corresponding EAP Response including a random number (NONCE) chosen by the supplicant.

After receiving the EAP Response/SIM/Start, the AAA server obtains *n* GSM triplets from the user's home operator's Authentication Centre (AuC) on the GSM network. From the triplets and other authentication parameters (Identity, EAP version, NONCE) the AAA server derives the keying material:

- The authentication key *K_aut* to be used with the MAC attributes;

ISSN 0085-7130©Telenor ASA 2007

- The encryption key *K_encr*, to be used with the ENCR_DATA attributes;

- Eventually, the master key and other application specific keys may also be derived.

The authentication key K_aut is used to compute the message authentication code (MAC) to be used in subsequent EAP messages. This MAC may contain message specific content (e.g. as shown in Figure 1, MAC (message | NONCE) will be the MAC of concatenation of the EAP message with the NONCE attribute).

The encryption key is used to encrypt the ENCR-DATA attributes. This encryption also uses an Initialization Vector (IV) that is a mandatory attribute in all EAP messages where any encrypted attribute is present. Finally, the master key can be used to protect the radio link depending on the different 802 security protocols used.

Once the key has been calculated, it is possible for the AAA server to send an EAP Request/SIM Challenge including the RAND of the GSM Triplets, an encrypted next client identity, and the MAC including the NONCE to be sent back to the supplicant.

Once the supplicant has received this challenge, it will run the GSM algorithm to obtain the GSM triplets, then derive the keys as done in the server, and compute the MAC to compare it with the server-calculated MAC. If the MACs match, the network is identified as one knowing GSM triplets and the client originated NONCE random number. If the network authentication is correct, the supplicant responds with the EAP Response SIM/Challenge, containing the MAC attribute that includes the client's SRES response values.

The AAA server verifies that the MAC is correct and sends an EAP-Success packet to the authenticator, indicating that the authentication was successful.

EAP-SIM Implementation for Authentication to Internet Services

Installation of EAP-SIM specific software on the client PC is not required. An ActiveX-compliant supplicant with [4] and [5] is automatically downloaded from the IDP. This ActiveX supplicant can run in the MS Internet Explorer. Its main functions are to

1 Receive the EAP request packets from the EAP Authenticator;

2 Send the contents of these packets to SIM as specified in [4] if the TS 102.310 EAP-SIM card appli-

Figure 8 Entering the session ID received from the mobile phone

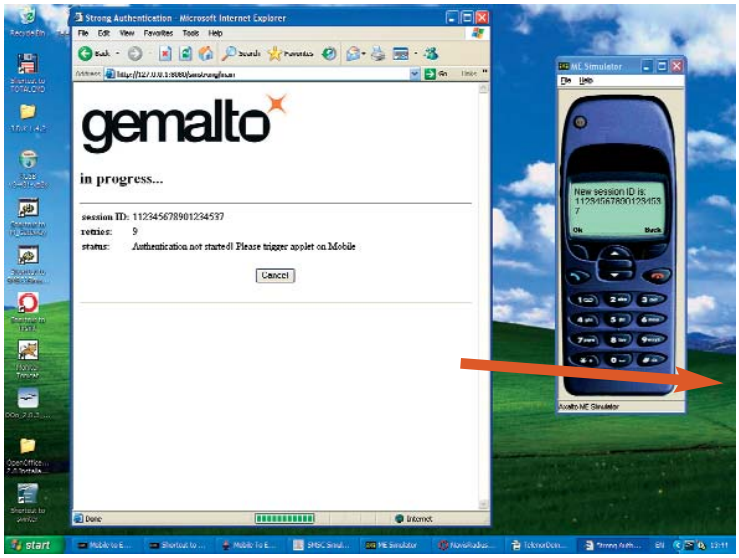


Figure 9 Initiation of the authentication of the mobile phone

Figure 10 Server authentication succeeded

ISSN 0085-7130

Figure 11 Authentication successful

cation is installed in SIM, or as specified in [5] if the WLAN-SIM card application is installed in SIM;

3 Build the EAP response packets from the SIM responses;

4 Send back the EAP-response packets to the EAP authenticator.

The EAP Authenticator is implemented as a Java servlet inside the Telenor IDP. This servlet communicates through the ActiveX module to the Gemalto SIM card. The EAP Authenticator first requests the EAP identity from the SIM (via the supplicant) and sends the identity received to the Lucent Technologies Vital*AAA* server. Note the first permanent EAP identity contains the IMSI.

The Telenor IDP communicates with the ActiveX supplicant running in the Microsoft Internet Explorer browser on the user's laptop, which in turn communicates with the SIM to request the IMSI (International Mobile Subscriber Identity) or a temporary identity.

After receiving the EAP Identity from the supplicant (relayed via the IDP Authenticator), the Vital*AAA* server sends a request for triplets corresponding to the SIM IMSI to the Telenor HLR. Because the current generation of HLRs are only accessible via the SS7 protocol, an IP to SS7 gateway is required for the AAA server to access data stored in an HLR. The Vital*AAA* server sends requests for HLR data to the Ulticom Signalware MAP Authentication Gateway via a special interface application. The Signalware gateway then formats MAP messages which are sent to the HLR and the resulting response is returned to the Vital*AAA* server.

Authentication triplets are not stored in the HLR, but are generated as needed by the Authentication Centre (AuC) in the HLR. After receiving the MAP request from the Signalware gateway, the Telenor HLR

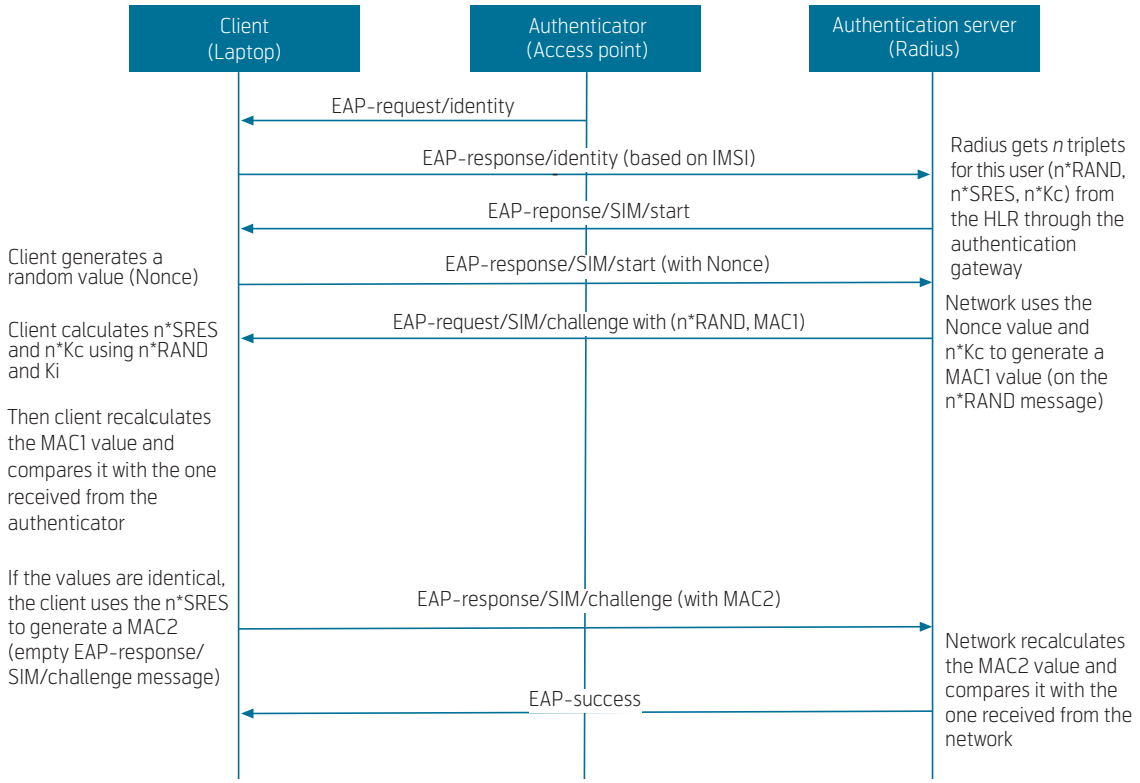


Figure 12 EAP SIM authentication

ISSN 0085-7130©Telenor ASA 2007

requests the triplets from the AuC and returns them to the AAA server through the MAP gateway.

The Vital*AAA* server then sends the challenges contained in the triplets to the Telenor IDP authenticator servlet that then forwards them to the ActiveX supplicant in the Microsoft Internet Explorer browser.

The browser ActiveX supplicant communicates with SIM that it will

1 Verify the MAC1 received from Radius in order to authenticate the server, and

2 Calculate a MAC2 to be sent to the AAA server (through the IDP servlet).

If the AAA server verifies that MAC2 is correct, it informs the IDP about the mutual authentication success and the IDP then redirects the browser back to the application provider, in this case myBank.no for Kari.

Upon successful authentication, the Telenor IDP will return a Single-Sign-On token to the browser and redirect it back to the service provider. The service provider will verify the Single-Sign-On token before granting access to its services.

Additional Protections

In addition to the inherent authentication capabilities provided by the SIM based identity, the Vital*AAA* server can also provide other means of controlling system access via authorization policy. For example:

- Users may be limited to a list of applications;
- Access may be limited to specific geographic locations;
- Time-of-Day and Day-of-Week controls can be applied;
- Accounts may be temporarily blocked for business purposes;
- Security measures may be applied to reject stolen or compromised SIMs;
- Each user access can easily be logged and logs are shared in real time with the appropriate application provider.

EAP-SIM Over SMS Implementation

As illustrated previously in the example of Kari, another solution available for StrongSim authorization is to use SMS. As with the other SIM Strong Authentication methods, SMS makes use of the SIM card for unique, secure identification, just as do the other StrongSim methods. The benefit of SMS is the elimination of the need for Bluetooth or dongle connectivity. SMS connectivity is enabled by a SIM application provided by Gemalto.

The architecture of the EAP-SIM over SMS solution is shown in Figure 13. The EAP-SIM protocol is now transported by SMS from the mobile phone to the Telenor IDP that forwards them to the Vital*AAA* server. The Vital*AAA* server will then communicate with the Telenor HLR/AuC to get the authentication triplets necessary for the authentication.

In order to ensure that the legitimate user is attempting to log in, the user is asked to start the authentication application on the mobile phone and to enter the Session ID received on the mobile phone on the web page. A confirmation on the mobile phone is then also required to initiate the authentication.

In this implementation, the number of exchanged short messages was optimized: only two mobile-originated short messages and one server-originated short message are required for an EAP-SIM full authentication.

Value Proposition

To End Users

The SIM Strong Authentication Service will deliver value to end users in the following ways:

- *Simple and better control and management of their identities:* The user does not have to manage a multitude of passwords. All the end user needs is a SIM card and a mobile phone, a USB dongle, or a PC GPRS/3G Data card with a SIM card.

- *Better protection and higher level of security:* The SIM Strong Authentication Service provides much better protection than passwords.

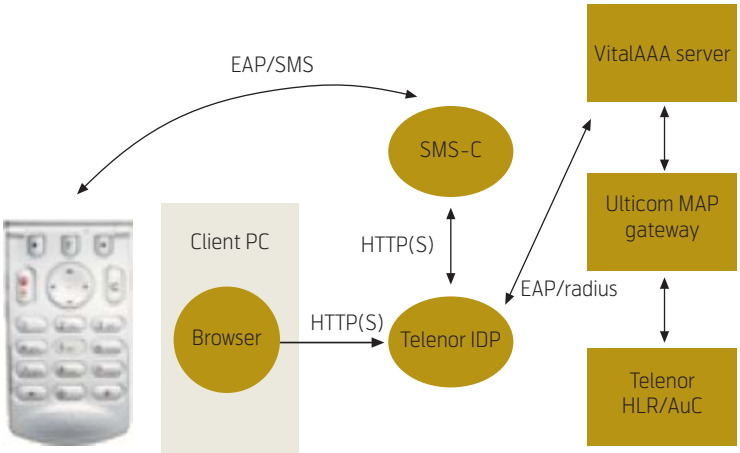


Figure 13 EAP-SIM over SMS

ISSN 0085-7130

- *Ease of use:* The SIM Strong Authentication Service is very simple to use and does not require any particular technical skill. The log-in is easy and quite intuitive.

- *Single-Sign-On:* After a successful authentication, the user does not have to log in again when visiting other service providers using the SIM Strong Authentication Service. The availability of Single-Sign-On access is time limited for security purposes.

- *Universal applicability:* The SIM Strong Authentication Service can be used for any service or application and the user need not use several different authentication solutions.

- *Global availability:* The SIM Strong Authentication Service can be used anywhere and even when there is no GSM coverage. Indeed, even with a non-operational phone due to lack of coverage, the SIM-based authentication can still be performed via Bluetooth.

To Service Providers

The SIM Strong Authentication Service will bring the following benefits to service providers:

- *Better protection and higher level of security:* The SIM strong and mutual authentication service provides higher protection of valuable assets and contributes to extending the availability of their services.

- *Cost savings:* By replacing their current password-based authentication schemes, service providers can save money on operation and maintenance costs due to the simplicity of the application

- *Lower threshold for deployment:* Service providers do not have to invest large amounts of money to deploy the SIM Strong Authentication Service because the mobile operator manages most of the infrastructure. No great technical expertise is required and the SIM Strong Authentication Service fits very well for larger enterprises and SMEs.

- *Simpler customer management:* Service providers do not have to take care of the password management since the mobile operators will assume this responsibility.

- *Reach more customers:* The service providers may also reach new customers that are subscribers at the mobile operators.

To Mobile Operators

For mobile operators, the SIM Strong Authentication Service will bring the following benefits:

- *New source of revenue:* The SIM Strong Authentication Service constitutes an additional source of revenue for mobile operators which is not based on the sale of air traffic. This source of revenue has large potential since it brings value to end users and service providers.

- *Reuse of existing infrastructure:* Because the SIM authentication solution uses the same SIM and HLR infrastructure used for normal GSM and GPRS services, it allows the reuse of the GSM expertise of the mobile operator.

- *Improved customer loyalty:* The SIM Strong Authentication Service will be a valuable service to end users and will hence contribute to improving customer loyalty and reducing churn.

- *New business customers:* As a compelling service, the SIM Strong Authentication Service will attract new customers for the mobile operator.

- *Strengthened position:* By extending the role and the value of the mobile phone and SIM to the computing world, the SIM Strong Authentication Service will contribute to considerably strengthening the mobile operator's position in the new converged ICT world.

- *Easy adaptability for the future:* Because the SIM strong authentication is based on easily changeable software elements (ActiveX supplicant, IDP Java Authenticator, Vital*AAA* server and Signalware gateway) it can easily be modified and upgraded to support emerging and future technologies. For example: UMTS USIMs, Smart Card based Certificates, Smart Card-based One-Time-Password (OTP) schemes, etc. Because of the flexibility of the platform described in this paper, it is quite possible to support multiple authentication schemes over a single authentication infrastructure.

Conclusion

In this paper, a SIM Strong Authentication Service is presented. By its usage simplicity, its high level of security, its universal applicability and its cost efficiency, the SIM Strong Authentication Service will most likely be a successful service in the near future. A proof-of-concept implementation has been completed by Telenor, Gemalto, Linus and Oslo University College in collaboration with Sun, Lucent Technologies and Ulticom. A demonstration of the service

ISSN 0085-7130©Telenor ASA 2007

was shown at the 3GSM World Congress in Barcelona, Spain in February 2006. The demo over SMS that enables any end-user, even with a low-end GSM handset, to perform a SIM strong authentication was presented at Cartes 2006, Paris, November 2006 and it also won a SESAMES award at the same event.

References

1 *The Liberty Alliance.* August 30, 2007 [online] – URL: http://www.projectliberty.org/

2 Haverinen, H, Salowey, J (eds). *Extensible Authentication Protocol Method for Global System for Mobile Communications (GSM) Subscriber Identity Modules (EAP-SIM)*. IETF, January 2006. (RFC 4186)

3 Arkko, J, Haverinen H. *Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement (EAP-AKA)*. IETF, January 2006. (RFC 4187)

4 ETSI 3GPP. *EAP-support on UICC – TS 102310 v060100p*. Sophia Antipolis, ETSI, August 2–5, 2004.

5 *WLAN-SIM Specification version 1.0*. WLAN Smart Card Consortium, October 20, 2003.

6 Aboba, B et al (eds). *Extensible Authentication Protocol (EAP)*. IETF, June 2004. (RFC 3748)

7 Rigney, C et al. *Remote Authentication Dial In User Service (RADIUS)*. IETF, June 2000. (RFC 2865)

8 Rigney, C, Willats, W, Calhoun, P. *RADIUS Extensions*. IETF, June 2000. (RFC 2869)

For a presentation of Ivar Jørstad, Do Van Thuan, and Do Van Thanh, please turn to page 10, 18, and 2, respectively.

Tore Jønvik is Associate Professor at Oslo University College, Faculty of Engineering. Jønvik holds an MSc in Physics and a PhD in Informatics from the University of Oslo. He has been guest researcher in Telenor R&I and has been working in the Eurescom project P1101: Device Unifying Service, and two Eureka-Celtic projects: ADPO and Fidelity. He is now engaged in the Eureka project Mobicome.

email: tore.jonvik@iu.hio.no

ISSN 0085-7130

Unifying CardSpace and Liberty Alliance with SIM Authentication

IVAR JØRSTAD, DO VAN THUAN, TORE JØNVIK, DO VAN THANH

Ivar Jørstad is CEO of Ubisafe AS

Do Van Thuan is Lead Scientist in Linus AS

Tore Jønvik is Associate Professor at Oslo University College

Do Van Thanh is Senior Research Scientist in Telenor R&I

This paper presents an innovative service that integrates both the Microsoft CardSpace and the Liberty Alliance Identity Management with the GSM SIM authentication. Telenor, Gemalto, Linus, Ubisafe and Oslo University College with the collaboration of SUN, Ulticom and Lucent have implemented a proof-of-concept of the service in Oslo. The goal is to demonstrate the feasibility of a convergent authentication service. The architecture is based on a multi-vendor environment where Sun provides the Identity Provider, Microsoft the Security Token Service (STS), Lucent Technologies the Radius server and Ulticom the SS7 MAP Authentication Gateway connecting the prototype to the Telenor mobile network.

A typical user flow for such a service would be the case of a user browsing on the World Wide Web from home, a customer premise, an Internet café, etc. When trying to access a protected resource such as Web mail, company portal, or bank account, he/she logs on to the requested secured site simply by approving the authentication on his/her mobile phone.

This service is available anywhere and can support any Internet services no matter whether they are offered by a Liberty Alliance Service Provider or by a CardSpace Relying Party. The Unified SIM strong authentication is both user-friendly and cost efficient, with a low deployment threshold. The technology is also capable of supporting other Smart-Card based identity services such as USIM (UMTS), certificate based schemes (e.g. TLS) and One Time Password schemes (OTP).

Introduction

Several frameworks for identity management have been proposed by different organisations and companies the last years. Microsoft has embedded their CardSpace into recent versions of their operating system, and Liberty Alliance [1] has developed open specifications for managing user identities in Internet-based services. However, the existence of several quite different identity mangement frameworks does not necessarily make life easier for the users. Currently, users must remember more and more usernames and passwords, and one of the goals of these frameworks is to remedy this situation by enabling single-sign-on and simpler management of user credentials. Liberty Alliance came up with a federated network identity solution that offers single sign-on enabling the user to visit several web sites without having to log in again. CardSpace from Microsoft provides a user-friendly solution to manage multiple identities. Unfortunately, these solutions are not interoperable. Telenor, Linus, Oslo University College, Ubisafe and Gemalto, in collaboration with Sun, Lucent Technologies and Ulticom, have designed and implemented a strong authentication service that integrates both the Microsoft CardSpace and the Liberty Alliance Identity Management. The idea is to integrate the current SIM authentication used in GSM with Liberty Alliance and CardSpace such that it can be used for Internet services. Indeed, this is a step further than earlier work that uses SIM authentication for WLAN (Wi-Fi – EAP-SIM). The idea of making the mobile phone and its SIM a universal authentication token is compelling, since the mobile phone is so common nowadays, and the GSM network is currently the largest mobile network and ubiquitous in much of the world.

This paper presents the Unified SIM Strong Authentication Service for Liberty Alliance and CardSpace. It starts with a short introduction of Liberty Alliance and CardSpace. Then follows an overview of our SIM Strong Authentication Service followed by a scenario showing how our SIM Strong Authentication Service works will be depicted. The value brought to users and service providers will be identified. The business opportunities for the mobile operators are also analysed.

Finally, the paper will explain how the Liberty Alliance Framework can be used to leverage this SIM-based strong authentication solution in a heterogeneous, multi-vendor environment that bridges Internet-based services and the GSM network.

Introducing Liberty Alliance

The Liberty Alliance [1] uses the concept of network identity which refers to the global set of attributes that are contained in an individual's various accounts with different service providers. Currently the user's network identities are like isolated islands and the user is responsible for remembering numerous user-

ISSN 0085-7130

names and passwords for each of these identity islands. The user will typically either try to always use the same password or to record the password somewhere. Either way, the result is a drop in the level of security.

The most logical solution to the problem caused by the isolated network identity is to build bridges that interconnect them and allow information flows between them. This is precisely what "Federation" is doing. *Federation* refers to the technologies that make identity and entitlements portable across autonomous policy domains. Consequently, the Federated Network Identity is a portable identity.

The establishment of federated relationships between service providers will hence allow the users to move more seamlessly from one service provider to another. However, if every service provider has to make alliance to each of the other service providers it will be time consuming and require tremendous efforts. For n service providers, it requires $n(n-1)/2$ established relationships.

To circumvent this problem, the Liberty Alliance proposed a new role called Identity Provider. The Identity Provider assumes the management of the user's Federated Network Identity and the user authentication.

A Circle of Trust is a group of service providers and identity providers that have business relationships based on Liberty architecture and operational agreements and with whom users can transact business in a secure and apparently seamless environment.

Figure 1 shows a Circle of Trust. The Principal is the user, employer, customer, game user, etc. whose Federated Network Identity is managed by the Identity Provider. Once federation is done, the user can enjoy Single Sign-On. As shown in Figure 3, Joe has logged in at his IDP as JoeSelf. When visiting Yahoo.com, he is "automatically" logged in as BostonDude.

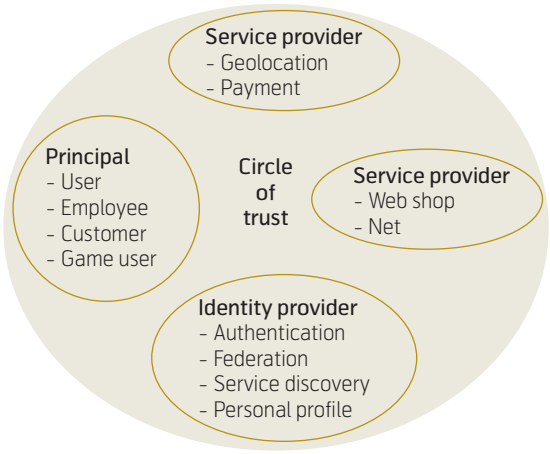


Figure 1 A Liberty Alliance Circle of Trust

Introducing CardSpace

CardSpace [2] is Microsoft's latest proposal for secure digital identities. CardSpace, originally code-named "InfoCard", lets any Windows application, including Microsoft's own applications such as the next release of Internet Explorer and those created by others, and its users to work with digital identities in a common way. As part of the .NET Framework 3.0, CardSpace will be available for Windows Vista, Windows XP, and Windows Server 2003.

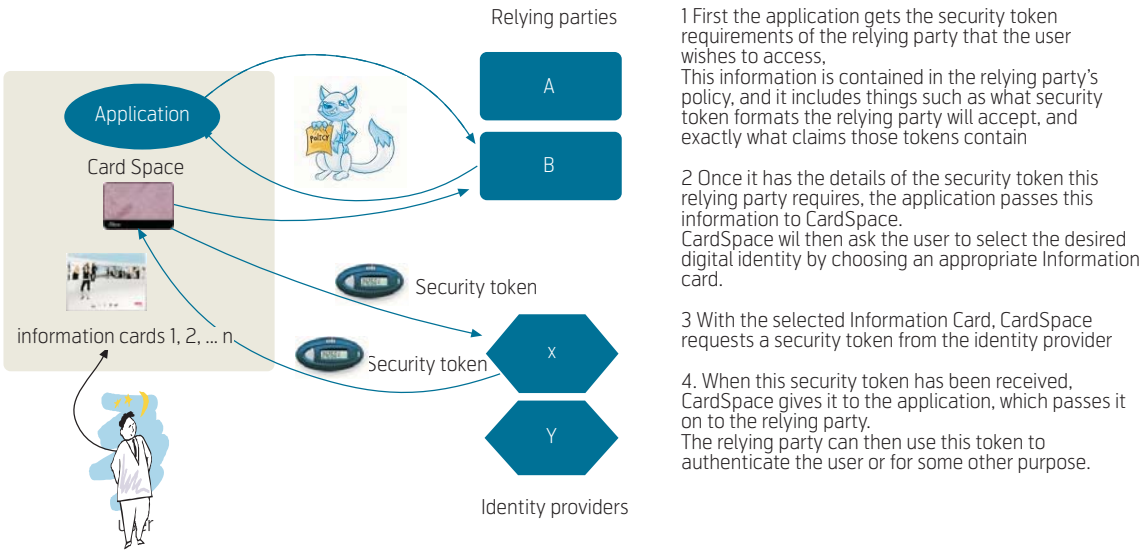


Figure 2 CardSpace and interaction among user, relying party and identity provider

ISSN 0085-7130

CardSpace provides the user with a consistent way to work with multiple digital identities, regardless of the kinds of security tokens they use. The user can create, use, and manage these diverse digital identities in an understandable and effective way. She might also be able to choose from a group of identity providers as the source of the digital identity she presents to the relying parties.

Our SIM Strong Authentication Service

A SIM-based strong authentication service that extends SIM card GSM authentication to Web services is proposed. It can briefly be described as follows:

- A user with a valid Telenor mobile subscription having a mobile phone with a SIM and SMS service may quite easily and securely log on to
  - An Internet bank
  - A corporate intranet
  - A commerce web shop
  - An Enterprise website
  - An *e*Government application
- At any time and anywhere in the world.

The advantages of the SIM Strong Authentication Service can be summarised as follows:

- Removes the user's burden of remembering passwords by using the SIM card;
- Provides an authentication service that is both strong and easy to use;
- Allows rapid deployment due to the high penetration of mobile phones;
- Reuses existing GSM authentication structures (SIM card and HLR);
- Allows the integration of all services and applications;
- Uses open standards and supports interoperability with other systems;
- Provides scalability and supports a large number of users and service providers.

The SIM strong authentication extends the usage of the EAP-SIM protocol [3] [4] [5] [6]in WLAN authentication to the Internet services.

Integration of Liberty Alliance and CardSpace

The Unified Strong SIM authentication, as indicated by its name, unifies the Liberty Alliance solution and Microsoft's CardSpace and provides the user with the possibility to log in using both schemes.

When visiting a Service Provider belonging to Telenor's Circle-of-Trust the user will be redirected to Telenor's Identity Provider for sign in. The user can use his mobile phone to authenticate himself. After successful authentication, the user is logged onto the Service Provider. After a while, if the user visits another Service Provider belonging to Telenor's Circle of Trust, he does not have to sign in again. Single Sign-on is provided.

Now, if the user visits a web site which does not belong to the Telenor Circle-of-trust but is a Relying Party, i.e. uses the Telenor's authentication service, he can use the Telenor ID card in CardSpace to do the authentication. Again, the authentication is carried out via his mobile phone.

To elucidate the Unified Strong SIM authentication service let us consider two cases:

Sign in to the Liberty Alliance Circle-of-Trust

As shown in Figure 3, the following actions are performed:

1 Kari connects her laptop on the Internet and visits a website, e.g. myBank.com.

2 When she attempts to log in, she is redirected to the Telenor Identity Provider website for authentication.

3 The Telenor IDP performs authentication via SMS, and Kari receives a message on her mobile phone. She approves the authentication.

4 Kari is now notified that her authentication is successful. The Telenor IDP redirects the browser back to myBank.com where Kari is now logged in, and a welcome page is displayed. Kari can carry out all her transactions.

5 After a while Kari decides to go to her enterprise, e.g. myEnterprise.com. There she is immediately recognised and receives a welcome page. She enjoys the convenience of single sign-on.

ISSN 0085-7130©Telenor ASA 2007

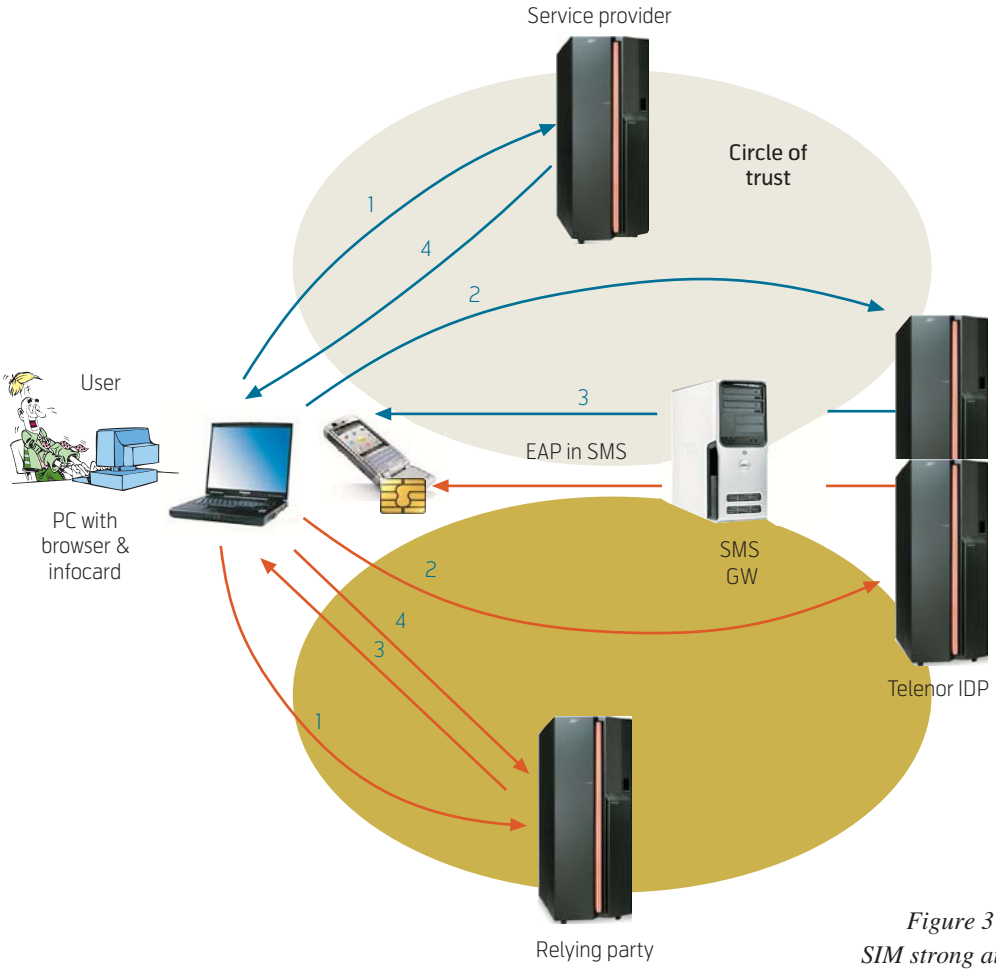


Figure 3 The Unified SIM strong authentication

Sign in with Card Space

1 Later, Kari goes and visits a web site, e.g. https://sim10.nta.no/zivasrv, which is a relying party of the Telenor's Identity Provider.

2 When she clicks on the log-in button she is redirected back to CardSpace in her PC.

3 She selects the Telenor ID card. CardSpace requests the Telenor IDP to initiate authentication.

4 The Telenor IDP carries out the authentication via SMS and Kari receives a message on her mobile phone. She approves the authentication.

5 The authentication is successful. Kari is re-directed back to the Relying Party where a welcome page is displayed to her.

Unified SIM Strong Authentication Implementation for Liberty Alliance and CardSpace

The architecture of the Unified SIM strong authentication is depicted in Figure 4. The heart of the system is Telenor's Identity Provider (IDP). It communicates with all the entities and supervises all the interactions:

- On the *Internet side*, it is able to communicate with
  - All the *Liberty Alliance Service Providers* that have joined Telenor's Circle-of-Trust and provide the SIM strong authentication service to them;
  - All the *CardSpace Relying Parties* that use Telenor's Identity Card and offers the SIM strong authentication service to them.
- On the *mobile network side*, it is able to communicate with
  - *The SMS (Short Message Service) gateway* to perform authentication using EAP-SIM protocol toward the users' mobile phones. More details about the EAP-SIM protocol is given in Annex A;
  - *The AAA (Radius) server* [7] [8] that again communicates with Telenor's HLR (Home Location Register) via the MAP gateway to carry out the users' authentication.

ISSN 0085-7130

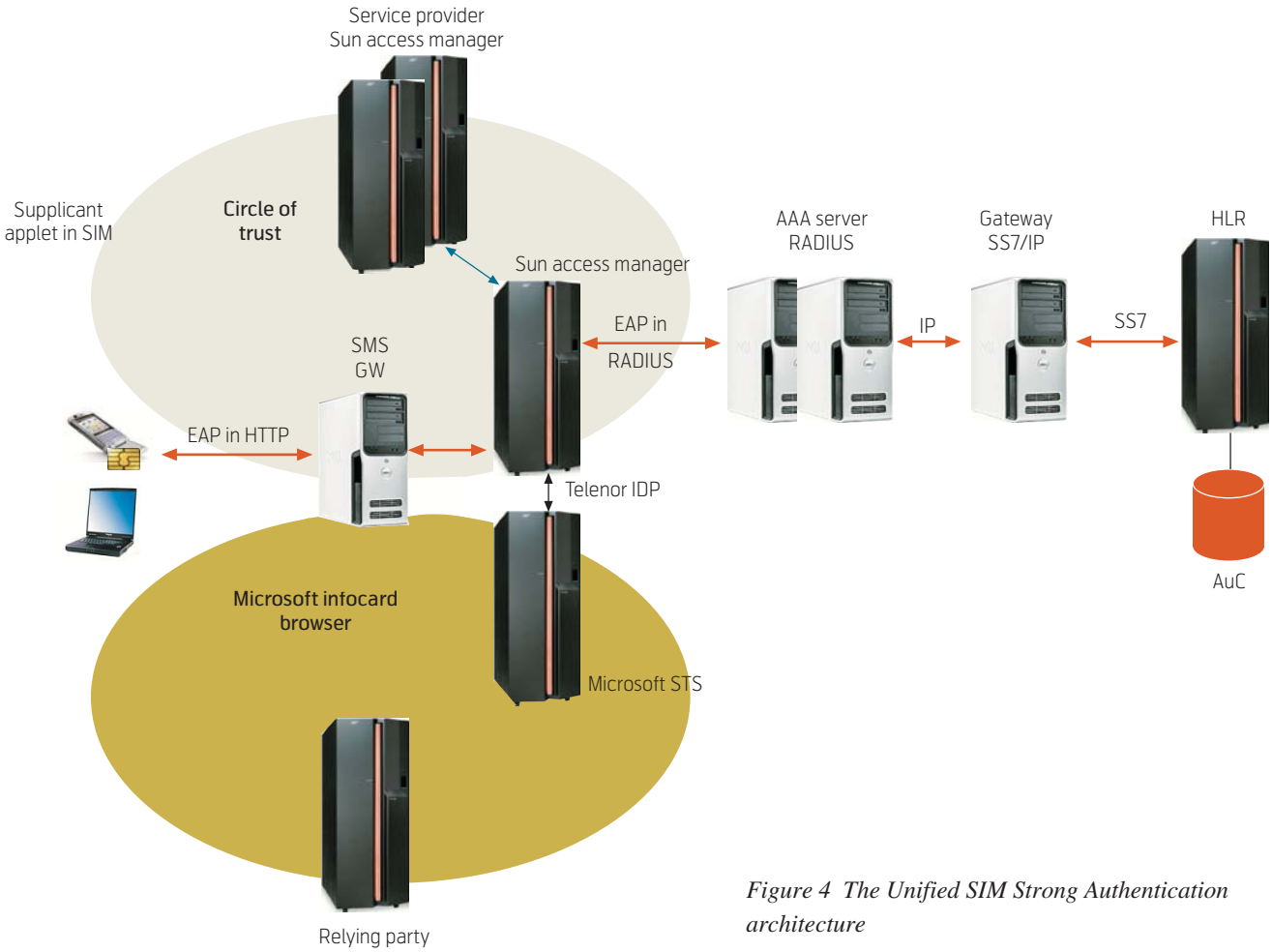


Figure 4 The Unified SIM Strong Authentication architecture

Telenor's IDP consists of two main elements:

- A *SUN Access Manager,* which is a Liberty Alliance compliant Identity Provider;
- A *Microsoft STS (Security Token Service).*

Since the Unified SIM Strong Authentication Service is an extension of the SIM Strong Authentication [9], which is offered in a Liberty Alliance Circle-of-Trust with the SUN Access Manager as the main element, an interface has been introduced to bridge with Microsoft's STS. In addition to management and information exchange methods, this interface offers an Authentication request method that allows the STS to initiate the entire authentication based on the SIM card.

Value Proposition

To End Users

The Unified SIM Strong Authentication Service will deliver value to end users in the following ways:

- *Simple and better control and management of their identities:* The user does not have to manage a multitude of passwords. All the end user needs is an operating mobile phone with a SIM card.
- *Better protection and higher level of security:* The Unified SIM Strong Authentication Service provides much better protection than passwords.
- *Ease of use:* The Unified SIM Strong Authentication Service is very simple to use and does not require any particular technical skill. The log-in is easy and quite intuitive.
- *Single Sign-On:* After a successful authentication, the user does not have to log in again when visiting other service providers using the Unified SIM Strong Authentication Service. The availability of Single Sign-On access is time limited for security purposes.
- *Universal applicability:* The Unified SIM Strong Authentication Service can be used for any service or application.
- *Global availability:* The Unified SIM Strong Authentication Service can be used anywhere, even where there is no GSM coverage. Indeed, even with a non-operational phone due to lack of cover-

ISSN 0085-7130©Telenor ASA 2007

age, the Unified SIM-based authentication can still be performed via Bluetooth.

To Service Providers and Relying Party

The Unified SIM Strong Authentication Service will bring the following benefits to service providers:

- *Better protection and higher level of security:* The Unified SIM strong and mutual authentication service provides higher protection of valuable assets and contributes to extending the availability of their services.

- *Cost savings:* By replacing their current password-based authentication schemes, service providers can save money on operation and maintenance costs due to the simplicity of the application

- *Lower threshold for deployment:* Service providers and Relying Partners do not have to invest large amounts of money to deploy the Unified SIM Strong Authentication Service because the mobile operator manages most of the infrastructure. No great technical expertise is required and the Unified SIM Strong Authentication Service fits very well for larger enterprises and SMEs.

- *Simpler customer management:* Service providers and Relying Parties do not have to take care of the password management since the mobile operators will assume this responsibility.

- *Reach more customers:* The Service Providers and Relying Parties may also reach new customers that are subscribers at the mobile operators.

To Mobile Operators

For mobile operators, the Unified SIM Strong Authentication Service will bring the following benefits:

- *New source of revenue:* The Unified SIM Strong Authentication Service constitutes an additional source of revenue for mobile operators which are not based on the sale of air traffic. This source of revenue has large potential since it brings value to end users and service providers.

- *Reuse of existing infrastructure:* Because the Unified SIM authentication solution uses the same SIM and HLR infrastructure used for normal GSM and GPRS services, it allows the reuse of the GSM expertise of the mobile operator.

- *Improved customer loyalty:* The Unified SIM Strong Authentication Service will be a valuable service to end users and will hence contribute to improving customer loyalty and reducing churn.

- *New business customers:* As a compelling service, the Unified SIM Strong Authentication Service will attract new customers for the mobile operator.

- *Strengthened position:* By extending the role and the value of the mobile phone and SIM to the computing world, the Unified SIM Strong Authentication Service will contribute to a considerable strengthening of the mobile operator's position in the new converged ICT world.

- *Easy adaptability for the future:* Because the Unified SIM strong authentication is based on easily changeable software elements (Active-X supplicant, IDP Java Authenticator, Vital*AAA* server and Signalware gateway) it can be easily modified and upgraded to support emerging and future technologies, for example UMTS USIMs, Smart Card based Certificates, Smart Card-based One-Time-Password (OTP) schemes, etc. Because of the flexibility of the platform described in this paper, it is quite possible to support multiple authentication schemes over a single authentication infrastructure.

Conclusion

Today, service providers have to choose between so many authentication and identity management schemes, and users are left struggling with a variety of digital identities. There are too many duplications and divergences in the digital identity world, and it must end. With the Unified SIM Strong Authentication Service, the mobile phone is indeed the point of convergence of CardSpace and Liberty Alliance identity frameworks. The user is offered the freedom and simplicity of participating and visiting all the web sites no matter whether they are a Liberty Alliance Service Provider or a Microsoft's Relying Party. In addition, a high level of security and convenience is ensured via the usage of the mobile phone as a security token.

A proof-of-concept implementation of the Unified Strong Authentication has been completed by Telenor, Gemalto, Linus, Ubisafe and Oslo University College in collaboration with Sun, Lucent Technologies and Ulticom. A demonstration of the service was shown at the 3GSM World Congress in Barcelona, Spain, February 2007.

ISSN 0085-7130

References

1 *The Liberty Alliance*. August 31, 2007 [online] – URL: http://www.projectliberty.org/

2 Microsoft. *CardSpace*. August 31, 2007 [online] – URL: http://cardspace.netfx3.com/

3 Haverinen, H, Salowey, J (eds). *Extensible Authentication Protocol Method for Global System for Mobile Communications (GSM) Subscriber Identity Modules (EAP SIM)*. IETF, January 2006. (RFC 4186)

4 Arkko, J, Haverinen, H. *Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement (EAP-AKA)*. IETF, January 2006. (RFC 4187)

5 *WLAN-SIM Specification version 1.0*. WLAN Smart Card Consortium, October 20, 2003.

6 Aboba, B et al (eds). *Extensible Authentication Protocol (EAP)*. IETF, June 2004. (RFC 3748)

7 Rigney, C et al. *Remote Authentication Dial In User Service (RADIUS)*. IETF, June 2000. (RFC 2865)

8 Rigney, C, Willats, W, Calhoun P. *RADIUS Extensions*. IETF, June 2000. (RFC 2869)

9 *SIM strong – Offering SIM strong authentication to Internet Services*. August 31, 2007 [online] – URL: http://www.simstrong.org

For a presentation of Ivar Jørstad, Do Van Thuan, Tore Jønvik, and Do Van Thanh, please turn to page 10, 18, 135, and 2, respectively.

ISSN 0085-7130©Telenor ASA 2007

The Mobile Phone as Authentication Token

IVAR JØRSTAD, DO VAN THANH

Ivar Jørstad is CEO of Ubisafe AS

This paper provides a study of the various ways the mobile phone can be used as an authentication token towards service providers on the Internet. It starts by discussing the need for a strong authentication scheme and the motivation for using the mobile phone to improve on several aspects of the current authentication processes. Then, the general architecture for authentication with mobile phones is presented. Several different authentication solutions using the mobile phone as authentication token are then described, where the solutions vary in complexity, strength and user-friendliness. The paper ends with an evaluation of the different solutions, and a discussion of the most probable attacks. A classification of the solutions is also provided, according to defined criteria.

Do Van Thanh is Senior Research Scientist in Telenor R&I

1 Introduction

A lot of value-added services on the Internet today require authentication, i.e. the proper verification of the user's identity, before the user gets access to the service itself. The most common authentication scheme which is employed is the use of username and password. However, the username/password scheme has several weaknesses which make it unsuitable for very many (most) services. Most importantly they reduce security and make service access inconvenient for the users due to:

- Username/password can be subject to phishing;
- Too many username/password pairs make it impossible to remember all;
- It is common to reuse the same username/password for different services;
- It is common to write down username/password where others can easily find them.

Due to these weaknesses of the username/password scheme, other solutions must be used. For services with particularly high security requirements, One-Time-Passwords (OTP) [1] or Smart Card solutions have been taken into use, e.g. for Internet-banking services or corporate service access. These more secure solutions usually increase the cost of the security solutions both in the deployment phase (provide all users with OTP-calculators or Smart Cards w/readers) but also in the maintenance phase (defective OTP-calculators, lost Smart Cards, invalidation of users etc.).

As a consequence of the weaknesses of existing authentication solutions, new solutions should be investigated. Using the mobile phone as an authentication solution towards Internet-based services is particularly compelling due to:

- Its widespread deployment (over 100 % market penetration in Norway);
- It usually follows the user at all times (you seldom leave home without).

Thus, by using the mobile phone as an authentication solution, the major weaknesses of existing solutions are reduced:

- More convenient for the user who now only requires one device;
- Usually higher level of security (depends on the specific authentication scheme implemented);
- Reduction in deployment and maintenance costs for service providers (reuses existing infrastructure, i.e. the mobile phone and the mobile network).

In addition, some such solutions can provide mobile network operators with new sources of revenue since they control much of the required infrastructure.

This paper starts with a discussion of authentication strength and discusses the concept of multi-channel authentication as a mechanism for stronger authentication. Then some different authentication schemes using the mobile phone as an authentication token will be described. An evaluation of the different schemes will then be performed according to specific criteria.

2 Strong Authentication

2.1 Two Factor Authentication

Two factor authentication means that the authentication process consists of two stages, where each stage uses a different credential. For example, the mobile phone authentication process is (usually) two factor

ISSN 0085-7130

because it relies on 1) the PIN-code to activate the SIM-card, and 2) the possession of the SIM-card itself with the appropriate keys and algorithms. Lacking one of these will usually mean that it is impossible to authenticate. However, it is possible to reduce the SIM authentication to one factor by disabling the PIN-code verification.

2.2 Multi-channel Authentication

Multi-channel authentication is the process of utilising more than one communication channel for the secure establishment of the user's identity. It is crucial to understand the properties of the different channels in order to implement authentication solutions with adequate properties. When using the mobile phone for authentication, it is today possible to use the connection between the mobile phone and a computer, which again communicates with an authentication server on the Internet, as one channel, e.g. to initiate the authentication process. The response to the authentication request can be sent on another channel back to the user, e.g. using an SMS message. The user can then again either complete the authentication process by responding with an SMS or by sending a message through the initial channel.

3 The Different Mobile Phone Authentication Schemes

To be able to authenticate users through a mobile phone, certain main components must be in place. Figure 1 shows the main components and the basic architecture needed for the solutions described later to work.

The user must have access to a computer connected to the Internet and be in possession of a mobile phone with a working SIM card. If the computer and mobile phone are equipped with Bluetooth, higher usability can be obtained. Through the Internet browser on the computer the user can access web services provided by service providers. The service provider (SP) is connected to an Authentication Server (AS) that will handle the authentication on behalf of the SP. The AS is connected to the GSM network which enables it to communicate with the user's mobile phone and the operator's Authentication Center (AuC). The AS is composed of two parts, an authenticator and an AAA server. The authenticator communicates with the client and relays messages to the AAA server which handles the authentication.

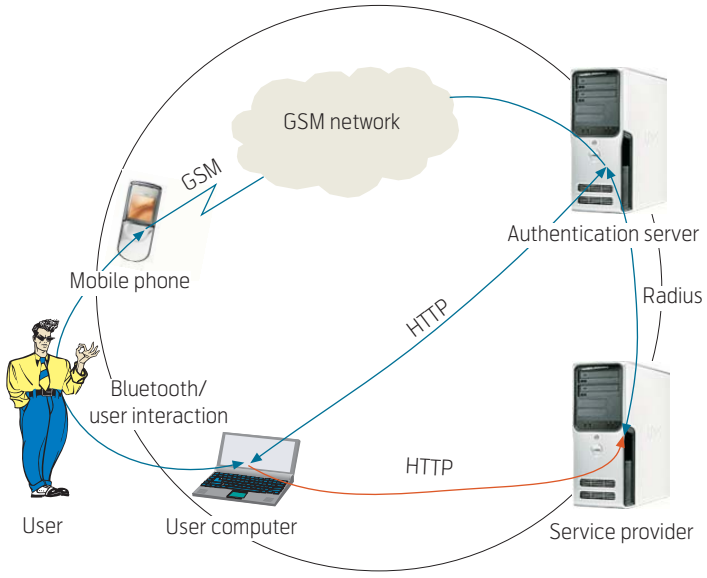


Figure 1 A general architecture of mobile authentication

When designing an authentication scheme which uses two separate devices that communicate over two different networks it is very important to ensure that it is the same user that controls both devices. This is done by ensuring that there is a "closed loop" going through all the components involved in the authentication as illustrated in Figure 1. The loop starts in the device requesting the service, the user's computer, goes through the network with SP and AS and then via the mobile phone and back to the initial device, either by user interaction or Bluetooth.

This closed loop can be realized in several ways as described in the solutions presented below.

3.2 SMS Authentication with Session-ID Check

This solution exploits that a user with a valid mobile subscription is already authenticated through the GSM system. Session IDs are used to ensure that it is the same user that controls both the computer and the mobile phone.

When the user accesses a service provider a unique session ID is created and sent both to the user's computer and to the mobile phone. The session ID is sent over the Internet to the computer and shown in the web browser and sent to the phone by SMS. Then the user can confirm that the session IDs match and send a confirmation by SMS to the service provider. When receiving the confirmation from the user the AS knows that the user is in possession of the phone and the authentication is successful.

The comparison of the session ID can be made by the user or automatically by software if the phone is connected to the computer by Bluetooth.

3.2.1 User Verification of Session ID

The user accesses the service provider through the Internet browser and chooses to be authenticated by his mobile phone. The user identifies himself with his

ISSN 0085-7130©Telenor ASA 2007

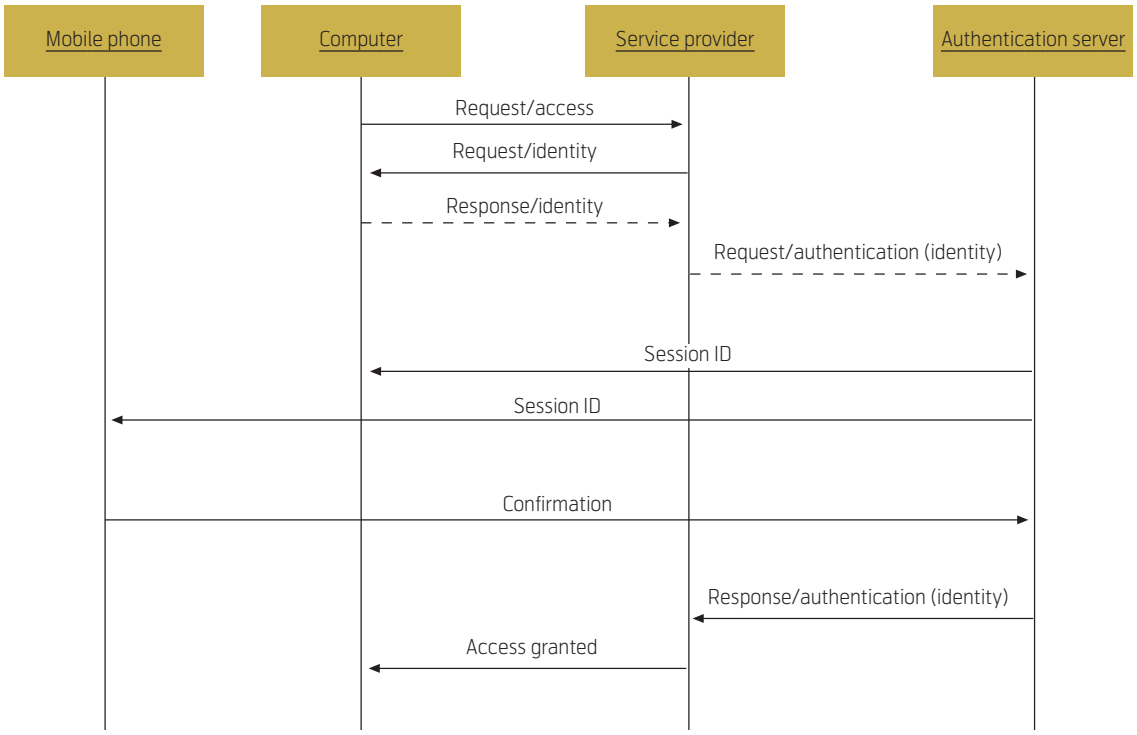


Figure 2 Session ID check

mobile phone number and the request is redirected to the authentication server. The authentication server then creates a unique session ID and sends this to the user by SMS and over the Internet to the computer. The user checks that the session ID in the SMS is the same as the one shown in the browser. If this is the case the user replies by sending an SMS back to the AS confirming that the session IDs match. When receiving the confirmation the authentication server redirects the browser back to the service provider and the user is given access to the service. The message exchange is shown in Figure 2.

3.2.2 Automatic Check of Session ID

To make it even easier for the user the session ID can be checked automatically. This can be done by pairing the mobile phone and the computer by Bluetooth.

When the computer receives the session ID from the AS a Java applet running in the browser contacts the SIM card using the Bluetooth SAP. SAP requires that a 16 digit pass-phrase is used during the pairing process. The applet checks the SIM for an SMS with an appropriate session ID. For this to function it must be ensured that the SMS received from the AS is stored on the SIM. This can be done by setting the TP-PID in the SIM header to require SIM data download [2]. This forces the mobile phone to store the SMS on the SIM. If a matching session ID is found the applet creates a confirmation message that is sent to the AS by SMS. This is done with a proactive SIM which can issue commands to the mobile phone as described in [3]. When the AS receives the confirmation it notifies the authenticator that the user is authenticated and redirects the browser back to the SP. The message exchange is shown in Figure 3.

3.3 One-Time-Password from PC to SMS

The next two solutions make use of the OTP principle [4] for authentication. The solutions describe different ways to securely use the mobile phone as an OTP token.

The authentication procedure for the OTP from PC to SMS solution is shown in Figure 4. When the client wants to access the SP the client's identity is requested. The client responds by typing his username in the browser. The message is relayed to the AS which handles the authentication. Upon receiving

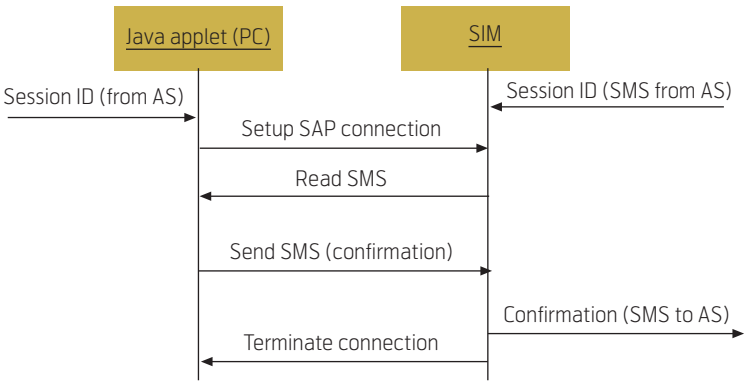


Figure 3 Automatic check of session ID

ISSN 0085-7130

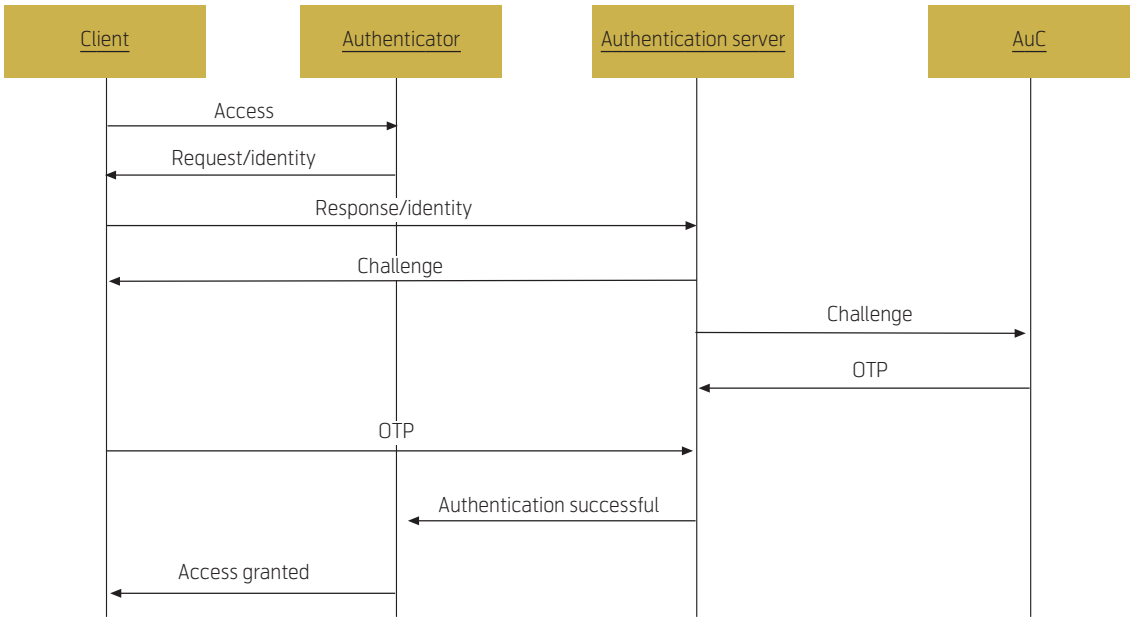


Figure 4 OTP from PC to phone

the client's identity the AS generates a challenge, typically a random number based on the client's profile, and a corresponding OTP. Then the AS sends the challenge to the client. The client enters the challenge on the mobile phone. The OTP applet on the SIM card generates an OTP from the challenge. The OTP is then sent to the AS by SMS. The AS compares the calculated and received OTP and notifies the authenticator that the client is authenticated. The browser is redirected back to the SP and the user is successfully logged in.

3.3.1 Manual Variant

When the client receives the challenge from the AS it is displayed in the browser. The user starts a MIDlet on the phone which can communicate with the SIM card through SATSA-APDU [5] as shown in Figure 5. The user is prompted for the challenge and enters it on the phone. The MIDlet transfers the challenge to the OTP applet which responds with an OTP. The user creates an SMS containing the OTP and sends it to the AS.

3.3.2 Automatic Variant

If the mobile phone and computer are connected through Bluetooth the challenge can be sent to the phone automatically. To realize this, a Java applet will run in the browser and communicates with the Java MIDlet on the mobile phone. The MIDlet on the phone also has to be expanded so that it can create and send an SMS with the OTP to the AS. This can be done by the Wireless Message API[6].

When the user wants to log in using the automatic solution the first thing to be done is to pair the mobile phone and the computer. If this is the first time these two devices are paired a pass-phrase must be entered on both devices. When the pairing is done the user can choose to log in with the automatic solution and after providing an identity the rest of the procedure will go automatically.

After providing the identity a challenge is sent to the user. The Java applet retrieves the challenge and contacts the MIDlet on the mobile phone. When the Bluetooth connection is set up the challenge is sent to the MIDlet. The MIDlet contacts the SIM card as shown in Figure 5. When the OTP is created the MIDlet creates an SMS containing the OTP and sends it by the GSM network to the AS.

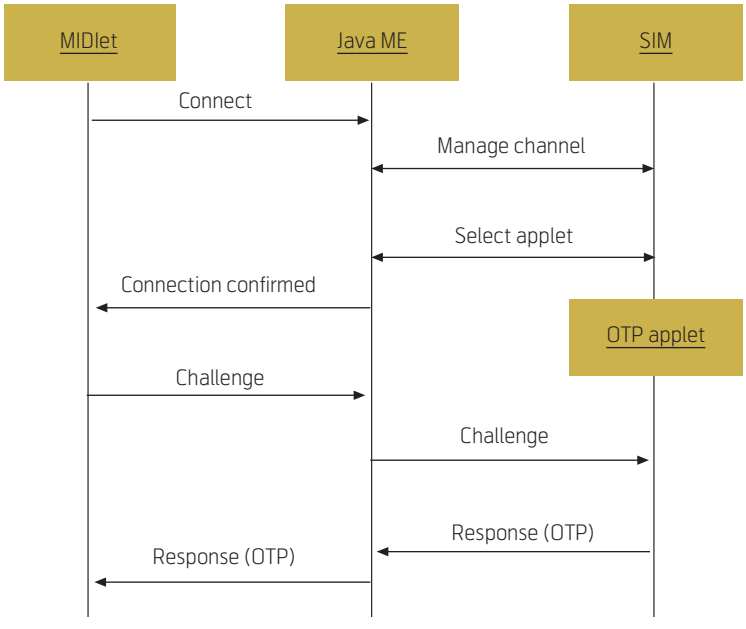


Figure 5 OTP Applet

ISSN 0085-7130©Telenor ASA 2007

3.4 One-Time-Password from SMS to PC

This solution builds on the same principle as the session ID check, that a user with a working phone is already authenticated through the GSM network. The difference is that the check is done by the server and therefore relieves the user from this burden.

The user starts the authentication procedure by entering his username. The session is redirected to the AS which creates an OTP based on the user's identity by a cryptographic hash function. The OTP is then sent to the user by SMS. When receiving the SMS the user types the OTP in the browser. The AS verifies that the OTP is correct and redirects the browser back to the service provider and the user is logged in. Figure 6 shows the message exchange of this solution.

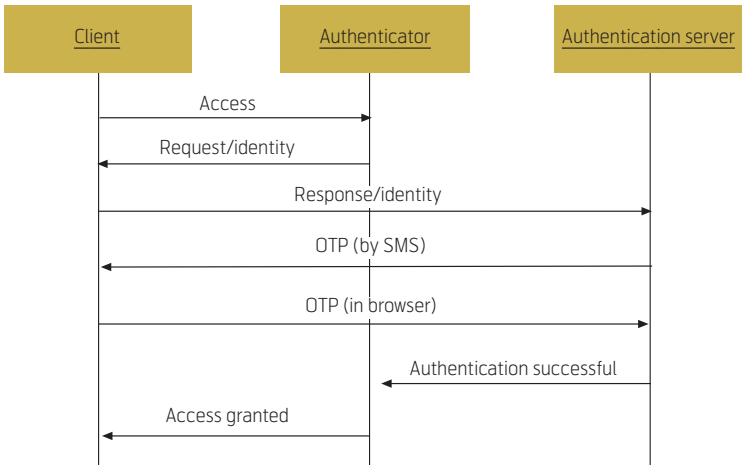


Figure 6 OTP SMS to PC

3.4.1 Manual Variant

When the user receives the SMS with the OTP from the AS he types the OTP in the browser and presses the log-in button. If the OTP is correct the user is successfully logged in.

3.4.2 Automatic Variant

If the mobile phone and the computer are paired through Bluetooth the OTP can be transferred automatically from the phone to the computer and then forwarded to the AS through the browser. This can be handled by a Java applet on the computer communicating with the SIM card through SAP. When the AS has sent the OTP by SMS it notifies the client that this has been done. When receiving this notification the Java application contacts the mobile phone and retrieves the SMS from the AS. The applet retrieves the OTP from the SMS and sends it to the AS through the Internet browser. As in the session ID solution the TP-PID field in the SMS header must be set to "SIM DATA DOWNLOAD" so that the SMS is guaranteed to be stored on the SIM card.

3.5 SIM Strong Authentication via Mobile Phones

This solution makes use of the EAP-SIM protocol [7] to authenticate the user. EAP-SIM is run between the SIM and the AS. The protocol can be run through the computer over Bluetooth and Internet or over the GSM network by SMS.

When the user accesses an SP the browser is redirected to an AS. If the user chooses to run the SIM strong authentication the rest of the procedure is hidden for the user as the SIM card and the AS authenticate each other. If the authentication is successful the browser is redirected back to the SP and the user is logged in.

In order to be able to run the EAP-SIM protocol between the AS and the client the AS needs to communicate with the SIM card. To avoid having to install specific software on the client's computer this communication is handled by a Java applet [8].

The applet will play the role as supplicant and run in the client's browser and relay messages between the authenticator and the SIM card.

Its main functions will be to

1 Receive EAP request from the EAP authenticator;
2 Send the content of the these packets to the SIM card;
3 Build EAP response packets from the SIM responses;
4 Send the EAP response packets to the EAP authenticator.

The applet communicates with the SIM card using Bluetooth SIM Access Protocol (SAP) [9]. The EAP authenticator is implemented as a Java Servlet running inside the AS. The authenticator first request an EAP identity from the supplicant. The supplicant translates this request and relays it to the SIM card. The SIM card responds with the international mobile subscriber identity (IMSI). The authenticator sends the identity received to the AAA server which contacts the user's AuC for GSM triplets. The challenges contained in the triplets are concatenated and sent back to the supplicant. The supplicant relays the challenges to the SIM card that

1 Authenticates the server by verifying that the MAC1 received is in order;
2 Runs the GSM algorithms to calculate a MAC2 that is sent back to the AS.

ISSN 0085-7130

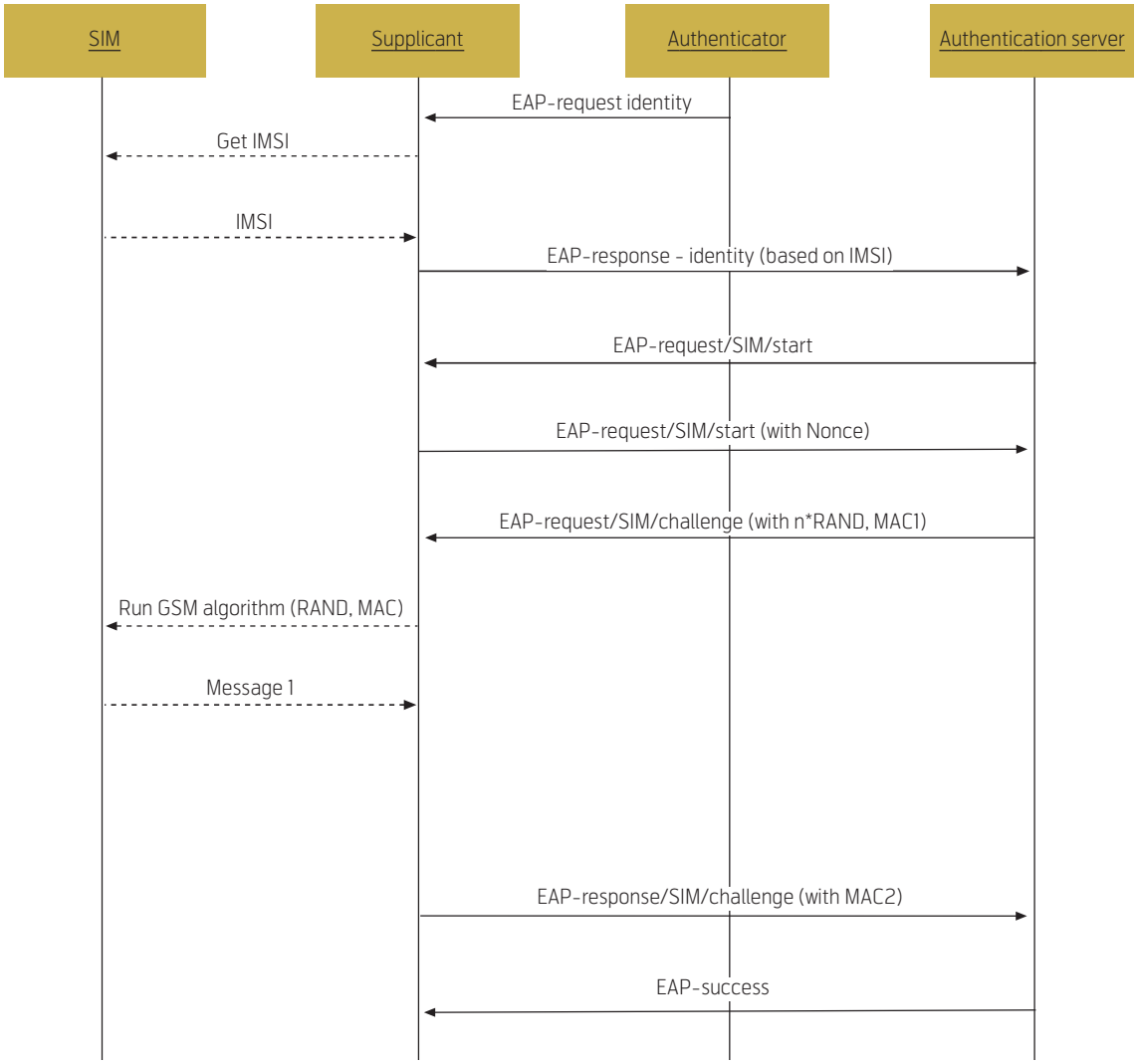


Figure 7 EAP SIM authentication

The supplicant receives the response from the SIM and sends them to the authenticator in EAP format. The authenticator relays the response to the AAA server that verifies that the MAC2 is correct. If MAC2 is correct the authentication is successful and the user is logged in. The EAP-SIM authentication is shown in Figure 7.

3.5.1 SMS Variant

This solution is a variant that does not require a Bluetooth enabled mobile phone. Instead the EAP-SIM protocol will run over SMS.

The user chooses to log in using the EAP over SMS solution. When this is done, a web page is presented asking for a session ID. An authentication applet on the SIM card is started by the AS through SAT and the user is asked to confirm that he wants to start the authentication. A session ID is displayed on the mobile phone screen and the user enters the session ID in the browser. If the session ID is correct the user must accept to start authentication by pressing an OK button on the phone. Then EAP-SIM authentication is performed between the SIM card and the AAA server by exchanging SMS messages. If the authentication is successful, the SP is notified that the user with the corresponding session ID is authenticated, and the user is provided access to the protected resources.

To enable the phone to perform EAP authentication over SMS an applet on the SIM card must be installed. The applet generates the session ID to be entered in the browser. Thereafter, the SIM applet performs authentication towards GSM using SMS messages which encapsulate the EAP-SIM protocol. This solution is similar to the previous, however it does not require the Bluetooth connection between the mobile phone and the user's PC since it instead uses SMS to close the authentication loop. SAP is used to control the SIM card and mobile phone during authentication.

The implementation can be optimized so that only two mobile-originated short messages and one server-originated short message are required for full EAP-SIM authentication.

ISSN 0085-7130©Telenor ASA 2007

4 Evaluation

4.1 Criteria

This section provides an evaluation of the different authentication schemes using the mobile phone as authentication token. The evaluation is performed according to the following criteria:

- Strength & vulnerability
- Cost
- User friendliness

As Figure 8 illustrates, the different authentication solutions can be subject to attacks in several areas, depending on the specific scheme:

1 On the mobile phone;
2 Across the Bluetooth connection;
3 On the computer;
4 Across the Internet connection;
5 Across the connection between service provider and authentication server;
6 On and between GSM network components and connections.

The following sections will discuss some of the most probable attacks on the different mobile authentication solutions.

4.2 Session-ID check with SMS

4.2.1 Manual

With regard to Figure 8, this solution is realistically most susceptible to an attack in the response phase, i.e. when the user sends an acknowledgement back to the AS through the GSM network. Consider a possible hijacker whose goal it is to steal a session from a valid user. If the hijacker establishes a session towards a service at the same time as the valid user, using the valid user's identity (i.e. cellular phone number), two messages with session-ids will be sent to the valid user. Being unaware of the hijack attempt, the valid user might respond with acceptance to the invalid user's request, instead of to the valid user's request.

However, such an attempt requires a hijacker to be familiar with the valid user's identity as well as being well synchronised with the valid user's activitity (e.g. through visual observation). The likeliness of such a successful hijack attempt is therefore in practice extremely low, since it also assumes that the user does not properly study each of the incoming SMS messages with session-ids. Another problem with the solution is that it is technically possible to spoof SMS sender addresses, i.e. send SMS messages with the valid user's number. Additional security mechanisms should be applied to prevent this.

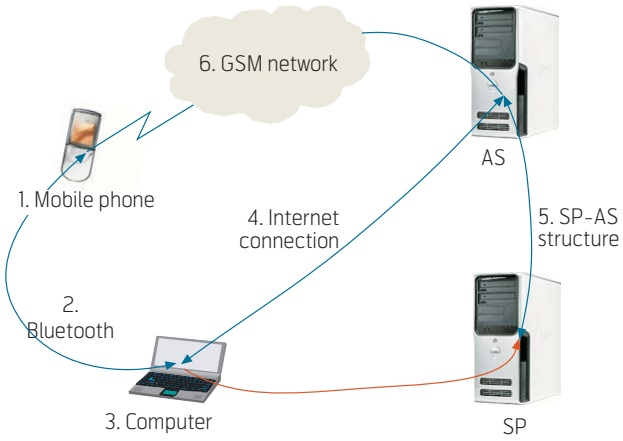


Figure 8 Possible points of attack in mobile authentication solutions

4.2.2 Automatic

This solution, being rather similar to the previous one, eliminates the human factor in the verification of the received session-id in the SMS. Instead, it introduces another point of possible attack; the Bluetooth connection (2) between the mobile phone and the user's computer. The equality of the session-id on the mobile phone and the user's computer is performed over this Bluetooth connection. Consider a hijacker which is able to intercept communication over this connection (man-in-the-middle attack), thus being able to receive requests over it as well as sending responses in return. A hijacker could then, similarly to in 4.2.1, establish a session towards the service in question with the user's identity (cellular phone number). An SMS will then be received by the valid user's mobile phone, with the appropriate session-id. However, if the hijacker is able to redirect his Java Applet so that it contacts the valid user's mobile phone, instead of his own, to access the valid session-id, he can in theory hijack a user session, and this could happen even when the valid user is not actually in the process of establishing a session towards the service at the same time. The Java Applet must also send the confirmation SMS through the valid user's mobile phone, since the AS will check that the confirmation comes from the appropriate cellular phone number.

The Bluetooth connection between the user computer and the mobile phone is protected by encryption using a 16 digit key (as required by SAP) which is used in the pairing process, so the feasibility to establish such an attack is very limited.

ISSN 0085-7130©Telenor ASA 2007

4.3 OTP PC-to-SMS

4.3.1 Manual

In this solution, it is the OTP which is the most crucial element. If an eavesdropper can get hold of the OTP, he could in theory hijack a user session. However, since a user's session (and OTP) is associated with a specific challenge, it is not enough to get hold of the OTP, it is also necessary to hijack the HTTP-session which has previously been created by the valid user. The solution could also in addition to the OTP include a check of the sender address of the SMS carrying the OTP, to further improve the strength and reduce the possibility of malicious attacks.

4.3.2 Automatic

The automatic variant of the OTP-solution mostly improves the user-friendliness. However, it also reduces the possibility of attacks due to a strong association between the user's phone and computer through an authenticated and encrypted Bluetooth connection. Attacks towards this solution could be launched on the Bluetooth-connection, but this is not trivial.

4.4 OTP SMS-to-PC

4.4.1 Manual

Since the OTP in this solution is sent to the user, attacks must be directed towards obtaining the OTP. However, the OTP is typically associated with an HTTP-session created by the valid user. Therefore, both the OTP and the HTTP-session must be attacked in order for a malicious user to get access to the valid user's services.

4.4.2 Automatic

This solution simplifies the work of the user, and prevents eavesdroppers from visually getting hold of the OTP which is received by the valid user's mobile phone. The OTP is never actually exposed to the user, but handled only by the system components.

The solutions are rated with relative weights from 1 to 6, with 6 as the best.

4.5 SIM Strong Authentication

The SIM strong authentication solution can be used with the SIM-card in the mobile phone, but also with the SIM-card in a USB-dongle or similar. The two different solutions differ in some of their properties.

4.5.1 SIM in Mobile Phone

From a user's perspective, this is the most ideal solution, since no additional device is required to perform authentication. Regarding the security properties, this solution requires a wireless Bluetooth connection between the phone and the user's computer, which in theory could be compromised, but as previously discussed, this is in reality unfeasible. The biggest security risk with the solution is loss of terminal, but the SIM is also protected by a PIN-code, which renders the device useless when unknown. The PIN-code must be used to activate the SIM when the phone is switched on, but in addition, the solution can require the user to also enter the PIN-code each time an authentication is to be performed, further strengthening the solution in case of loss of mobile phone.

4.5.2 SIM in USB-dongle

This solution is slightly less user friendly than the previous one due to the need for an additional device, and in theory a bit more secure since it requires a physical connection between the user's computer and the SIM-card. This solution also requires a PIN-code for activation of the SIM-card prior to authentication, which prevents use by arbitrary users if the dongle with SIM is lost.

4.5.3 SIM Strong Authentication with SMS

This solution combines the strength of SIM strong authentication with the session-id check, and the result is a relatively strong authentication solution. However, it requires the installation of a special applet on the user's SIM-card. The solution also requires a check of session-id, and could also possibly include the PIN-verification procedure to provide further protection in case of loss of mobile phone.

All the SIM strong authentication solutions above are based on well-proven, standardised technologies and protocols where the strength and weaknesses have been scrutinized by computer and communication network security experts.

4.6 Comparison of Authentication Solutions

Table 1 illustrates some of the security properties of the major categories of authentication solutions discussed in this paper. It shows what type of attacks the different solutions can be susceptible to.

Figure 9 shows a comparison of the different authentication solutions with respect to four parameters; security, cost, infrastructure and user-friendliness. The comparison is informal and only meant to provide some insight into the properties of the various solutions in relation to each other. As with all such solutions, the actual implementations might have other properties than illustrated here. The scores are relative from 1 to 10, with 10 as the best rating. High rating in e.g. cost and infrastructure does not mean high cost, nor a lot of infrastructure, but rather the opposite.

ISSN 0085-7130

5 Conclusion

This paper has suggested several ways to employ the mobile phone as an authentication token, with the purpose of addressing shortcomings of existing authentication solutions on the Internet today. The solutions show that there are many different possibilities, and that the authentication process can take one single path back and forth, or exploit several channels (multi-channel authentication) to complete the authentication. The paper also informally illustrates the security properties of the solutions and provides a comparison between the solutions with respect to four important criteria.

Security requirement	Session ID	OTP solutions	SIM strong solutions
Online guessing	n/a	√	n/a
Replay	√	√	√
Eavesdropping		√	√
Shared secrets not revealed	n/a	√	√
Multi-factor authentication	√	√	√
Man-in-the-middle			√
Session hijacking			√
Authenticated data transfer			√

Table 1 Security properties

References

1 *OATH – An industry roadmap for Open Strong Authentication*. October 22, 2007 [online] – URL: http://www.openauthentication.org/

2 3GPP. *Technical realization of the Short Message Service*. 2003. (3GPP TS 23.040)

3 ETSI. *GSM 11.14 Specification of the SIM Application Toolkit for the Subscriber Module – Mobile Equipment (SIM – ME) Interface*. Sophia Antipolis, 1996.

4 Haller, N, Metz, C, Nesser, P, Straw, M. *A One-Time Password System*. IETF, 1989. (RFC 2289)

5 Cricco, R, Vinson, Y. *Integrating the SIM Card into J2ME as a Security Element*, 2005. October 2006 [online] – URL: http://www.gemplus.com/pss/telecom/download/jsr177_whitepaper_april05.pdf

6 Ortiz, C E. *The Wireless Messaging API*, 2002. March 2007 [online] – URL: http://developers.sun.com/techtopics/mobility/midp/articles/wma/

7 Aboba, B et al. *Extensible Authentication Protocol*. IETF, 2004.

8 Lunde, L, Wangensteen, A. Using SIM for strong end-to-end application authentication. In: *Telematics*. Trondheim, NTNU, 2006. (MSc thesis)

9 Bluetooth-SIG. *Bluetooth specification SIM Access Profile v.10*, 2005.

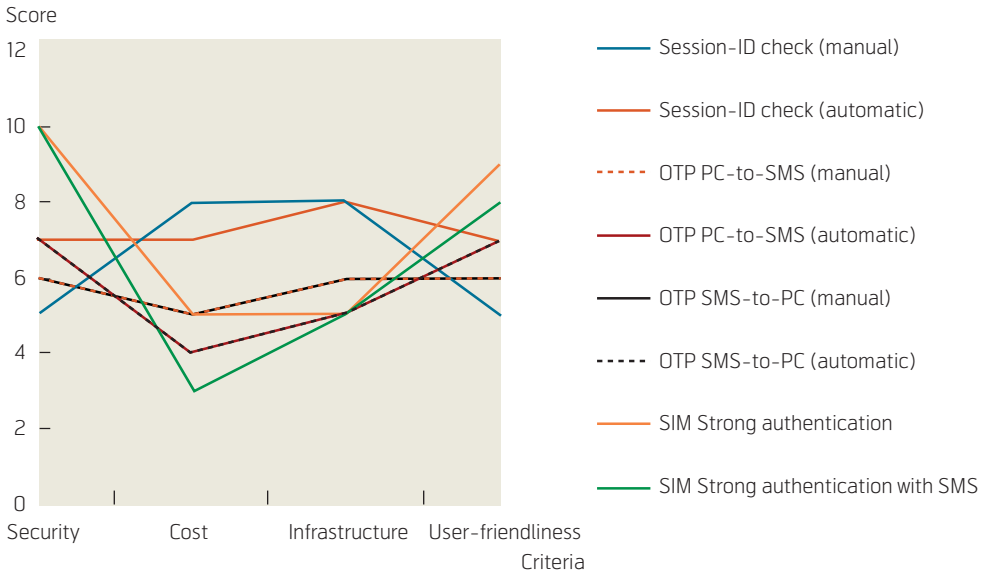


Figure 9 Comparison of the different authentication solutions based on the mobile phone

For a presentation of Ivar Jørstad and Do Van Thanh, please turn to page 10 and 2, respectively.

ISSN 0085-7130©Telenor ASA 2007

Identity Management in a Fixed-Mobile Convergent IMS Environment

BONING FENG, DO VAN THUAN, IVAR JØRSTAD, TORE JØNVIK, DO VAN THANH

Boning Feng is Associate Professor at Oslo University College

Do Van Thuan is Lead Scientist in Linus AS

Ivar Jørstad is CEO of Ubisafe AS

Tore Jønvik is Associate Professor at Oslo University College

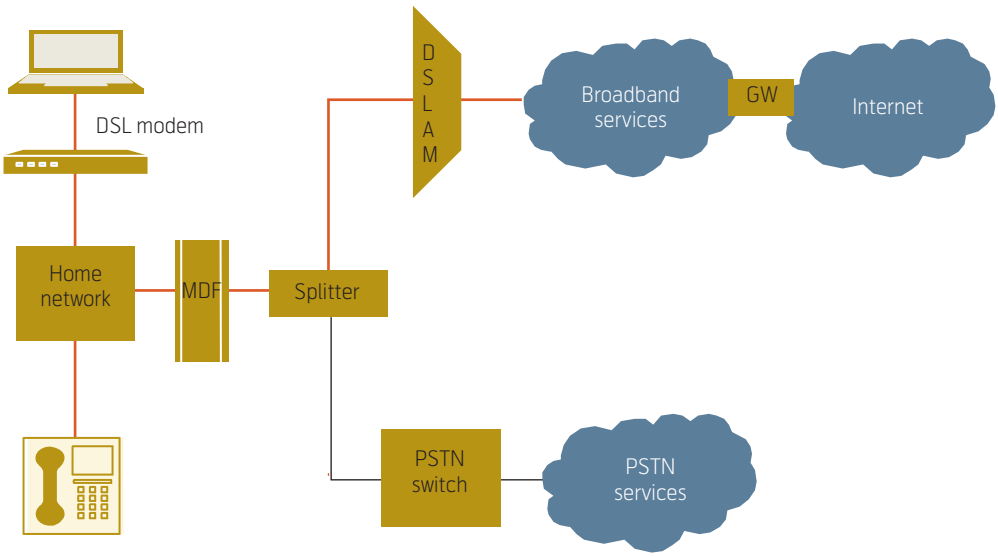
Do Van Thanh is Senior Research Scientist in Telenor R&I

Although originally intended for mobile networks, IMS (IP Multimedia Subsystem) [1] is now proposed also for the fixed network. However, the convergence dream cannot be realised before some serious problems are solved. One of them is the fundamental difference in how the user's identities are viewed and defined in the fixed and mobile world. The goal of this paper is to shed light on the identity management in the fixed and mobile worlds. In this paper, the fundamental differences between the mobile network and fixed one regarding identity pose a great deal of problems for the establishment of a uniform and consistent fixed mobile convergent IMS environment. Fortunately, by appropriate federation of identities, it is possible to realize a complete subscription for a household that comprises both mobile and fixed network. By federation, it is also possible to unify two originally separate subscriptions offered by different operators. Last but not least, interoperability and service continuity between SIP [2] and IMS can also be realized using identity federation. In the paper, only the federation of identity is considered.

1 Introduction

Fixed-mobile convergence has been an objective for the telecommunication world in the past two decades but so far it is still an unrealized dream. The first attempt proposed to equip mobile terminals with DECT (Digital Enhanced Cordless Telecommunications) [3] in the 1990s which would allow them to connect to the PSTN (Public Switched Telephone Network) base station at home or also the ones in the city. Unfortunately, neither this attempt nor any of its successors managed to succeed. The motivations behind the realization of a fixed-mobile convergent system are, on the one hand the convenience for the users, and on the other hand the ease of management and cost savings for the operators since the same infrastructure can be used in both the fixed and mobile environments. The convergence dream is suddenly revived with the arrival of IMS (IP Multimedia Subsystem) which is originally intended for mobile systems. It is definitely an ingenious idea to extend the IMS usage to the fixed broadband IP network. However, there are certainly many challenges to overcome, and one of them lies probably on the fundamental difference in how the user's identities are viewed and defined in the fixed and mobile worlds.

The goal of this paper is to shed light on the identity management in the fixed and mobile worlds. A proposed solution for integrating the two identity management schemes will also be presented. This paper captures the current results of the EUREKA Mobicome (MObile Fixed COnvergence in Multiaccess



MDF: Main Distribution Frame
DSLAM: Digital subscriber line access multiplexer

Figure 1 Lines (Local loop) sharing by PSTN (ISDN) and DSL

ISSN 0085-7130©Telenor ASA 2007

Environment) project which is aiming to resolve the challenging issues implementing IMS in a fixed-mobile convergent environment.

2 Identity Management in the Fixed Environment

In both Public Switched Telephone Network (PSTN) and Integrated Services Digital Network (ISDN), the subscription is based on the physical lines, either analogue (PSTN) or digital (ISDN) going to the subscriber's premise. Before deregulation the subscription price includes the price of line rental and the price of delivering traffic to and from the customers plus universal service commitments and funds for PSTN/ISDN service providers.

No identification or authentication of the user is necessary since the identity of the subscriber is assimilated to the identity of the physical lines which are fixed. The identity of the lines is a telephone number by which it can be uniquely identified. The phone number is conformed to a format specified by the ITU-T in the recommendation E.164 [4]. A person A calls a person B by dialling his telephone number. The network will use this number to connect the call. Person B might have equipment presenting A's telephone number to check before accepting the call. It is worth noting that person B may have several phones and the call may be accepted by any of them.

In PSTN/ISDN, there is no subscriber identity, no user identity and no terminal identity. The line identity is used as identity for the subscriber and many users can share the same subscription.

The PSTN/ISDN network can determine the traffic to and from a local loop but is not able to identify and determine which user is using which telephone.

With the arrival of *xDSL (Digital Subscriber Line)* [5] access technologies, the same telephone lines are also used for faster data transmissions. A typical configuration is shown in Figure 1. A splitter is used to separate the voice and data traffic and direct them to the respective network. An ADSL (Asymmetric DSL) subscription is in the same way as a PSTN based on the physical lines. In some ADSL configurations, identification and authentication of the subscription are carried out by the DSL modem which is equipped with a username and password for authentication. Both the username and the password must be configured by the subscriber or installer at installation.

An ADSL subscription defines the downstream and upstream bit rate but does not limit how many computers or users are sharing the ADSL connection. In fact, it is common that all members of a household share the same subscription and they can also grant access to visiting friends.

In ADSL, a subscriber identity is defined, but this is also mapped to the physical line identity. A subscription may be shared by several users.

The ADSL network can determine the traffic to and from an ASDL line but is not able to identify and determine which user is using which computer.

3 Identity Management in the Mobile Environment

In the mobile networks, each user may have one or more subscriptions. The subscription may be paid by a third party such as employer, parent, etc. However, it is very seldom that several users share the same subscription on a long term basis.

Each subscription has an identity called *IMSI (International Mobile Subscriber Identity)* which identifies the subscriber uniquely. A subscription is also allocated a public identifier, called *MSISDN (Mobile Subscriber Integrated Services Digital Network Number)* or simply phone number that can be used to address or call the subscriber. The IMSI is stored in a portable tamper-resistant smart card called *SIM (Subscriber Identity Module)* [6] [7] [8]. The mobile phone itself, also called *Mobile Equipment (ME)* has its own identity called *IMEI (International Mobile Equipment Identity)* but cannot operate by itself without a SIM card. In fact, an operating mobile phone consists of an ME and a SIM. At power-on authentication is carried on towards the SIM and the mobile phone is granted connection to the mobile network only after successful authentication.

The SIM card enables *personal or user mobility* because it allows the user to change terminal simply by moving the SIM card to the new terminal. It also enables *terminal mobility* because the mobile phone is allowed to move and change access networks.

In the mobile network, the service provisioning and charging are purely based on the subscriber identity and neither on the network identity nor the terminal identity in use.

In the mobile network, a subscriber identity and a terminal identity are defined. The subscriber identity is usually assimilated to the user identity. A subscription is used by one user.

The mobile network can determine exactly the traffic from and to a mobile phone.

ISSN 0085-7130

4 IMS Identity Management for Mobile Networks

The IP Multimedia Subsystem (IMS) is a mobile network infrastructure, defined by the 3rd Generation Partnership Project (3GPP), enabling Voice over IP (VoIP) and multimedia services over an IP-based infrastructure. The key technology behind IMS is the SIP (Session Initiation Protocol) which is used for multimedia session negotiation and session management.

As shown in Figure 2 the IMS service architecture is composed of three different layers: Transport Layer, Session Control Layer, and Service/Application Layer.

The *Transport layer* initiates and terminates SIP signalling to set up sessions and provide bearer services such as conversion of voice from PSTN or other circuit switched networks to IP packets in the IMS network using Realtime Transport Protocol (RTP).

The *Session Control Layer* is the core of the IMS network. The main component in the core is the *Call Session Control Function (CSCF).*

The CSCF, in essence the SIP proxy server(s) in charge of processing SIP signaling in IMS, comprises of three different servers: the Proxy-CSCF (P-CSCF), the Interrogating-CSCF (I-CSCF) and the Serving-CSCF (S-CSCF).

The *P-CSCF* is the first contact point within the Session Control Layer (the IMS core) from the Transport and Service/Application layers. It acts as a service broker handling all inbound/outbound signalling from an IMS terminal to the rest of the network. The functions handled by the P-CSCF include security & authentication, SIP message verification and compression/decompression of SIP messages (to reduce the time and network load needed to transmit the messages), generation of charging information and possibly a Policy Decision Function (PDF) for resource authorization and Quality of Service (QoS) control. The P-CSCF can be located in both the visiting network and the home network.

The *I-CSCF* is the contact point for all connections destined for that particular network operator and is located at the edge of its administrative domain. It is made visible to other domains by the use of DNS as its address is listed in the DNS records of its home domain. In order to obtain user location information and routing SIP messages to the correct destination, the I-CSCF provides an interface to the network's HSS and Subscriber Location Function (SLF) through the Diameter protocol.

The *S-CSCF* is the central node of the IMS signalling plane and performs session control and registration services (SIP Registrar services) in addition to its SIP proxy server functionality. It retrieves user information, typically authentication vectors and user profiles, from the HSS through the same Diameter interface as the I-CSCF.

The *Service/Application Layer* comprises application and content servers to execute value-added services for the user. Generic service enablers as defined in the IMS standard (such as *presence* and *group list management*) are implemented as services in a SIP Application Server.

The *Home Subscriber Server (HSS)* is the central repository for user subscription data. This includes user profile information, registration data (e.g. authorization, authentication and location data) and the current S-CSCF allocated to each particular user.

In IMS, each user may have one or more IMS subscriptions. Each subscription addresses one subscriber that can be the user or a third party responsible for the payment of the subscription. It is also very seldom that an IMS subscription is shared by several users.

Each IMS subscription is associated with an *IP Multimedia Private Identity (IMPI)* and one or more *IP Multimedia Public Identity (IMPU)* [10]. To securely store the IMS subscriber identity, an *IP Multimedia Service Identity Module (ISIM)* is specified by 3GPP [10]. The ISIM is an application running on a UICC (Universal Integrated Circuit Card) smart card in a 3G mobile telephone in the IP Multimedia Subsystem (IMS). The ISIM application can co-exist with SIM and USIM on the same UICC making it possible to use the same smart card in both GSM networks and earlier releases of UMTS.

The *IMPI* is used for registration, authentication, authorization, accounting and administration. It is only visible to control nodes inside the IMS core network. It is stored in the ISIM and is not accessible to the user.

The *IMPI* has a globally unique identifier assigned by the home network. It has the format of a Uniform Resource Identifier (URIs), that can be digits (a tel-uri, like tel: +47-909 77 102) or alphanumeric identifiers (a sip-uri, like sip:john.doe@example.com). It may also have as identifier an IMSI (International Mobile Subscriber Identity). It contains also the encryption functions and keys necessary for authentication.

ISSN 0085-7130©Telenor ASA 2007

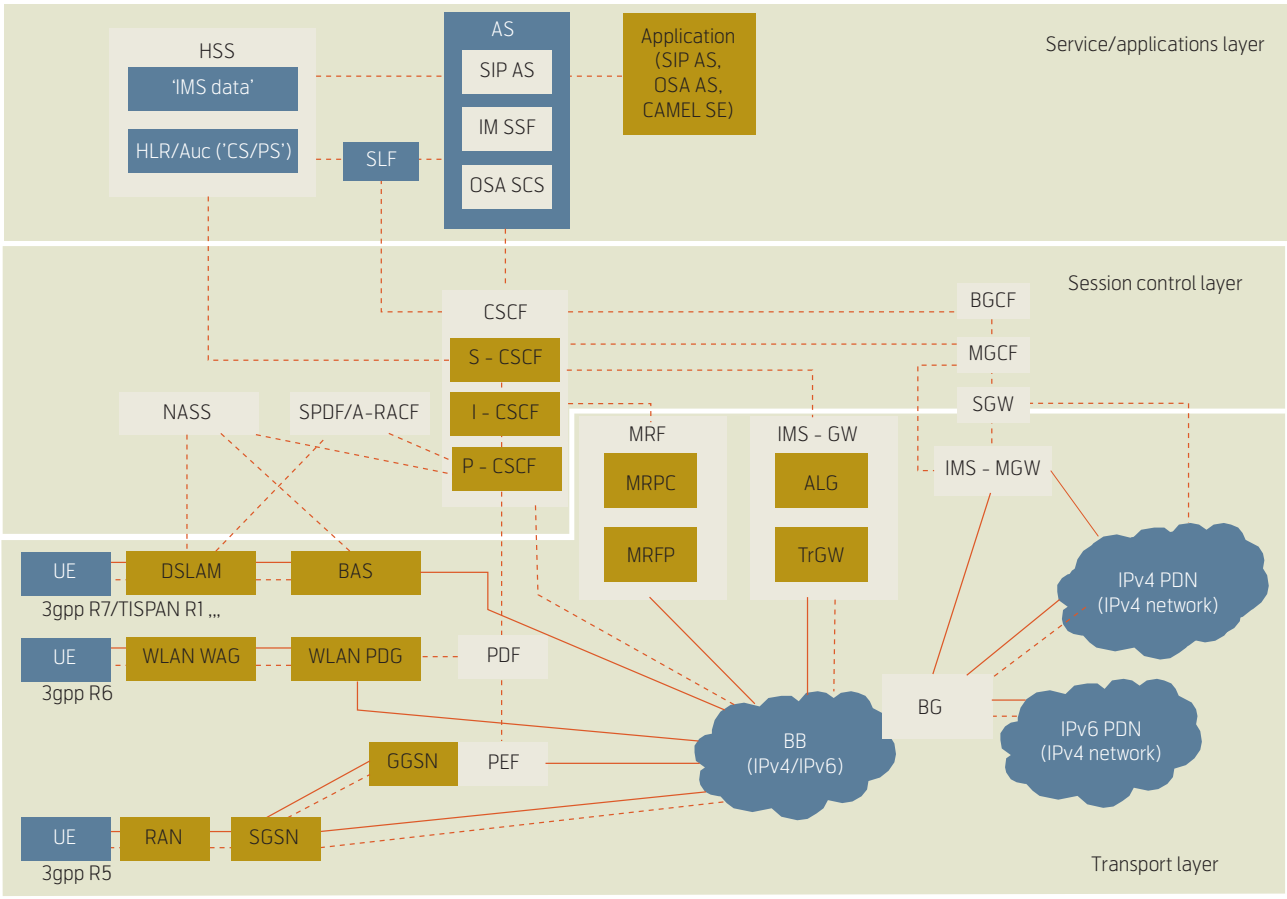


Figure 2 IMS service architecture [12]

The IMPU is a public identity that is visible to the outside world and is used to request and receive sessions. At least one IMPU must be stored in the ISIM and must not be changeable by the user. The IMPU is also in URI format.

As shown in Figure 3 there may be a many-to-many mapping between IMPIs and IMPUs. Each IMPU gets allocated exactly one Service Profile but a Service Profile may be allocated to more than one IMPU. The Service Profile defines the services a user may currently use, e.g. presence service, IP telephony, etc.

4 Identity Challenges in an IMS Fixed-Mobile Convergent Environment

IMS can support WLAN hotspots that are connected directly to the mobile network. In this configuration WLAN can be considered as an alternative access technology to GSM and 3G and a mobile phone with WLAN will not have any problem reaching the IMS core network.

IMS can also be used for fixed broadband networks. As shown in Figure 4, a mobile phone with WLAN when moving to a home WLAN zone may register

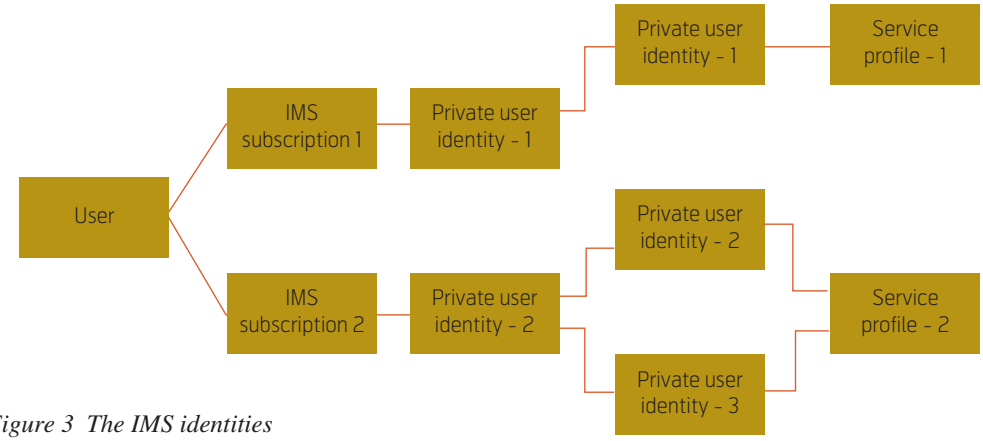


Figure 3 The IMS identities

ISSN 0085-7130

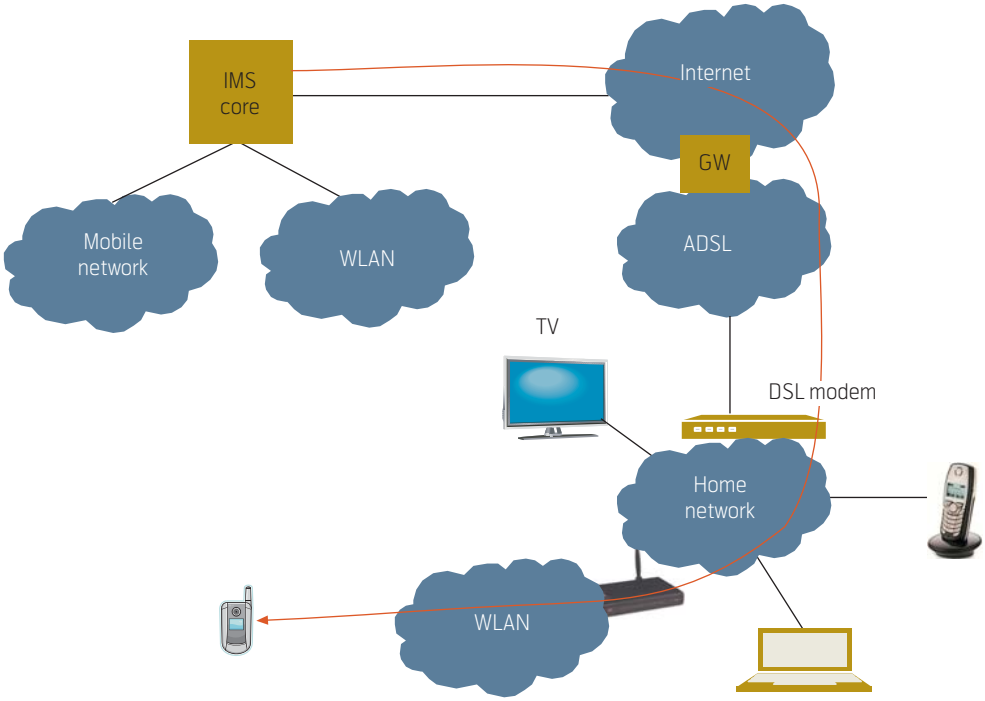


Figure 4 Fixed mobile IMS environment

itself to the IMS core network to receive and make calls without any big problem if WLAN access has been granted. However, fixed mobile convergence is not only about supporting mobile devices in fixed environments but it is also about supporting fixed devices and about offering a unified subscription including both mobile and fixed networks.

The following requirements must be satisfied:

- A subscription may consist of both mobile and fixed services.
- A subscription may include several users.
- A subscription may include several mobile and stationary heterogeneous devices. The number of mobile devices is defined at subscription, while the number of stationary devices can be dynamically changed by the subscriber, i.e. stationary devices can be added or removed according to the subscriber's needs.
- Each user must have a defined identity such that services can be delivered to him or her.
- Each user must be able to use several mobile or stationary devices at the same time or interchangeably, i.e. to make or receive calls interchangeably on different mobile or stationary devices.
- The cost when the services are delivered through the fixed network on mobile or stationary devices should be low or nothing since the connection fees are already paid through the fixed network subscription.

The IMS identity scheme specified by both 3GPP and TISPAN (Telecoms & Internet converged Services & Protocols for Advanced Networks) is very generic and does not prescribe any specific implementation that satisfies the requirements mentioned above. The Eureka Mobicome project has proposed an identity scheme for IMS fixed mobile environment described in Figure 5. A household has a subscription which comprises several users. Each user may have one or more IMS Public User Identity (IMPU) that identifies uniquely the user in the subscription.

Each mobile or stationary device is associated with an IMS Private User Identity (IMPI). The IMPI can be contained in an UICC (Universal Integrated Circuit Card) installed in the device or in common tamper-resistant store as in the case of stationary devices.

Since each mobile device belongs to one user it should be permanently associated with the user's IMPU. Stationary devices on the other hand are usually common devices shared by all the members of the household and should be associated to a main user representing the household, e.g. The Simpsons. It should also be possible to dynamically associate these devices with each member of the Simpsons.

ISSN 0085-7130©Telenor ASA 2007

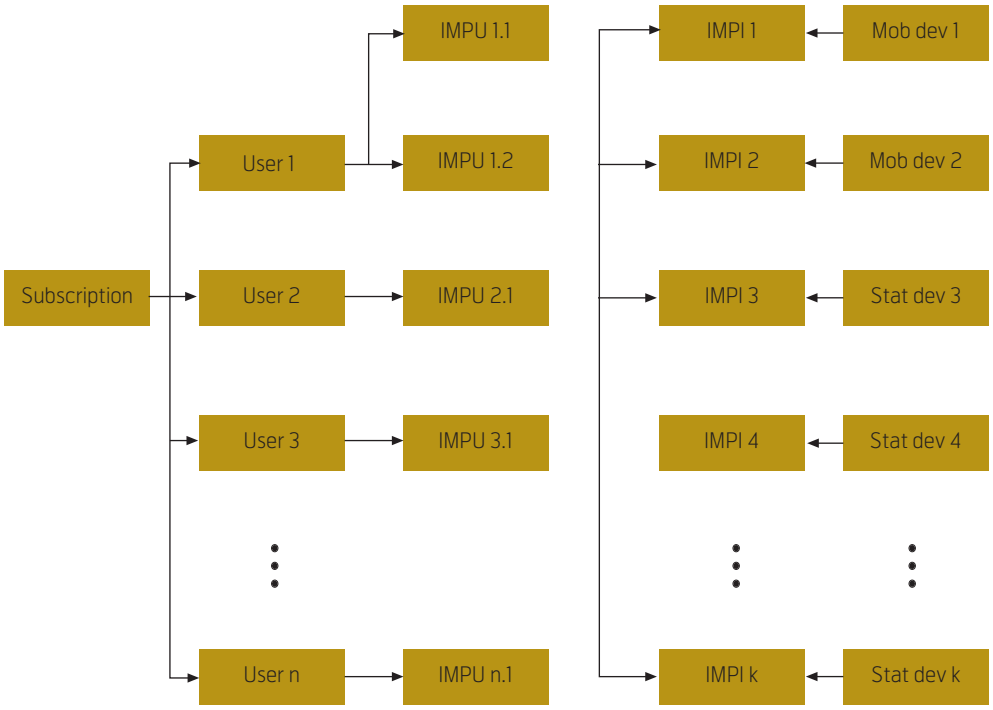


Figure 5 Mobicome identity scheme for fixed mobile IMS convergent environment

To elucidate the Mobicome identity scheme let us take the example of the Simpson family with Homer, Marge, Bart, Lisa and Maggie.

The Simpson's subscription consists of 6 users as shown in Figure 6:

1 Simpsons (main user associated with the Simpson's household), having two IMPUs: simpson@telenor.com and 55599555;

2 Homer, having two IMPUs: homer@telenor.com and 55599556;

3 Marge, having two IMPUs: marge@telenor.com and 55599557;

4 Bart, having two IMPUs: bart@telenor.com and 55599558;

5 Lisa, having two IMPUs: lisa@telenor.com and 55599559;

6 Maggie, having two IMPUs: maggie@telenor.com and 55599560.

Homer, Marge and Bart each have a cellular phone with own IMPIs as follows:

- Homer's cellular phone has the IMPI simpson-mobile1@telenor.com and gets associated permanently with Homer's IMPU 55599556;

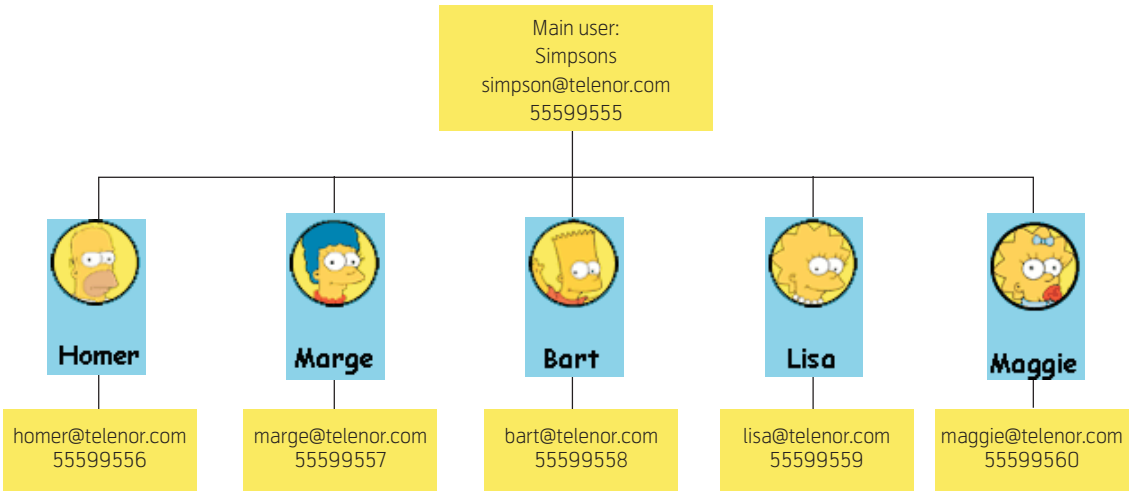


Figure 6 The Simpsons' subscription

ISSN 0085-7130

- Marge's cellular phone has the IMPI: simpson-mobile2@telenor.com and gets associated permanently with Marge's IMPU 55599557;

- Bart's cellular phone has the IMPI: simpson-mobile3@telenor.com and gets associated permanently with Homer's IMPU 55599558.

The Simpsons have three stationary devices that can be shared by all the members as follows:

- An IPTV with IMPI: simpson-stat1@telenor.com;
- A PC with IMPI: simpson-stat2@telenor.com;
- A fixed telephone with IMPI: simpson-stat3@telenor.com.

All the three stationary devices are associated to the main user Simpsons, i.e. to the two IMPUs: simpson@telenor.com and 55599555. This means that when someone calls to simpson@telenor.com or 55599555, all the three devices will ring at the same time or subsequently.

Calls addressed to each individual can be delivered on the respective mobile phone or a registered stationary device. For example, when entering home, Bart can register himself to the IPTV and receive phone calls addressed to him on it. The phone calls to the Simpsons' phone number 55599555 are still delivered subsequently to the stationary device. After a while, he decides to go out. He can deregister himself at the IPTV and the calls to him will be delivered to his mobile phone.

5 Unification of Separate Subscriptions

To improve friendliness and ease of use there should be a possibility to unify separate subscriptions and to get services delivered to the same identities on different devices and networks.

To illustrate let us take the example of Homer. At his workplace, Springfield Nuclear Power Plant, Homer has another IMS subscription with two IMPUs: homer@springfield-power.com and 55577666. He has also two stationary devices: a PC with IMPI, stat1@springfield-power.com and a fixed telephone with IMPI, stat2@springsfield-power.com. He also gets a cellular phone with IMPI, mobile1@ springfield-power.com.

At his office Homer may want to receive calls addressed to his private IMSI, 55599556, on his office fixed telephone or his office cellular phone. He can simply *federate* [11] his home public identities to his office public identities as follows:

- homer@telenor.com *is_federated* to homer@springfield-power.com;

- 55599556 *is_federated* to 55577666.

It is worth noting that the *is_federated* relation is non-symmetric. This means that homer@telenor.com *is_federated* to homer@springfield-power.com does not imply homer@springfield-power.com *is_federated* to homer@telenor.com.

Indeed, calls on homer@telenor.com will be redirected to homer@springfield-power.com but calls on homer@springfield-power.com will not be redirected.

Alternatively, Homer may want to federate his private IMSI to a specific device such as the PC as follows:
- 55599556 *is_federated* to stat1@springfield-power.com.

Adequate authentication must be performed to ensure that the user is the owner of the identities to be federated before allowing federation. In fact, by looking at the office subscription, it is not possible to determine whether 55599556 should be allowed to be federated with stat1@springfield-power.com.

It is crucial that the identity federation is sufficiently flexible to embrace all forms of federation at any time. The IMS user must be equipped with user-friendly tools to carry out the federation in an intuitive way.

After the federation is completed, registration must be carried out in order to activate the transfer of services between the identities. Registration should be done automatically without or with minimum involvement of the user. Indeed, the federation enforcement can be based on a predefined time table. It could be location based, i.e. dependent on the location of the user. It could also be event-based, i.e. it is based on the occurrence of certain events like PC login, detection of mobile phone in the room, etc. A simple client application running on the user's mobile phone may be an ideal tool enabling the user's federation enforcement.

The federation of identities is even more challenging if the user has subscriptions at different operators. Business agreements must be established between the operators before federation can be done.

ISSN 0085-7130©Telenor ASA 2007

6 Interoperability with SIP

No matter how successful IMS is it is reasonable to assume that IP telephony service based on native IETF SIP will co-exist and interoperability issues between IMS and native IETF SIP should be considered carefully. In this section, we focus on the roaming of the IMS mobile phone to native SIP environment.

Before continuing, it is worth noting that IMS is based on SIP but the necessary adaptation for the mobile network makes the 3GPP SIP different and incompatible with the IETF SIP. Consequently, an IMS client running on the mobile phone will not be able to communicate with a SIP server. Currently, many mobile phones with WLAN access are equipped with a SIP user agent and are therefore capable of making VoIP calls.

Let us suppose now that the Simpsons subscribe at home to an IP telephony Service Provider that uses SIP while IMS is offered at Springfield Nuclear Power Plant. Without identity federation, Homer's cellular phone operates as a SIP phone with SIP URI homer@telenor.com. At work, it acts as an IMS phone with IMPU homer@springfield-power.com and mobile1@springfield-power.com. Calls addressed to the home cellular phone will not be delivered at work, and calls addressed to the office cellular phone will not be delivered at home.

To remedy the situation the two identities, homer@telenor.com and homer@springfield-power.com should be federated such that calls can be forwarded to each other (see Figure 7).

7 Conclusion

In this paper, the fundamental differences between the mobile network and the fixed one regarding identity pose a great deal of problems for the establishment of a uniform and consistent fixed mobile convergent IMS environment. Fortunately, by appropriate federation of identities, it is possible to realize a complete subscription for a household that comprises both mobile and fixed networks. By federation, it is also possible to unify two originally separate subscriptions offered by two different operators. Last but not least, interoperability and service continuity between SIP and IMS can also be realized using identity federation. In the paper, only the federation of identity is considered. The functions and capabilities necessary for the federation are not yet studied and they will be taken care of in the Mobicome project.

References

1 3rd Generation Partnership Project. *Technical Specification Group Services and Systems Aspects; Network architecture (Release 8)*, June 2007. (3GPP TS 23.002 V8.0.0)

2 IETF Network Working. *SIP: Session Initiation Protocol*, June 2002. (RFC 3261)

3 ETSI. *Radio Equipment and Systems (RES); Digital European Cordless Telecommunications (DECT) – Common Interface Part 1: Overview*. Sophia Antipolis, 1999. (ETS 300 175-1 V1.4.2-1999-06)

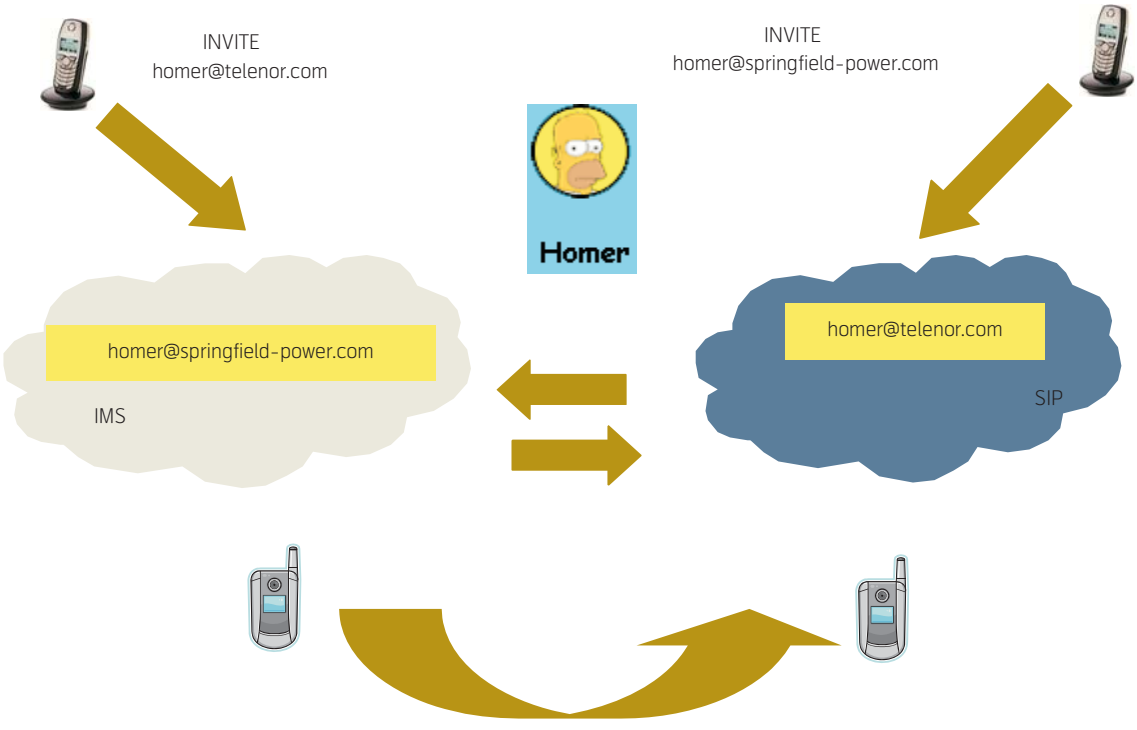


Figure 7 Federation of SIP and IMS identities

ISSN 0085-7130©Telenor ASA 2007

4 ITU. *Series E: Overall Network Operation, Telephone Service, Service Operation and Human Factors – International operation – Numbering plan of the international telephone service*. Geneva, International Telecommunication Union, October 2003. (ITU-T E.164.1)

5 CISCO. *Internetworking Technology Overview – Chapter 15: Digital Subscriber Line*. Cisco, June 1999.

6 3GPP. *Digital cellular telecommunications system (Phase 2+); Subscriber Identity Modules (SIM); Functional characteristics*. November 1999. (3GPP GSM 02.17 V8.0.0)

7 3GPP. *Digital cellular telecommunications system (Phase 2); Specification of the Subscriber Identity Module – Mobile Equipment (SIM – ME) interface*. December 1999. (3GPP GSM 11.11 V4.21.1)

8 3GPP. *Specification of the SIM Application Toolkit for the Subscriber Identity Module – Mobile Equipment (SIM – ME) interface*. 1999. (3GPP TS 11.14 V8.17.0)

9 3GPP. *Technical Specification Group Services and Systems Aspects – Network architecture (Release 8)*. September 2007. (3GPP TS 23.002 V8.1.0)

10 3GPP. *Technical Specification Group Core Network and Terminals; Characteristics of the IP Multimedia Services Identity Module (ISIM) application (Release 7)*. September 2007. (3GPP TS 31.103 V7.3.0)

11 Wikipedia. *Federated identity*. October 22, 2007 [online] – URL: http://en.wikipedia.org/wiki/Federated_identity

12 Wikipedia. *IP Multimedia subsystem*. October 22, 2007 [online] – URL: http://en.wikipedia.org/wiki/IP_Multimedia_Subsystem#Architecture

Boning Feng received his ME and PhD degrees in telematics from the Norwegian Institute of Technology (NTH), Trondheim, Norway in 1985 and 1990, respectively. Since 2005 he has been Associate Professor at Oslo University College (OUC), Faculty of Engineering. Prior to that he was Senior Research Scientist in Telenor R&D (1998–2004), and Associate Professor at Norwegian University of Science and Technology (NTNU, 1990–1997). He has worked in various fields associated with telecommunication services, internet protocols, traffic analysis, control aspects of networks, and optical communication. His current research interests include network architecture and services, IP Multimedia subsystem (IMS), and fixed mobile convergence.

email: Boning.Feng@iu.hio.no

For a presentation of Do Van Thuan, Ivar Jørstad, Tore Jønvik, and Do Van Thanh, please turn to page 18, 10, 135, and 2, respectively.

ISSN 0085-7130©Telenor ASA 2007

Access Control and Privacy Enhancement through Role-based Identity Management

MOHAMMAD M.R. CHOWDHURY, JOSEF NOLL

Mohammad M.R. Chowdhury is a PhD student at UniK, Kjeller

Josef Noll holds a professor stipend at University of Oslo/ UniK

Managing user identities for information security and privacy in today's connected systems is a crucial issue. This paper focuses on the access control and privacy problems in a project based business environment to access project resources and to maintain privacy of members. In this regard, a semantic ontology is proposed which formalizes roles of the members, and controls access to project resources by means of formalized privacy policies and rules.

1 Introduction

Currently, managing various forms of identities to represent people on the web is crucial for secure service access and privacy. Today's connected systems often contain sensitive information; there is an increased need for adequate security and privacy support. We believe that capabilities of semantic technology can contribute to providing solutions to these problems. This paper is proposing such a solution which is expected to handle the identity management and privacy issues in business organizations. Each person possesses certain privileges to access resources based on the roles he/she plays in an organisation. To provide these services, we have formulated policies and rules to control access to resources such as project documents. These role-based policies need to be computer readable, which is achieved through a formal representation of a domain. Semantic technologies enable such a computer readable presentation, and are introduced here to handle the growing need of identity management and privacy support in corporate network.

2 Motivation

A project oriented working culture is a common scenario in a business environment. Projects are often set up across organizations, which limits the usability of company internal content management systems. Members in a project have certain roles and based on these roles they enjoy certain rights and privileges in a project. Role-based identity management can facilitate access control and privacy enhancement provisions for service access in a business project.

Having these aspects in mind, we have designed a use case scenario (Figure 1) targeting a business environment. A fictive project named UMTS Release 9 roll-out (Rel9) is created by Telenor and Ericsson. Telenor members in this project are György Kalman and Josef Noll, and Ericsson is represented by Erik Swansson. These members have their own supervisors in their parent companies. The project has resources like documents, deliverables, member details etc. With Josef Noll, Telenor has the project leader in the Rel9 project, while the others are ordinary project members. Visitors (example, Geir Egeland) are any persons from Telenor or Ericsson who are not members of the Rel9 project but want to know about the project. Based on the roles, members have differential rights to access project resources. For example, supervisors have the authority to read project documents and have visibility of project member details.

On the basis of this scenario, we have developed access control and privacy enhancement mechanism through roles, policy and rules using semantic web technology which will be discussed in detail in Section 5.

3 Semantics for Access Control

The significance of adding privacy-enhancing technologies (PET) in virtual networks is overwhelming [1], [2]. The project scenario (Section 2 and Figure 1) which is introduced in this paper is similar to a community in business environment having higher access control and privacy requirements. Not much research has focused on providing access control and privacy support in community environments involving semantic technology. Krug et al. provided a solution for community-aware identity management with access rights delegation [3], [4]. Instead of maintaining a centralised access control list, a trust based access right group has been proposed to delegate access rights. A private key based signature scheme was proposed to ensure the privacy of networks and users, which requires secure distribution and maintenance of keys. A similar concept of trust has been used by [5] to create and access community resources. A distributed trust management approach is also considered as one of the main components to secure the Semantic Web [6]. The authors intended to provide access to community and privacy solutions only by means of trust or reputation management; however, this does not provide adequate security in business contexts, which require a higher level of

ISSN 0085-7130

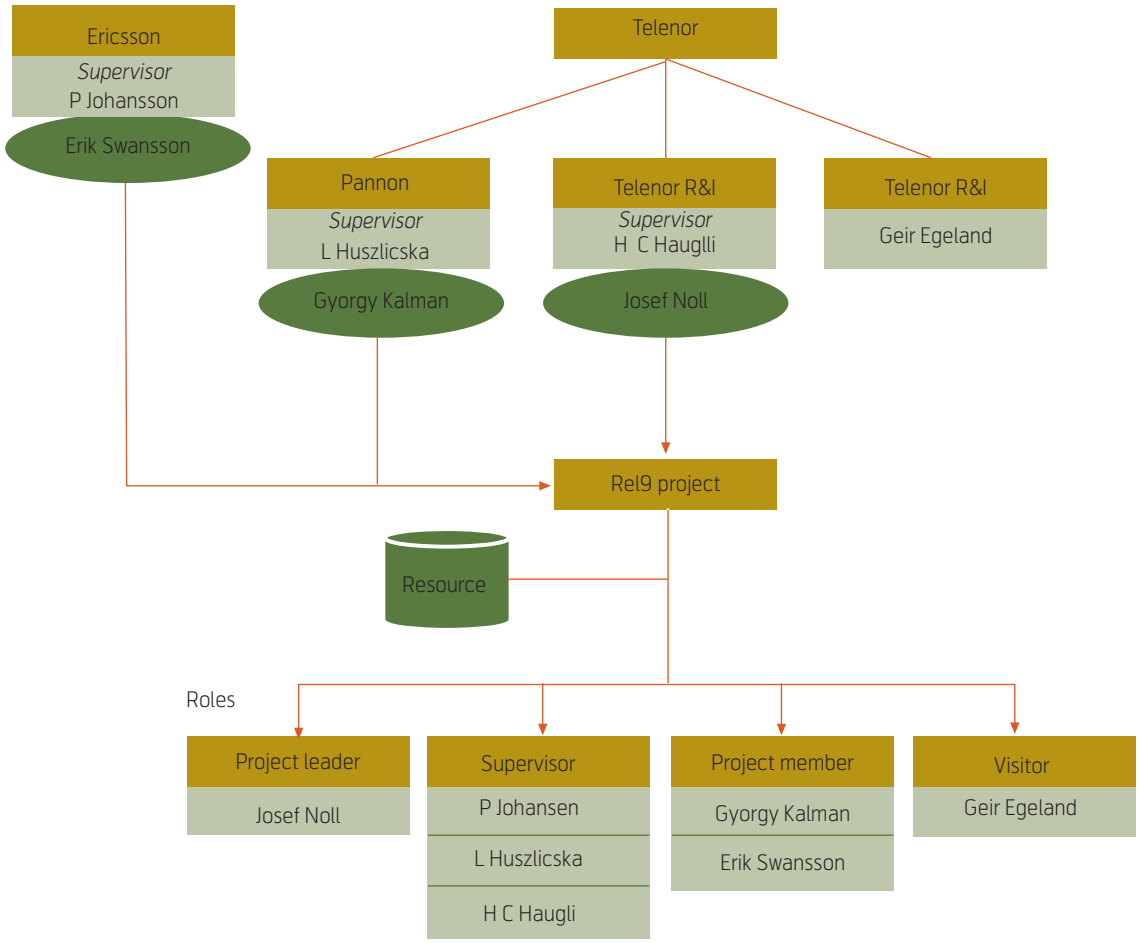


Figure 1 Use case scenario: Release 9 project

security. Trust is affected by various factors and therefore difficult to quantify.

FOAF (friend of a friend) was used by Krug et al. to delegate trust in a community [3]. Instead of FOAF, this paper uses the Web Ontology Language (OWL) which has more facilities for expressing meaning and semantics than FOAF. Finini proposed to use semantic languages such as OWL for constructing ontologies which define policies in [6]. In another paper [7], Smith introduced role-based access control (RBAC) policy management concepts. The National Aeronautics and Space Administration (NASA) uses semantic technologies like OWL to manage policies and access mechanisms across the organisation. Smith used the algorithms introduced by Kolovski [8]. Few of these concepts have been applied to express policies and rules such as those in our paper. To simplify access control, the Liberty Alliance Project[1)] introduced Circle-of-Trust (COT) to establish a legal framework for identity federation. But it lacks finer granularity of service access rights (based on differential access rights) and privacy. The notion of community-aware service access and privacy assurance can be addressed in a similar manner through the proposed architecture of this paper.

4 User, Device and Service Environment

This section focuses on the challenges of a ubiquitous service environment, supporting the preferences and context of the user and his communication devices.

4.1 Service Environment Scenario

Historically a service centric architecture was introduced to let services communicate with each other. The user- or I-centric approach, postulated by the Wireless World Research Forum (WWRF), is based on the transition of access delivery to service delivery [9]. Current rule-based algorithms become too complex when handling user context and preferences, thus asking for new mechanisms allowing dynamic adaptability of services.

The service centric world was introduced based on service level agreements (SLA) between trusted partners. In a more dynamic service provisioning world,

1) *Liberty Alliance Project, http://www.projectliberty.org/*

ISSN 0085-7130©Telenor ASA 2007

as envisaged in a Semantic Web Services environment, privacy becomes one of the key issues [10]. Our approach is to take advantage of the developments in both worlds, using the privacy and security mechanisms of the I-centric world and combining them with the semantic representation of data as known from the Semantic Web (Services) World [11].

The key challenge in a user-centric approach is the handling of user preferences, context, devices, and connectivity with proper privacy assurance required by these features. The European project ePerSpace introduced personal service delivery in the home segment, based on user profiles and preferences [12]. Experiences from this and similar projects showed that managing and updating preferences is a tedious work. While the home is a rather controlled environment, with trusted and known constellations of devices, the mobile world is more vulnerable. The ever increasing connectivity to the Internet with these mobile devices introduces various security and privacy threats. Service delivery in the mobile/wireless world is more complex. Louis V. Gerstner, Jr. of IBM said: *Picture a day when a billion people will interact with a million e-Businesses via a trillion interconnected, intelligent devices*. Pervasive systems do not just mean computers everywhere; it means computers, networks, applications, and services everywhere. To build personalised services is a challenge to the system design in pervasive environment from a security and privacy point of view.

Service access is coupled to user identity, or a way of proof that *I am the person who is allowed to access/ purchase the service*. Identity is verified through an authentication mechanism. Personalisation is based on handling the user's identity. Approaches for a mathematical description of identities have a long tradition. Khoshaflau claimed back in 1986 the need for a 'strong support of identity', and described identities through a graphical representation [13].

The introduction of semantics and the representation in .rdf and .xml allows describing user preferences and relations to characterise the roles of the users as indicated in the business use case scenario illustrated in Section 2.

4.2 Semantic Service Delivery

New methodologies, techniques and tools are necessary to develop and maintain services for the future that are attractive, easy to use and sufficently cheap. Concepts and technologies like Service Oriented Architectures (SOA), Web Services (WS), Semantic Web (SW) and Semantic Web Services (SWS) have gradually grown up to show their viability, especially if they are used in combination. Semantic Web-based technologies are widely acknowledged to play an important role in solving the interoperability problem between applications; the usage of semantic description in the context of advanced services delivery is expected to support easy access to the services. Not only do such formal and explicit descriptions enable easy service integration, but they will also support the exchange of preferences, profiles and context information of users.

According to the OASIS framework SOA is an architectural paradigm (model) that does not necessarily mean usage of Web Services, although Web Service is a popular implementation [14]. One prototypical implementation of a Semantic SOA platform was performed in the European Research project Adaptive Services Grid[2)] (ASG) in order to dynamically create services for the end user. While a technical implementation of a semantic service platform might be expected in the time frame 2009/2010, issues like privacy and protection of user requests and dynamic service level agreements between service providers might hamper the time to market [15]. Kagal et al. pointed out similar findings and claimed the necessity to extend Web Services in privacy and security [10]. They suggested extending Semantic Web Services with policies, representing security requirements for service discovery and privacy protection of user requests. These mechanisms of semantic technologies are used there to address privacy and security concerns in service delivery. We suggest extending the usage of semantic descriptions to user preferences and context, thus allowing to dismiss only the required information for a specific service request.

The mobile service world has made the move to a Web service oriented architecture. Noll et al. used a semantic annotation of advanced Telecom services to achieve exchange of roaming information on a dynamic basis [16]. The main findings of the approach were the cost reductions in service delivery, due to reduced effort for testing and updating of Web services in a semantic service world.

4.3 Authentication Mechanism

Extending the user preferences and context description in a semantic manner supports the disclosure of just the relevant user information for secure service access. In our scenario, access to Rel9 documents should only be granted to members and superiors of the Rel9 project. Today such access is often secured through directory access mechanisms which have limited functionality and are complex to manage.

2) *http://asg-platform.org*

ISSN 0085-7130

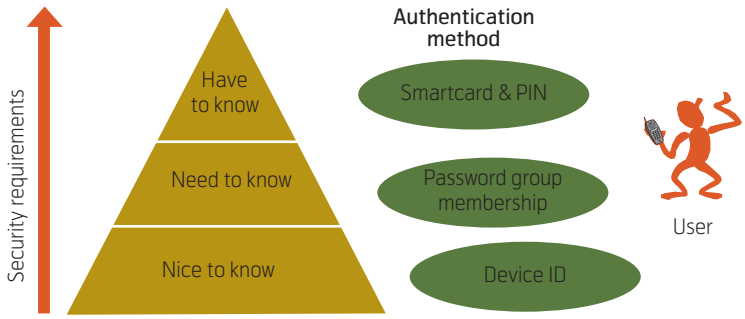


Figure 2 Security requirements for service access

Our approach is to define access rights through corporate relations, e.g. all superiors of Rel9 project members have read access to technical documents produced in the project. Depending on the security requirements of the specific service (see Figure 2), identification can be of type *nice to know*, e.g. using the device identity; *need to know,* through e.g. password, or *have to know,* through e.g. smartcard and pin code. Different identification mechanisms for the variety of services are defined and realise the mechanisms suggested by the Initiative for open authentication[3]; (i) SIM authentication (SIM), (ii) Public Key Infrastructure (PKI), and (iii) One-Time-Password (OTP).

The difference from today's authentication is that a person does not identify himself to a specific service, but is asked to verify his role (e.g. corporate relationship) providing him with the service access. Information about the user will not be disclosed; the service provider will just receive a certificate ensuring that the user has sufficient rights to use the service.

4.4 Mobile Supported Service World

Service access includes more and more the mobile phone, examples of which are admittance and payment services through contactless cards. Near Field Communications (NFC) enables these services on the mobile phone; the technology is prototyped worldwide, e.g. from MasterCard in Dallas [17]. One goal of these field trials is to demonstrate interworking between wireless technologies and NFC, another goal is to address security issues like potential threats as well as identity, privacy and simplicity. Adding NFC capabilities to the mobile phone opens for key exchange through near field and through the mobile network, thus providing a principle way of delivering authentication information. It is assumed that members of the company/project are authenticated through keys. These keys are distributed between the members using short messages (SMS) service or beforehand through NFC technology. This capability of in-band or out-of-band delivery of authentication keys makes the mobile phone a preferred device in administrating access rights.

5 Identity Management

A dynamic service request, taking into account the privacy requirements of a user, can be treated as identity administration. Identity is reputation: *what I say about me and what others say about me* [18]. My reputation is different, depending on whether I am at work, doing sports, or enjoying membership awards in a club. In the virtual world identity handling is more difficult, taking into account the dynamic service requests and privacy requirements of a user. Roccas introduced this in 2002 through the term social identity complexity, defining a new theoretical construct that refers to an individual's subjective representation of the interrelationships between his or her multiple group identities [19].

The Internet was built without an identity layer. In the current Web2.0[4] discussion Identity2.0[5] is introduced to interconnect people, information and software. Various institutes and industries are working to provide better identity management solutions. The Web community has defined Laws of Identity, providing a unifying identity meta-system that can offer the Internet the identity layer it needs [20]. It claims to handle minimal disclosure for Constrained Use, thus the claim to protect the privacy of the user. In Liberty Alliance[6], members are working to build federated identity and interoperability mechanisms in multiple federations. Within this, they are focusing on end user privacy and confidentiality issues and solutions against identity theft. Another solution, Sxip[7] has been designed to address the user-centric identity architecture. It provides user identification, authentication and internet form fill solutions using web interfaces for storing user identity, attribute profiles and facilitating automatic exchange of identity data over the Internet. To access online services, Windows CardSpace[8] uses various virtual cards (mimic physi-

3) *OATH, http://www.openauthentication.org/*
4) *http://en.wikipedia.org/wiki/Web_2*
5) *http://identity20.com/media/OSCON2005/*
6) *http://www.projectliberty.org/*
7) *http://www.sxip.com/*
8) *http://cardspace.netfx3.com/*

ISSN 0085-7130©Telenor ASA 2007

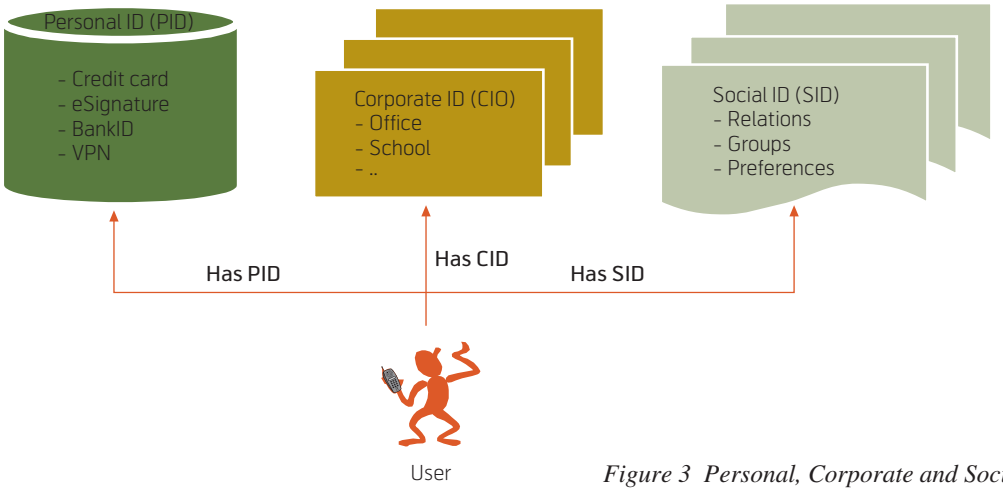


Figure 3 Personal, Corporate and Social Identities

cal cards) issued by the identity providers for user identifications and authentication, each retrieving identity data from an identity provider in a secure manner.

Most of these mechanisms are tailored to foster the usage of identity based web services. Our scheme focuses on the identity management in business community based on the relationships of its actors with the community facilitating the access to its contents and privacy enhancement.

5.1 Representing Identity

The proposed integrated identity mechanism consists of certificates, keys and preferences stored in a personal device and in the network. These identities are categorized in three groups of identity; personal identity (PID), corporate identity (CID) and social identity (SID), based on the roles exercised by a person in real life [21]. Figure 3 shows example applications of PID, CID and SID.

Our approach suggests a decentralised identity architecture, consisting of network components and the personal device of the user. Such an approach brings the user in control of his services, allowing him to accept or deny access to privacy information. The mechanism builds on a personal user device, typically a mobile phone, providing the underlying infrastructure. With the identity subscription certificate users can access the network identity repository, e.g. service references located in the SID. Identities stored in this repository can give access to services (remote or proximity) that need medium or low level of security requirements. The main reason to store service and user preferences in the network is the availability of the network repository and the short response time, avoiding the costly and varying mobile/wireless link. Personal identities (PID) require high security, and will thus be stored in the personal device of the user, allowing him to control when and what PID information is released to service providers.

Semantic Web technology is proposed to represent role-based identity management in areas of business/social resource access.

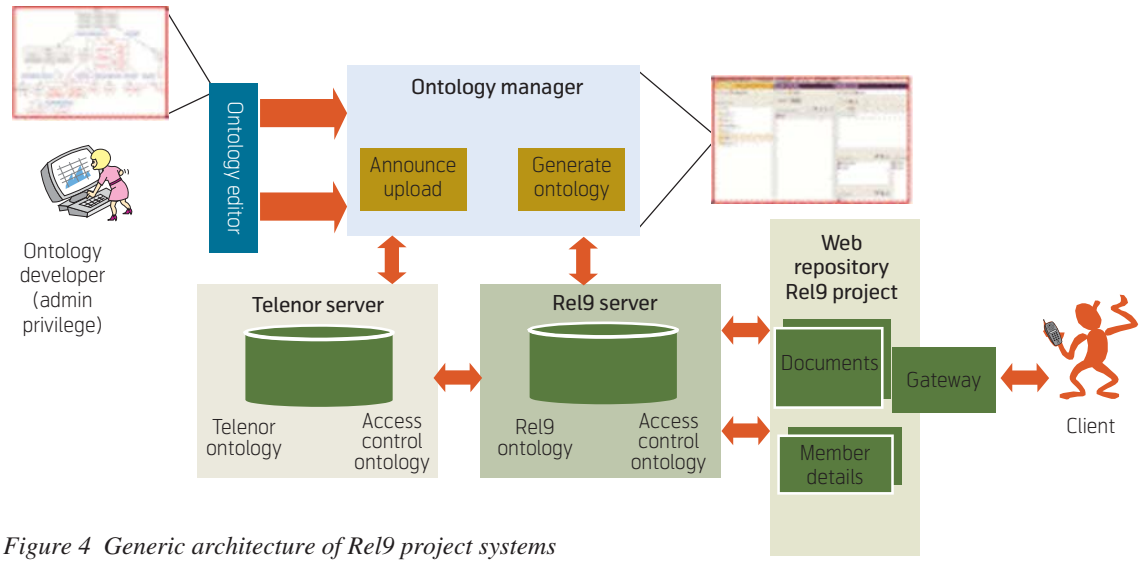


Figure 4 Generic architecture of Rel9 project systems

ISSN 0085-7130©Telenor ASA 2007

5.2 Generic Architecture of the System

Users are authenticated by the identity providers of corporate organisations using keys. These keys are assumed distributed only among the members of the company/project using the mobile environment or near-field communications. They then access the company/project contents using the proposed role-based ontology with differential access rights. The generic architecture of this paper is illustrated in Figure 4. The project membership and access rights management to the project contents is handled through a project ontology. While the project ontology handles the project members' access to content, the company ontology will allow the supervisors of project members to do so. Our service scenario builds on the relationship between the actors and establishes access rights to content.

5.3 Role-based Identity Management

A business project group is composed of several different types of project members. They can be categorized based on the roles they play in the project. Each of the roles has different privileges or rights to access different project resources. Table 1 shows examples of such scenarios. Access control to project resources and maintaining privacy of project member information are the objectives of the proposed ontology. 'He is leader of the Rel9 project' refers to his corporate identity (CID) in professional life and his role in the Rel9 project. The project leader has administrative and final approval privileges which the project members do not have. One of the crucial requirements of privacy is to ensure that a visitor should not be allowed to see the project member details (for example; contact address, email, phone number, etc.). Project leader and members have visibility of member details. Project members (including project leader) have their own supervisors in parent organizations. Supervisors may want to know about project status, see (or even write) project documents, deliverables and member details. The role as supervisors ensures these features.

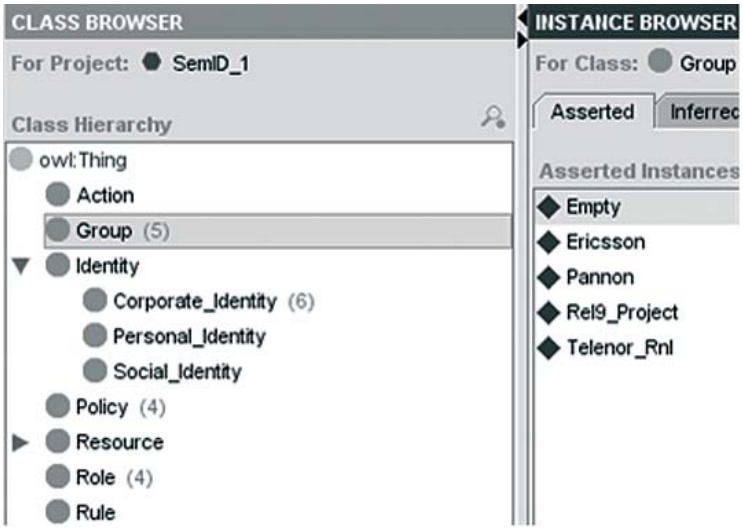


Figure 5 Classes and subclasses of ontology

Figure 1 of Section 2 illustrated our use case scenario. The Semantic Identity Management (SemID.org) ontology has been developed based on this scenario.

5.3.1 Policy and Rule

The corporate identity of each project member, the project group to which he belongs and the role he plays are defined in the ontology. Each role has certain policy (or policies). A policy (P) represents the privilege reserved for each role in a community and expressed through a set of rules (R_1, R_2, ..., R_n). Therefore a policy can be presented as

$$P = \{R_1, R_2, ..., R_n\}.$$

A rule is a function that takes an access request as input and results in an action (permit, deny or not-applicable). A rule is composed of the triple Subject (S), Resource (R) and Action (A) that must be met for a rule to apply to a given request. In the proposed SemID ontology, Subjects are the identities that play specific Roles (which is predefined in the ontology) like project leader, supervisor, project member and visitor. Resources are the project resources like deliverables, documents, etc. So, the rule is simplified as

$$R = \{S, R, A\}.$$

If Josef Noll is the project leader and he wants to write over a project deliverable, the corresponding rule will be defined as

$$R = \{ProjectLeader, Deliverables, Submit\}.$$

For the same purpose, the corresponding rule of the visitor Geir Egeland will be defined as

$$R = \{Visitor, Deliverables, Deny\}.$$

These example rules belong to the policy: *write*. However, these access control rules have not been explicitly defined in the modelled ontology which will be implemented in later works. It is assumed that relevant subjects (individuals defined as CIDs in the ontology) are going to be authenticated to their company systems through secure means. In our case it

9) *Protégé, http://protege.stanford.edu/*

ISSN 0085-7130©Telenor ASA 2007

means that Geir Egeland is authenticated as Telenor R&I member, which through Telenor's participation in Rel9 gives him guest status in the Rel9 project.

In conclusion, access control to project resources is maintained through policies and rules.

5.3.2 Ontology of the System

We model the ontology of the use case scenario with OWL-DL using the Protégé ontology editor platform[9]. In the left part of Figure 5 classes and subclasses of the SemID ontology are presented, modelling the proposed use case and meeting the requirements. Figure 5 also illustrates the instances of four different groups described in the use case scenario. An *empty* group has been created to support privacy of a group when visitors try to access resources.

In this ontology, the corporate identity (CID) of each representative is defined by corresponding names. These are the instances of Identity subclass: CID. Anyone whose Identity instance is not defined explicitly in SemID will be considered as 'Visitor'. Each instance has four properties: *hasGroup*, *hasVisibility*, *hasRole* and *hasSupervisor*. The group to which a member belongs is explicitly identified using *hasGroup* property. *hasVisibility* points to the groups a member needs general purpose visibility to. The Role a person plays in a project is identified by *hasRole*. *hasSupervisor* explicitly defines 'who is supervisor of whom'. Example source codes (RDF/XML) of corporate identities and their properties are as follows:

```
<Corporate_Identity rdf:ID=
"Erik_Swansson">
  <hasGroup rdf:resource="#Ericsson"/>
  <hasGroup rdf:resource="#Rel9_
  Project"/>
  <hasVisibility rdf:resource=
  "#Ericsson"/>
  <hasVisibility rdf:resource=
  "#Rel9_Project"/>
  <hasRole rdf:resource=
  "#Project_Member"/>
  <hasSupervisor
rdf:resource="#Peter_Johansson"/>
</Corporate_Identity>
```

There are four possible roles in the Rel9 project. Each Role has specific policy/policies. These are expressed by *hasPolicy* property. The following source code illustrates four different policies of these roles: Administrator, FinalApproval, Read and ReadWrite. Administrator policy is introduced to represent the administrative privilege of the project leader and similarly, FinalApproval policy refers to the final approval of project deliverables.

Project roles	Privileges	Project resources
Project leader	Administrator	Membership details
	Final Approval	Deliverables
	Read/Write	Documents
	Visibility	Member details
Supervisors	Read/Write	Deliverables
		Documents
	Visibility	Member details
Members	Read/Write	Documents
	Visibility	Member details
Visitors	Read only	Documents
	No visibility	Member details

Table 1 Roles in a project and their privileges to access resources

```
<Policy rdf:ID="Administrator"/>
  <Policy rdf:ID="FinalApproval"/>
  <Policy rdf:ID="Read"/>
  <Policy rdf:ID="ReadWrite"/>
```

Four different roles of the project are represented by four instances of Role: Project leader, Supervisor, Project member, and Visitor. Appropriate policies are added to each instance of Role. According to the use case scenario, the project leader has the policies: *Administrator*, *FinalApproval* and *ReadWrite*. 'Visibility' privilege deals with the privacy of the project and is satisfied through two properties: *hasVisibility* and *hasVisibilityOfGroup*. In order to fulfill the requirements of group member's privacy, a *hasVisibilityOfGroup* property has been created. The Visitor instance has visibility of group called '*empty*' (an instance of class: Group) to ensure that 'as a visitor one should be allowed to read the documents of the project, but he does not have the permission to see the member details of the visited project'. Example codes to represent roles and corresponding properties in the ontology are as follows:

```
<Role rdf:ID="Project_Leader">
  <hasVisibilityOfGroup
rdf:resource="#Rel9_Project"/>
  <hasPolicy rdf:resource="#Administrator"/>
  <hasPolicy rdf:resource="#Final Approval"/>
  <hasPolicy rdf:resource="#ReadWrite"/>
</Role>
```

In this ontology, we have defined ten properties. Each property has its domain and range. The classes to which a property is attached are called domain. Allowed classes for properties are often called a range of a property. Sample source codes of the implementation of these properties and corresponding domain and range are given as follows:

ISSN 0085-7130

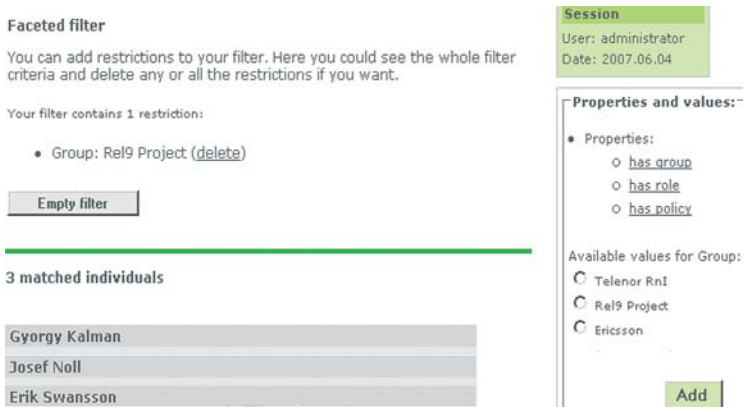


Figure 6 Screen shot of SEDO for users with administrative rights

```
<owl:ObjectProperty rdf:ID="hasAction">
  <rdfs:domain rdf:resource="#Rule"/>
  <rdfs:range rdf:resource="#Action"/>
</owl:ObjectProperty>
  ……..
<owl:ObjectProperty rdf:ID="hasGroup">
  <rdfs:domain rdf:resource="#Identity"/>
  <rdfs:domain rdf:resource="#Group"/>
</owl:ObjectProperty>
```

hasVisibility and *hasVisibilityOfGroup* ensure the privacy of groups. *hasSubject*, *hasResource* and *hasAction* create the simplified rule.

5.3.3 Privacy Enhancement

Privacy requirements are satisfied in the SemID ontology using two properties (*hasVisibility* and *hasVisibilityOfGroup*). *hasVisibilityOfGroup* is attached to class: Role and *hasVisibility* are attached to class: Identity (subclass: CID). The latter is rather a general visibility property which ensures that anyone belonging to at least one group has visibility of resources of those groups. Role based visibility (*hasVisibilityOfGroup*) represents the visibility of specific resources like the project member details. Leader, members and supervisors of the Rel9 project have visibility of project members' details which the visitors cannot see. This privacy feature is introduced to protect member details, such as email and phone number to visitors. We introduced the role of supervisor in order to satisfy the information requirements of participating companies, namely Telenor and Ericsson. Supervisors have access to member details, ensured through SemID.

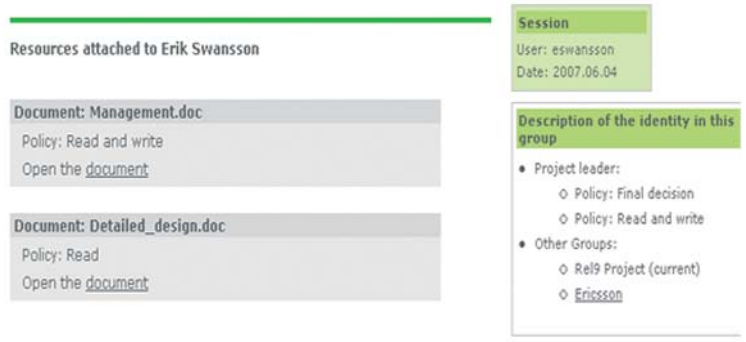


Figure 7 Screen shot of the SEDO software when the user is logged-in as normal project member

6 Semantic-based Enterprise Content Management

The ontology developed above was integrated by Universidad Carlos III de Madrid in SEDO, a Semantics-based Enterprise Content Management System [22]. In a nutshell, an Enterprise Content Management System (ECMS) is a software application used to manage computer files, media, audio files, electronic documents and web content inside the boundaries of a company, specifying different levels of access for those business resources. The idea behind the ECMS is to make these files available both within the company as well as over the web. Figures 6 and 7 show screen shots of the implemented software.

SEDO is implemented by means of Ruby on Rails (RoR)[10]. The SemID ontology has been used as a conceptual backbone for the permissions and access levels of the different users of the system. There are a number of things that can be achieved by the SEDO system thanks to the conceptual backbone of the SemID ontology and they are summarised as follows:

- *Creating and annotating resources and SEDO users*

Semantic descriptions are added to the resources, together with a number of rules and policies. The description is formalised through the SemID ontology.

- *Navigating and Searching through semantics*

Filtering the information depending on the properties of the ontology is what is called "Faceted Search and Browsing" [23]. In a nutshell, facets are orthogonal conceptual dimensions of the data, and SEDO allows us to see, for example, which other users are eligible to access a particular resource like "Document 1" of the Rel9 project.

7 Conclusion

This paper addresses the complexity of content handling in projects involving multiple organisations. Project content is typically stored within one company system, making it difficult to share the content across company borders.

[10] *Ruby on Rails, http://www.rubyonrails.org*

ISSN 0085-7130©Telenor ASA 2007

We introduced semantic technologies to describe in a formal way the roles and corresponding access policies. Roles and access rights for project members and their superiors in the parent companies are formalised in the Semantic Identity (SemID.org) ontology.

The prototypical implementation in a semantic-based enterprise content management system demonstrates the capabilities of the approach. An identity management based on roles represents a flexible, efficient and secure way to ensure that only relevant content is dissolved.

References

1 Chewar, C M, McCrickard, D S, Carroll, J M. *Persistent virtual identity in community networks: Impact to social capital value chains*. Virginia Tech, Computer Science, 2003. (Technical Report TR-03-01)

2 Walters, G J. Privacy and Security: An Ethical Analysis. *Computers and Society*, 2001, 8-23.

3 Kruk, S R, Grzonkowski, S, Gzella, A, Woroniecki, T, Choi, H-C. D-FOAF: Distributed Identity Management with Access Roghts Delegation. *1st Asian Semantic Web Conference*, Beijing, China, 2006.

4 Kruk, S R, Gzella, A, Grzonkowski, S. *D-FOAF Distributed Identity Management based on Social Networks*. In demo session of ESWC 2006.

5 Choi, H-C et al. Trust Models for Community-Aware Identity Management. *Identity, Reference and the Web IRW2006. WWW2006 Workshop*, Scotland, May 23, 2006.

6 Finin, T, Joshi, A. Agents, Trust, and Information Access on the Semantic Web. *ACM SIGMOD*, 31 (4), 30-35, 2002. (Special Issue: Special section on semantic web and data management)

7 Smith, M A et al. Mother, May I? OWL-based Policy Management at NASA. *European Semantic Web Conference 2007, ESWC2007*.

8 Kolovski, V, Hendler, J, Parsia, B. Analyzing Web Access Control Policies. *16th International World Wide Web Conference, WWW2007*, Alberta, Canada, May 8-12, 2007.

9 Kellerer, W et al. *Systems beyond 3G – Operators' vision*. Eurescom Project P1203, December 2002.

10 Kagal, L et al. Authorization and Privacy for Semantic Web Services. *IEEE Int. Systems*, 19 (4), 50-56, 2004.

11 McIlraith, S A, Cao Son, T, Zeng, H. Semantic Web Services. *IEEE Int. Systems*, 16 (2), 46-53, 2001.

12 Danet, P Y. ePerSpace: A European Project for the Seamless and Personalised Digital Communicating Home of the Future. *European VPN Services Forum Conference*, London, June 15-17, 2006.

13 Khoshaflau, S N, Copeland, G P. Object Identity. *Proceedings OOPSLA'86*, Sept. 1986, 406-416.

14 MacKenzie, C M et al. OASIS, Reference Model for Service Oriented Architectures 1.0. August 2, 2006. January 10, 2007 [online] – URL: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev= soa-rm

15 Noll, J, Lillevold, E. Roadmap to ASG based Semantic Web Services. In: *Proc. of The International Conference on Internet & Web Applications and Services 2006, ICIW*, February 23-25, 2006.

16 Noll, J et al. Estimating business profitability of Semantic Web Services for Mobile Users. In: Schaffert, S, Sure, Y. *Semantic Systems, From Visions to Applications, Proc. of the Semantics 2006*. Österreichische Computer Gesellschaft, 195-204.

17 Cellular-news. *MasterCard Tests NFC Payments with Nokia Handsets*. December 10, 2006 [online] – URL: http://www.cellular-news.com/story/20211.php

18 Hardt, D. Identity 2.0. *OSCON 2005*. November 2, 2007 [online] – URL: http://www.identity20.com/media/OSCON2005/

19 Roccas, S, Brewer, M B. Social Identity Complexity. *Personality and Social Psychology Review*, 6 (2), 88-106, 2002.

20 *The Laws of Identity*. October 11, 2007 [online] – URL: http://www.identityblog.com/stories/2004/12/09/thelaws.html

21 Chowdhury, M M R, Noll, J. Distributed Identity for Secure Service Interaction. P*roc. Third Intern. Conference on Wireless and Mobile Communica-*

ISSN 0085-7130©Telenor ASA 2007

tions, ICWMC07, Gouadeloupe, French Caribbean, March 4-9, 2007

22 Chowdhury, M M R, Gomez, J M, Noll, J, Crespo, A G. SemID: Combining Semantics with Identity Management. *Intern. Conf. on Emerging Security Information, Systems and Technologies, SECURWARE 2007*, Valencia, Spain, October 2007.

23 Oren, E et al. (2007, May). Activerdf: Object-oriented semantic web programming. In: *WWW '07: Proceedings of the 16th international conference on World Wide Web*, Banff, Alberta, Canada, May 2007, 817-824. New York, NY, ACM Press.

Mohammad M.R. Chowdhury is a PhD student at the University Graduate Center at Kjeller (UniK), Norway, in the area of User Mobility and Service Continuity. He received his MSc from Helsinki University of Technology in Radio Communication. Before joining his current position, he was radio planning engineer at GrameenPhone, a Telenor subsidiary in Bangladesh.
His current areas of interest are identity, identity representations and identity based service interactions, seamless user experience in heterogeneous wireless networks and development of innovative service concepts for mobile.

email: mohammad@unik.no

Josef Noll holds a professor stipend from the University of Oslo in the area of Mobile Services. Working areas include Mobile Authentication, Wireless Broadband Access, Personalised Services, Mobile-Fixed Integration and the Evolution to 4G systems. He is also Senior Adviser in Movation, Norway's leading innovation company for mobile services.
Previously, he was Senior Adviser in Telenor R&I in the Products and Markets group, and was use-case leader in the EU FP6 'Adaptive Services Grid (ASG)' projects, and has initiated i.a. the EU's 6th FP ePerSpace and several Eurescom projects.

email: josef@unik.no

ISSN 0085-7130©Telenor ASA 2007

Terms and Acronyms in Identity Management

Acronym /term	Definition	Explanation	Web-resources
2G	Second Generation mobile technology	Refers to the family of digital cellular telephone systems standardised in the 1980s and introduced in the 1990s. They introduced digital technology and carry both voice and data conversation. CDMA, TDMA and GSM are examples of 2G mobile networks.	
3G	Third Generation mobile technology	The generic term for the next generation of wireless mobile communications networks supporting enhanced services like multimedia and video. Most commonly, 3G networks are discussed as graceful enhancements of 2G cellular standards, like e.g. GSM. The enhancements include larger bandwidth, more sophisticated compression techniques, and the inclusion of inbuilding systems. 3G networks will carry data at 144 kb/s, or up to 2 Mb/s from fixed locations. 3G comprises mutually incompatible standards: UMTS FDD and TDD, CDMA2000, TD-CDMA.	
3GPP	Third Generation Partnership Project	Group of the standards bodies ARIB and TTC (Japan), CCSA (People's Republic of China), ETSI (Europe), T1 (USA) and TTA (Korea). Established in 1999 with the aim to produce and maintain the specifications for a third generation mobile communications system called UMTS. Note that 3GPP itself is not a standardization organization and that all produced standards must be ratified by a standardization organizations. A permanent project support group called the "Mobile Competence Centre (MCC)" is in charge of the day-to-day running of 3GPP. The MCC is based at the ETSI headquarters in Sophia Antipolis, France.	http://www.3gpp.org
AAA	Authentication, Authorization and Accounting	Key functions to intelligently controlling access, enforcing policies, auditing usage, and providing the information necessary to do billing for services available on the Internet. The term AAA is used to denote an internet security service architecture that provides the AAA services. The architecture includes AAA servers and AAA protocols. The AAA protocols include RADIUS and DIAMETER. Defined in IETF RFC 2903.	http://www.ietf.org, http://tools.ietf.org /html/rfc4303
AAA server, EAP server, or backend authentication server		These three terms are used interchangeably in this document. AAA stands for Authentication, Authorization, and Accounting. A backend authentication server is an entity that provides an authentication service to an authenticator. RADIUS is an AAA server.	
A3	Algorithm 3	Authentication algorithm; used for authenticating the subscriber in GSM.	
A5	Algorithm 5	Cipher algorithm; used for enciphering/deciphering data in GSM.	
A8	Algorithm 8	Cipher key generator; used to generate the cryptographic key Kc in GSM.	
Access manager		Sun Java System Access Manager delivers open, standards-based access control across intranets and extranets. It is a security foundation that helps organizations to manage secure access to enterprises' Web applications both within the enterprise and across business-to-business (B2B) value chains. It provides open, standards-based authentication and policy-based authorization with a single, unified framework. It secures the delivery of essential identity and application information to meet today's needs and to scale with growing business needs by offering single sign-on (SSO) as well as enabling federation across trusted networks of partners, suppliers, and customers.	
AuC	Authentication Centre	The AuC is the authentication centre in 2G and 3G cellular networks. The AuC is co-located with a HLR. It is the network element that provides the authentication triplets for authenticating the subscriber.	
Authenticator		The component that initiates the EAP authentication. In this document the authenticator is running in IDP.	
BankID		A PKI concept for Norwegian banks.	
BSI	British Standards Institute	BSI Group, also known in its home market as the British Standards Institution (or BSI) is a multinational business services provider whose principal activity is the production of standards and the supply of standards-related services.	http://www.bsi-global .com/
CA	Certification Authority	In cryptography, a certificate authority or certification authority is an entity which issues digital certificates for use by other parties. It is an example of a trusted third party.	

ISSN 0085-7130

Acronym /term	Definition	Explanation	Web-resources
CO	Content Object (in DRM)		
CoT	Circle of Trust	A federation of service providers and identity providers that have business relationships based on Liberty architecture and operational agreements and with whom users can transact business in a secure and apparently seamless environment. Also known as a Trust Circle.	
CRL	Certificate Revocation List	In the operation of some cryptosystems, usually public key infrastructures (PKIs), a certificate revocation list (CRL) is a list of certificates (more accurately: their serial numbers) which have been revoked, are no longer valid, and should not be relied on by any system user. There are different revocation reasons defined in RFC 3280.	http://www.ietf.org, http://tools.ietf.org /html/rfc3280
DB	Data base	A collection of data structured and organized in a disciplined fashion so that access is possible quickly to information of interest.	
DRM	Digital Rights Management	Any of several technologies used by publishers (or copyright owners) to control access to and usage of digital data (such as software, music, movies) and hardware, handling usage restrictions associated with a specific instance of a digital work.	
EAP	Extensible Authentication Protocol	An authentication framework that enables clients to authenticate with a central server. EAP can be used with several authentication mechanisms (EAP methods), such as: EAP-AKA, EAP-SIM, EAP-MD-5, etc. Defined in IETF RFC 3748.	http://tools.ietf.org /html/rfc3748
EAP-AKA	Extensible Authentication Protocol – Authentication and Key Agreement	An extension to the EAP proposed by the IETF enabling authentication and session key distribution using the UMTS AKA mechanism. UMTS AKA is based upon symmetric keys and runs typically on a USIM (UMTS Subscriber Identity Module). EAP/AKA Authentication includes optional user anonymity and re-authentication procedures. EAP AKA is defined in IETF RFC 4187.	http://www.ietf.org, http://tools.ietf.org /html/rfc4187
EAP-SIM	Extensible Authentication Protocol – Subscriber Identity Module	An extension of the EAP using the GSM SIM. EAP-SIM is defined in RFC 4186.	http://www.ietf.org, http://tools.ietf.org /html/rfc4186
EMV	Europay, MasterCard and VISA	EMV is a standard for interoperation of IC cards ("Chip cards") and IC capable POS terminals and ATMs for authenticating credit and debit card payments. The name EMV comes from the initial letters of Europay, MasterCard and VISA, the three companies which originally co-operated to develop the standard. Europay International SA was absorbed into MasterCard in 2002. JCB (formerly Japan Credit Bureau) joined the organisation in December 2004. IC card systems based on EMV are being phased in across the world, under names such as "IC Credit" and "Chip and PIN". The EMV standard defines the interaction at the physical, electrical, data and application levels between IC cards and IC card processing devices for financial transactions. Portions of the standard are heavily based on the IC Chip card interface defined in ISO 7816. EMVCo is the organisation responsible for developing and maintaining the EMV standard.	http://www.emvco.com/
ETSI	European Telecommunication Standards Institute	A non-profit membership organization founded in 1988. The aim is to produce telecommunications standards to be used throughout Europe. The efforts are coordinated with the ITU. Membership is open to any European organization proving an interest in promoting European standards. It was e.g. responsible for the making of the GSM standard. The headquarters are situated in Sophia Antipolis, France.	http://www.etsi.org
ETSI SCP	ETSI Smart Card Platform		
GPD	GlobalPlatform Device		
GPS	Global Positioning System	The Global Positioning System (GPS) is a worldwide radio-navigation system formed from a constellation of 24 satellites and their ground stations. GPS uses these "man-made stars" as reference points to calculate positions accurate to a matter of metres.	http://www.gps.gov/, http://www.navcen.uscg .gov/gps/default.htm

ISSN 0085-7130©Telenor ASA 2007

Acronym /term	Definition	Explanation	Web-resources
GSM	Global System for Mobile communications	A digital cellular phone technology system that is the predominant system in Europe, but is also used around the world. Development started in 1982 by CEPT and was transferred to the new organisation ETSI in 1988. Originally, the acronym was the group in charge, "Group Special Mobile" but later the group changed its name to SMG. GSM was first deployed in seven countries in Europe in 1992. It operates in the 900 MHz and 1.8 GHz band in Europe and 1.9 GHz band in North America. GSM defines the entire cellular system, from the air interface to the network nodes and protocols. As of October 2006, there were more than 2.1 billion GSM users in more than 200 countries worldwide. The ubiquity of the GSM standard makes international roaming very common between mobile phone operators enabling phone users to access their services in many other parts of the world as well as their own country. GSM differs significantly from its predecessors in that both signalling and speech channels are digital, which means that it is seen as a second generation (2G) mobile phone system. This fact has also meant that data communication was built into the system from very early on. GSM is an open standard which is currently developed by the 3GPP.	http://www.gsmworld.com/, http://www.etsi.org, http://www.3gpp.org
GSMA	GSM Association	The world's leading wireless industry representative body, consisting of more than 660 second and third-generation wireless network operators and key manufacturers and suppliers to the wireless industry.	http://www.gsmworld.com/
HLR	Home Location Register	The Home Location Register or HLR is a central database that contains details of each mobile phone subscriber that is authorized to use the GSM core network. More precisely, the HLR stores details of every SIM card issued by the mobile phone operator. Each SIM has a unique identifier called an IMSI which is one of the primary keys to each HLR record. The next important items of data associated with the SIM are the telephone numbers used to make and receive calls to the mobile phone, known as MSISDNs. The main MSISDN is the number used for making and receiving voice calls and SMS, but it is possible for a SIM to have other secondary MSISDNs associated with it for fax and data calls. Each MSISDN is also a primary key to the HLR record.	http://www.etsi.org
HSS	Home Subscriber Server	The home subscriber server contains all operative subscriber data, including information on subscribed services, location/roaming information and security credentials. Includes HLR/AuC and AAA services.	http://www.3gpp.org
HTTPS	Hyper Text Transfer Protocol Secure sockets	A Uniform Resource Identifier (URI) scheme used to indicate a secure HTTP connection. It is syntactically identical to the http:// scheme normally used for accessing resources using HTTP. Using an https: URL indicates that HTTP is to be used, but with a different default TCP port (443) and an additional encryption/authentication layer between the HTTP and TCP. This system was designed by Netscape Communications Corporation to provide authentication and encrypted communication and is widely used on the World Wide Web for security-sensitive communication such as payment transactions and corporate logons.	
IANA	Internet Assigned Numbers Authority	IANA (Internet Assigned Numbers Authority) is the organization under the Internet Architecture Board (IAB) of the Internet Society that, under a contract from the US government has overseen the allocation of Internet Protocol addresses to Internet service providers (ISPs). IANA has also been responsible for the registry for any "unique parameters and protocol values" for Internet operation. These include port numbers, character sets, and MIME media access types. Partly because the Internet is now a global network, the US government has withdrawn its oversight of the Internet, previously contracted out to IANA, and lent its support to a newly-formed organization with global, non-government representation, the Internet Corporation for Assigned Names and Numbers (ICANN). ICANN has now assumed responsibility for the tasks formerly performed by IANA.	http://www.iana.org/
ICCID	Integrated Circuit Card Identity	A SIM Serial Number, which is normally printed on the SIM-cards used in GSM and 3G phones. The numbering of the card is based on International Standard ISO/IEC 7812. The maximum length of the visible card number is 19 or 20 characters (see remark below) and is composed of the following subparts: Issuer Identification number (max. 7 digits): Major Industry Identifier (MII), 2 digits, 89 for telecommunication purposes, country code, 1-3 digits, as defined by ITU-T recommendation E.164, issuer identifier, variable. Individual account identification: individual account identification number, parity check digit.	
ID	Identity		
IdM	Identity Management		

ISSN 0085-7130

Acronym /term	Definition	Explanation	Web-resources
IdP	Identity Provider	According to Liberty Alliance specifications an Identity Provider creates and manages the identity of the users and authenticates them to the service providers. It manages identity information on behalf of Principals and provides assertions of Principal authentication to other providers.	
IEEE	The Institute of Electrical and Electronics Engineers	USA-based organisation open to engineers and researchers in the fields of electricity, electronics, computer science and telecommunications. Established in 1884. The aim is to promote research through journals and conferences and to produce standards in telecommunications and computer science. IEEE has produced more than 900 active standards and has more than 700 standards under development. Divided into different branches, or 'Societies'. Has daughter organisations, or 'chapters' in more than 175 countries worldwide. Headquarters in Piscataway, New Jersey, USA.	http://www.ieee.org
IETF	Internet Engineering Task Force	A large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. It is open to any interested individual. The technical work of the IETF is done in its working groups, which are organized by topic into several areas (e.g. routing, transport, security, etc.). Much of the work is handled via mailing lists. The IETF holds meetings three times per year. The IETF working groups are grouped into areas and managed by Area Directors (AD). The ADs are members of the Internet Engineering Steering Group (IESG). Providing architectural oversight is the Internet Architecture Board (IAB). The IAB also adjudicates appeals when someone complains that the IESG has failed. The IAB and IESG are chartered by the Internet Society (ISOC) for these purposes. The General Area Director also serves as the chair of the IESG and of the IETF, and is an ex-officio member of the IAB. The Internet Assigned Numbers Authority (IANA) is the central coordinator for the assignment of unique parameter values for Internet protocols. The IANA is chartered by the Internet Society (ISOC) to act as the clearinghouse to assign and coordinate the use of numerous Internet protocol parameters. IETF's mission statement is given in IETF RFC 3935.	http://www.ietf.org, http://tools.ietf.org /html/rfc3935
IMEI	International Mobile Equipment Identity	A number unique to every GSM and UMTS mobile phone. It is usually found printed on the phone underneath the battery and can also be found by dialling the sequence *#06# into the phone. The IMEI number is used by the GSM network to identify valid devices and therefore can be used to stop a stolen phone from accessing the network. For example, if a mobile phone is stolen, the owner can call his or her network provider and instruct them to "ban" the phone using its IMEI number. This renders the phone useless, regardless of whether the phone's SIM is changed. Unlike the Electronic Serial Number or MEID of CDMA and other wireless networks, the IMEI is only used to identify the device, and has no permanent or semi-permanent relation to the subscriber.	
IMPI	IP Multimedia Private Identity	Identifier required by the IP Multimedia Subsystem (IMS). It is a Uniform Resource Identifier (URI), which can be digits (a tel-uri, like tel:+1-555-123-4567) or an alphanumeric identifier (a sip-uri, like sip:john.doe@example.com).	http://www.3gpp.org, http://www.ietf.org, http://www.imsforum.org
IMPU	IP Multimedia Public Identity	Identifier required by the IP Multimedia Subsystem (IMS). It is a Uniform Resource Identifier (URI), that can be digits (a tel-uri, like tel:+1-555-123-4567) or an alphanumeric identifier (a sip-uri, like sip:john.doe@example.com). There can be multiple IMPU per IMPI (often a tel-uri and a sip-uri). The IMPU can also be shared with another phone, so both can be reached with the same identity (for example, a single phone number for an entire family).	http://www.3gpp.org, http://www.ietf.org, http://www.imsforum.org
IMS	IP Multimedia Subsystem	A standardised Next Generation Networking (NGN) architecture for telecom operators that want to provide mobile and fixed multimedia services. It uses a Voice-over-IP (VoIP) implementation based on a 3GPP standardised implementation of SIP, and runs over the standard Internet Protocol (IP). Existing phone systems (both packet-switched and circuit-switched) are supported. IMS was originally defined by an industry forum called 3G.IP (www.3gip.org) formed in 1999. 3G.IP developed the initial IMS architecture, which was brought to 3GPP for industry standardization as part of their standardization work for 3G mobile phone systems in UMTS networks. It first appeared in release 5 (evolution from 2G to 3G networks), when SIP-based multimedia was added. Support for the older GSM and GPRS networks was also provided. "Early IMS" was defined to allow for IMS implementations that do not yet support all "Full IMS" requirements. 3GPP2 (a different organisation) based their CDMA2000 Multimedia Domain (MMD) on 3GPP IMS, adding support for CDMA2000.	http://www.3gpp.org, http://www.ietf.org, http://www.imsforum.org

ISSN 0085-7130©Telenor ASA 2007

Acronym /term	Definition	Explanation	Web-resources
IMSI	International Mobile Subscriber Identity	The principal subscriber identity in 2G/3G systems. Structure and definition of IMSI is given both in ITU-T recommendations (E.212) and in 3GPP specifications (TS 23.003). Note that in ITU-T E.212 the acronym is defined as "International Mobile Station Identity", but the structure is otherwise identical.	http://www.itu.int, http://www.3gpp.org /ftp/Specs/html-info /23003.htm
IP	Internet Protocol	A protocol for communication between computers, used as a standard for transmitting data over networks and as the basis for standard Internet protocols. Originally defined in IETF RFC 791.	http://www.ietf.org, http://tools.ietf.org /html/rfc791
IPv4	Internet Protocol v4	IPv4 is version 4 of the Internet Protocol (IP) and it is the first version of the Internet Protocol to be widely deployed. IPv4 is the dominant network layer protocol on the internet. It is described in IETF RFC 791 (September 1981) which obsoleted RFC 760 (January 1980). IPv4 is a data-oriented protocol to be used on a packet switched internetwork (e.g. Ethernet). It is a best effort protocol in that it does not guarantee delivery. It does not make any guarantees on the correctness of the data; it may result in duplicated packets and/or packets out-of-order. All of these things are addressed by an upper layer protocol (e.g. TCP, UDP). See also IP and Ipv6.	http://www.ietf.org
ISDN	Integrated Services Digital Network	A digital telecommunications network that provides end-to-end digital connectivity to support a wide range of services, including voice and non-voice services, to which users have access by a limited set of standard multi-purpose user-network interfaces. The user is offered one or more 64 kb/s channels.	http://www.itu.int
ISIM	IP Multimedia Services Identity Module	An application running on a UICC smart card in a 3G mobile telephone in the IP Multimedia Subsystem (IMS). It contains parameters for identifying and authenticating the user to the IMS. The ISIM application can co-exist with SIM and USIM on the same UICC making it possible to use the same smartcard in both GSM networks and earlier releases of UMTS.	http://www.3gpp.org /ftp/Specs/html-info /31103.htm
ITU-T	International Telecommunication Union – Standardization Sector	A sector of the ITU whose mission is to ensure an efficient and on-time production of standards (Recommendations) covering all fields of telecommunications. It was created on 1 March 1993, replacing the former International Telegraph and Telephone Consultative Committee (CCITT).	http://www.itu.int/ITU-T/
Kc		Cryptographic key; used by the cipher A5.	
Ki		Subscriber authentication key; the cryptographic key used by the authentication algorithm, A3, and cipher key generator, A8.	
LAN	Local Area Network	A network shared by communicating devices, usually on a small geographical area. A system that links together electronic office equipment, such as computers and word processors, and forms a network within an office or building.	
LDAP	Lightweight Directory Access Protocol	A networking protocol for querying and modifying directory services running over TCP/IP. Its current version is LDAPv3, as defined in RFC 3377.	http://www.ietf.org
MAC	Medium Access Control	The lower of the two sub layers of the Data Link Layer. In general terms, MAC handles access to a shared medium, and can be found within many different technologies. For example, MAC methodologies are employed within Ethernet, GPRS, and UMTS.	
MAC	Message Authentication Code	A MAC function computes a cryptographic signed integrity checksum over an arbitrary length input string under the control of a secret key. MAC functions are quite similar to hash functions, but the MAC function output can only be computed with knowledge of the secret key. MAC functions can be used to provide the message origin authentication and data integrity security services.	http://en.wikipedia.org /wiki/Message _authentication_code
MAP	Mobile Application Part	A protocol that enables real time communication between nodes in a mobile cellular network. A typical usage of the MAP protocol would be for the transfer of location information from the VLR (Visitor Location Register) to the HLR (Home Location Register). Defined in 3GPP TS 09.02 for GSM and in 3GPP TS 29.002 for UMTS.	http://www.3gpp.org /ftp/Specs/html-info /0902.htm, http://www.3gpp.org /ftp/Specs/html-info /29002.htm
MNO	Mobile Network Operator		
MSISDN	Mobile Station Integrated Services Digital Network	MSISDN refers to the 15-digit number that is used to refer to a particular mobile station. It is the mobile equivalent of ISDN. The ITU-T recommendation E.164 defines the international numbering plan that MSISDN is based on.	http://www.itu.int
NAI	Network Access Identifier	In computer networking, a standard way of identifying users who request access to a network. The standard syntax is "user@realm". NAIs were originally defined in RFC 2486, which has been superseded by RFC 4282. The latter RFC is the current standard for the NAI. NAIs are commonly found as user identifiers in the RADIUS and DIAMETER network access protocols and the EAP authentication protocol.	http://tools.ietf.org /html/rfc4282

ISSN 0085-7130

Acronym /term	Definition	Explanation	Web-resources
NFC	Near Field Communication Technology	NFC, jointly developed by Sony and Philips was approved as an ISO/IEC standard on December 8, 2003. It was approved as an ECMA standard earlier on. On March 18, 2004, Nokia, Sony and Philips formed NFC-forum to advance NFC development. NFC is essentially about data sharing between devices using short-range radio technologies. Near Field Communication Technology holds the promise of bringing true mobility to consumer electronics in an intuitive and psychologically comfortable way since the devices can hand-shake only when brought literally into touching distance.	http://www.nfc-forum.org/home
NIST	National Institute of Standards and Technology	From 1901 to 1988 known as the National Bureau of Standards (NBS), a non-regulatory agency of the United States Department of Commerce. The institute's mission is to promote US innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve quality of life. NIST's headquarters are in Gaithersburg, Maryland. It also has laboratories in Boulder, Colorado.	http://www.nist.gov/
OCSP	Online Certificate Status Protocol	An Internet protocol used for obtaining the revocation status of an X.509 digital certificate. It is described in RFC 2560 and is on the Internet standards track.	http://tools.ietf.org/html/rfc2560
OID	Object Identifier	In computing, an OID is an identifier used to name an object (compare URN). Structurally, an OID consists of a node in a hierarchically-assigned namespace, formally defined using the ITU-T's ASN.1 standard. Successive numbers of the nodes, starting at the root of the tree, identify each node in the tree. Designers set up new nodes by registering them under the node's registration authority. In computer security, OIDs serve to name almost every object type in X.509 certificates.	
OMA	Open Mobile Alliance	A standards body which develops open standards for the mobile industry. The OMA was created in June 2002 as an answer to the proliferation of industry forums each dealing with a few application protocols: the WAP Forum (focused on browsing and device provisioning protocols), the Wireless Village (focused on instant messaging and presence), the SyncML Consortium (focused on data synchronization), the Location Interoperability Forum, the Mobile Games Interoperability Forum and the Mobile Wireless Internet Forum. Each of these forums had its bylaws, its decision-taking procedures, its release schedules, and in some instances there was some overlap in the specifications, causing duplication of work. The OMA was created to gather these initiatives under a single umbrella. Members include traditional wireless industry players such as equipment and mobile systems manufacturers and mobile operators, but also software vendors.	http://www.openmobilealliance.org/
OS	Operating Systems	The software that manages the sharing of the resources of a computer and provides programmers with an interface used to access those resources. An operating system processes system data and user input, and responds by allocating and managing tasks and internal system resources as a service to users and programs of the system. At the foundation of all system software, an operating system performs basic tasks such as controlling and allocating memory, prioritizing system requests, controlling input and output devices, facilitating networking and managing file systems. Most operating systems come with an application that provides a user interface for managing the operating system, such as a command line interpreter or graphical user interface. The operating system forms a platform for other system software and for application software. The most commonly-used contemporary desktop OS is Microsoft Windows, with Mac OS X also being well-known. Linux and the BSD derivatives are popular Unix-like systems.	
OTA	Over the Air	Over-the-air programming (OTA) may refer to either free-to-air, terrestrial television, or in the mobile content world, over-the-air service provisioning (OTASP), over-the-air provisioning (OTAP) or over-the-air parameter administration (OTAPA), methods of distributing new software updates to cell phones or provisioning handsets with the necessary settings with which to access services such as WAP or MMS. Some phones with this capability are labelled as being "OTA capable". When OTA is used to update a phone's operating firmware, it is sometimes called "Firmware Over The Air" (FOTA). For service settings, the technology is often known as Device Configuration. Various standardization bodies were established to help develop, oversee, and manage OTA. One of them is the Open Mobile Alliance (OMA).	
Peer or Supplicant		The end-user software that responds to the authenticator. In this document, the supplicant is the ActiveX running in MS Internet Explorer.	

ISSN 0085-7130©Telenor ASA 2007

Acronym /term	Definition	Explanation	Web-resources
PIN	Personal Identification Number	A secret numeric password shared between a user and a system that can be used to authenticate the user to the system. Typically, the user is required to provide a non-confidential user identifier or token (such as a banking card) and a confidential PIN to gain access to the system. Upon receiving the User ID and PIN, the system looks up the PIN based upon the User ID and compares the looked-up PIN with the received PIN. The user is granted access only when the number entered matches with the number stored in the system. PINs are most often used for ATMs (Mini banks) but are increasingly used at the Point of sale, especially for debit cards. Apart from financial uses, GSM mobile phones usually allow the user to enter PIN between 4 and 8 digits length. The PIN is recorded in the SIM card. In 2006, James Goodfellow, the inventor of the personal identification number, was awarded an OBE in the Queen's Birthday Honours List.	
PKI	Public Key Infrastructure	An arrangement which provides for third-party vetting of, and vouching for, user identities. It also allows binding of public keys to users. This is usually carried by software at a central location together with other coordinated software at distributed locations. The public keys are typically in certificates. The term is used to mean both the certificate authority and related arrangements as well as, more broadly and somewhat confusingly, to mean use of public key algorithms in electronic communications. The latter sense is erroneous since PKI methods are not required to use public key algorithms.	
POS	Point-Of-Sale		
RA	Registration Authority	A body given the responsibility of maintaining lists of codes under international standards and issuing new codes to those wishing to register them. A local registration authority (LRA) is an optional part of a public key infrastructure that maintains users' identities from which certification authorities can issue digital certificates.	
RADIUS	Remote Authentication Dial-In User Service	An authentication and accounting system used by many (W)ISPs. When logging in to a public Internet service you must enter your user name and password. This information is passed to a RADIUS service, which checks that the information is correct, and then authorizes access to the WISP. RADIUS is an AAA protocol. It is intended to work in both local and roaming situations. The RADIUS specification is maintained by a working group of the IETF. Defined in IETF RFC 2865.	http://www.ietf.org/, http://tools.ietf.org /html/rfc2865
RAND		A random challenge issued by the network. A number from a pseudorandom number generator function.	
RO	Rights Object (DRM)		
RSA	Rivest, Shamir and Adleman	An algorithm for public-key cryptography. It was the first algorithm known to be suitable for signing as well as encryption, and one of the first great advances in public key cryptography. RSA is widely used in electronic commerce protocols, and is believed to be secure given sufficiently long keys and the use of up-to-date implementations. The algorithm was publicly described in 1977 by Ron Rivest, Adi Shamir, and Leonard Adleman at MIT; the letters RSA are the initials of their surnames. RSA involves a public key and a private key. The public key can be known to everyone and is used for encrypting messages. Messages encrypted with the public key can only be decrypted using the private key.	http://www.rsa.com/, ttp://www.rsa.com /rsalabs/node.asp?id =2125
SAML	Security Assertion Markup Language	An XML-based standard defining a means for making assertions about events, attributes, and policy evaluations concerning subjects. In Liberty usage, SAML subjects are typically Principals.	
SAT	SIM Application Toolkit	The SIM Application Toolkit is a set of commands which defines how the card should interact with the outside world and extends the communication protocol between the card and the handset. With SIM Application Toolkit, the card has a proactive role in the handset (this means that the SIM initiates commands independently of the handset and the network).	
SD	Security Domain (in SIM)		

ISSN 0085-7130

Acronym /term	Definition	Explanation	Web-resources
SIM	Subscriber Identity Module	The SIM is a subscriber identity module for GSM/GPRS subscriptions. In 2G systems the term SIM is used for a dedicated smartcard with subscriber identity information (including security credentials and algorithms). In 3G systems a SIM is an application running on the UICC (smartcard). Although the terms UICC and SIM are often interchanged, UICC refers to the physical card, whereas SIM (in 3G) refers to a single application residing in the UICC that collects GSM/GPRS user subscription information. The corresponding UMTS subscriber application is the USIM (which is always present on a UICC). The SIM provides secure storing of the key identifying a mobile phone service subscriber but also subscription information, preferences and storage of text messages. The equivalence of a SIM in UMTS is a Universal Subscriber Identity Module (USIM). Defined in 3GPP specification series 31.	http://www.3gpp.org /ftp/Specs/html-info /31-series.htm
SIM-Card	See UICC		
SME	Small and Medium Enterprise	Micro, small and medium-sized enterprises represent 99 % of all enterprises in the European Union. The European Commission published in 2003 a revised definition of SMEs. According to this definition, micro-sized enterprises have less than 10 employees and a turnover less than € 2 mill. A small enterprise has less than 50 employees and a turnover of less than € 10 mill. Medium-sized enterprises have less than 250 employees and a turnover of less than € 50 mill.	http://europa.eu.int /comm/enterprise /enterprise_policy /sme_definition /index_en.htm
SMS	Short Message Service	A means by which short messages can be sent to and from digital cellular phones, pagers and other handheld devices. Alphanumeric messages of up to 160 characters can be supported.	
SOAP	Simple Object Access Protocol	SOAP is a protocol for exchanging XML-based messages over a computer network, normally using HTTP. SOAP forms the foundation layer of the web services stack, providing a basic messaging framework that more abstract layers can build on. SOAP facilitates the Service-Oriented architectural pattern.	
SP	Service Provider	A role donned by system entities. In the Liberty architecture, Service Providers interact with other system entities primarily via vanilla HTTP. From a Principal's perspective, a Service Provider is typically a website providing services and/or goods.	
SS7	Signalling System no 7	A CCS (Common Channel Signalling) system defined by the ITU-T. SS7 is used in many modern telecom networks and provides a suite of protocols that enables circuit and non-circuit related information to be routed about and between networks. A set of telephony signalling protocols which are used to set up the vast majority of the world's PSTN telephone calls. The main protocols include MTP (Message Transfer Part), SCCP (Signalling Connection Control Part) and ISUP (ISDN User Part).	http://www.itu.int
SSL	Secure Sockets Layer	A cryptographic protocol which provides secure communications on the Internet for such things as web browsing, e-mail, Internet faxing, and other data transfers. SSL provides endpoint authentication and communications privacy over the Internet using cryptography. In typical use, only the server is authenticated (i.e. its identity is ensured) while the client remains unauthenticated; mutual authentication requires public key infrastructure (PKI) deployment to clients. The protocols allow client/server applications to communicate in a way designed to prevent eavesdropping, tampering, and message forgery. SSL was developed by Netscape, SSL version 3.0 was released in 1996, which later served as the basis for TLS version 1.0, an IETF standard protocol first defined in RFC 2246.	http://tools.ietf.org /html/rfc2246
SSO	Single Sign On	From a Principal's perspective, single sign-on encompasses the capability to authenticate with some system entity in the Liberty context, an Identity Provider and have that authentication honored by other system entities, termed Service Providers in the Liberty context. Note that upon authenticating with an Identity Provider, the Identity Provider typically establishes and maintains some notion of local session state between itself and the Principal's user agent. Service Providers may also maintain their own distinct local session state with a Principal's user agent.	
STIP	Small Terminal Interoperable Platform		
STS	Security Token Service	Microsoft Live Labs Security Token Service is a part of Microsoft's Windows Live range of services. It is an online identity management service which provides an Information Card that enables offloading authentication functions, irrespective of whether the agent signing in is a user logging-in to web sites and services, or a site or service owner to authenticate users. It requires an Information Card compliant store and identity selector to use.	http://sts.labs.live.com/
SWP	Single Wire Protocol		

ISSN 0085-7130©Telenor ASA 2007

Acronym /term	Definition	Explanation	Web-resources
TIMSI (TMSI)	Temporary Mobile Subscriber Identity	TMSI is a 4 octet (byte) unstructured temporary subscriber identity used in the GSM/GPRS/UMTS systems. Subsequent to initial successful location updating and after encryption has commenced the VLR/SGSN may (should) assign a TMSI to the MS. The TMSI is subsequently to be used as replacement for IMSI. The TMSI is assigned in encrypted form and only used in cleartext, and thus there is no externally apparent binding between the IMSI and the TMSI. In effect this provides a (weak) measure of location- and identity privacy for the mobile subscriber. Defined in 3GPP TS 23.003.	http://www.3gpp.org /ftp/Specs/html-info /23003.htm
TLS	Transport Layer Security	TLS is a protocol that ensures privacy between communicating applications and their users on the Internet. When a server and client communicate, TLS ensures that no third party may eavesdrop or tamper with any message. TLS is the successor to the Secure Sockets Layer (SSL).	http://www.whatis.com
UICC	UMTS Integrated Circuit Card	A physically secure device, an IC card (or 'smart card'), that can be inserted and removed from the terminal equipment. It may contain one or more applications. One of the applications may be a USIM. Defined in 3GPP specification series 31.	http://www.3gpp.org /ftp/Specs/html-info /31-series.htm
UMTS	Universal Mobile Telecommunication System	The European member of the IMT 2000 family of 3G wireless standards. UMTS supports data rates of 144 kb/s for vehicular traffic, 384 kb/s for pedestrian traffic and up to 2 Mb/s in support of in-building services. The standardisation work began in 1991 by ETSI but was transferred in 1998 to 3GPP as a corporation between Japanese, Chinese, Korean and American organisations. It is based on the use of WCDMA technology and is currently deployed in many European countries. As of October 2006 there are more than 90 million subscribers worldwide. The first European service opened in 2003. In Japan NTT DoCoMo opened its "pre-UMTS" service FOMA (Freedom Of Mobile multimedia Access) in 2000. The system operates in the 2.1 GHz band and is capable of carrying multimedia traffic.	http://www.3gpp.org/, http://www.umts-forum .org
USB	Universal Serial Bus	USB is a plug-and-play interface between a computer and add-on devices (such as audio players, joysticks, keyboards, telephones, scanners, and printers). With USB, a new device can be added to your computer without having to add an adapter card or even having to turn the computer off. The USB peripheral bus standard was developed by Compaq, IBM, DEC, Intel, Microsoft, NEC, and Northern Telecom and the technology is available without charge for all computer and device vendors.	http://www.whatis.com
USIM	Universal Subscriber Identity Module	An application residing on the UICC used for accessing services provided by mobile networks, which the application is able to register on with the appropriate security. Defined in 3GPP specification series 31.	http://www.3gpp.org /ftp/Specs/html-info /31-series.htm
VLR	Visitor Location Register	The Visitors Location Register or VLR is a temporary database of the subscribers who have roamed into the particular area which it serves. Each Base Station in the network is served by exactly one VLR, hence a subscriber cannot be present in more than one VLR at a time. The data stored in the VLR has either been received from the HLR, or collected from the MS. In practice, for performance reasons, most vendors integrate the VLR directly to the V-MSC and, where this is not done, the VLR is very tightly linked to the MSC via a proprietary interface.	
VoIP	Voice over Internet Protocol	Voice over Internet Protocol is the routing of voice conversations over the Internet or any other IP-based network. The voice data flows over a general-purpose packet-switched network, instead of traditional dedicated, circuit-switched voice transmission lines. Several standards exist to support VoIP, like H.323 from ITU-T and SIP (IETF RFC 3261).	http://www.itu.int, http://www.ietf.org
Wi-Fi	Wireless Fidelity	A term for certain types of wireless local area network (WLAN) that use specifications in the 802.11 family. The term Wi-Fi was created by an organization called the Wi-Fi Alliance, which oversees tests that certify product interoperability. A product that passes the alliance tests is given the label "Wi-Fi certified" (a registered trademark).	http://www.wifialliance .org
WiMAX	Worldwide Interoperability for Microwave Access	A specification for fixed broadband wireless metropolitan access networks (MANs) that use a point-to-multipoint architecture. Based on the IEEE 802.16 WMAN. Published on April 8, 2002, the standard defines the use of bandwidth between the licensed 10 GHz and 66 GHz and between the 2 GHz and 11 GHz (licensed and unlicensed) frequency ranges and defines a MAC layer that supports multiple physical layer specifications customized for the frequency band of use and their associated regulations. 802.16 supports very high bit rates in both uploading to and downloading from a base station up to a distance of 50 km to handle such services as VoIP, IP connectivity and TDM voice and data.	http://www.ieee802 .org/16/, http://www.wimaxforum .org/

ISSN 0085-7130

Acronym /term	Definition	Explanation	Web-resources
WLAN	Wireless Local Area Network	This is a generic term covering a multitude of technologies providing local area networking via a radio link. Examples of WLAN technologies include Wi-Fi (Wireless Fidelity), 802.11b and 802.11a, HiperLAN, Bluetooth and IrDA (Infrared Data Association). A WLAN access point (AP) usually has a range of 20 – 300 m. A WLAN may consist of several APs and may or may not be connected to Internet.	
WPA	Wi-Fi Protected Access	An improved version of WEP (Wired Equivalent Privacy). It is a system to secure wireless (Wi-Fi) networks, created to patch the security of WEP. As a successor, WPA implements the majority of the IEEE 802.11i standard, and was intended as an intermediate measure to take the place of WEP while 802.11i was being prepared.	http://www.ieee802.org, http://www.wifialliance .org
XML	eXtensible Markup Language	The Extensible Markup Language (XML) is a W3C-recommended general-purpose markup language for creating special-purpose markup languages, capable of describing many different kinds of data. It is a simplified subset of SGML. Its primary purpose is to facilitate the sharing of data across different systems, particularly systems connected via the Internet. Languages based on XML (for example, RDF/XML, RSS, MathML, XHTML, SVG, and cXML) are defined in a formal way, allowing programs to modify and validate documents in these languages without prior knowledge of their form.	http://www.w3c.org /XML/

ISSN 0085-7130©Telenor ASA 2007

Kaleidoscope

Technical Slogans as Seen in the Negroponte Switch[1)]

RICH LING

Rich Ling is a sociologist at Telenor R&I

This paper examines the role of slogans in the process of organizing the implementation of technical change. There have been several examples of how slogans have affected the way that we approach the implementation of technical innovation. There was Sarnoff's "radio music box" used to assist the commercialization of the radio. Moore's Law describing the wild expansion of computing power has helped in the marshalling of resources, policy and development in the area of computing. In addition, the Negroponte Switch was posited as a prognosis regarding the future of wire and wireless mediation. There are also those slogans that are off the mark, such as Grosch's Law. Unlike Moore's Law Grosch suggested, "Computer performance increases as the square of the cost. If you want to do it twice as cheaply, you have to do it four times slower." In other words, some slogans have a life span and help us to imagine how the future might be, others are simply wrong. Some, like the Negroponte Switch might provide a broad outline but be wrong in the details. To be successful a slogan needs to encapsulate a complex technical/policy issue, it needs to be pithy and concise and it needs to be uttered by a person with legitimacy in the area. If a slogan is successful it helps to organize institutional capacity for the implementation of technological development.

Introduction

In the 1980s and 90s we were presented with the idea of the so-called Negroponte Switch. The idea was that communication that had gone via cable (telephony) would soon go via radio and that communication that had gone via radio transmission (TV and radio) would soon go via cable. The idea arose as we saw the dawn of both cable TV and mobile communication.

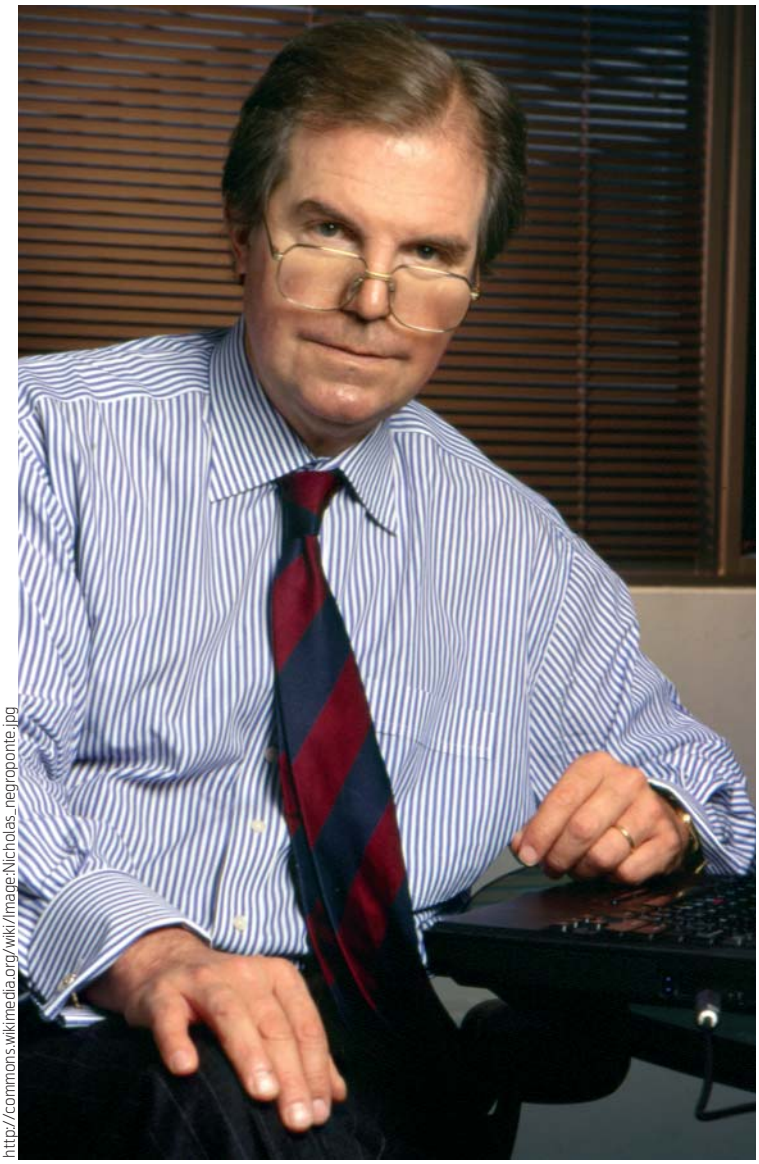
What, however, is the ultimate effect of visionary statements such as the Negroponte Switch? Are they the insights of wise and well reasoned persons who can somehow see better than others? Are they rallying cries that move industry in one direction or another as they crystallize complex issues into a comprehensible insight? Do they have careers where they wax and wane with time? This chapter will look into the history of one such statement, namely the so-called Negroponte switch. I will look at this in particular from the perspective of its role in the development of mobile telephony.

I would like to be so bold as to state the Sawhney principle. To whit, Harmeet Sawhney has suggested that access to unused capacity in a technical system results in creativity. While that may be true, I will posit corollary that the sudden access to new technical possibilities, be this via technical development or regulatory fiat, unleashes creativity, a round of catchphrases, slogans and clichés and, in some, the majority of the seven cardinal sins. Indeed Winston posits the "'law' of the suppression of radical potential" saying that between the initial invention of a technology and its mass acceptance there is a period of slower development where the pre-existing structures need to rearrange themselves in relation to the new arrival (Winston 1998, 11). Winston however, discounts the sloganeering associated with the development of technology as rodomontade or pretentious self-importance. Thus, he points to the same situation as Sawhney but puts another spin on it. He also downplays the importance of clichés in the readjustment process.

Packing this out a bit more, the paper will examine three episodes in the history of technology where either a technical advancement and/or regulatory contortions resulted in new possibilities for mediation, in the first instance the development of broadcast radio and the second is the rise of the integrated circuit and the third as indicated in the title of the paper, the potential of HDTV and the co-temporal development of mobile telephony. In each case, there was some type of technical advancement either on the doorstep, or in the recent past. In each case there was also a crystallizing description of the situation that helped to organize the institutional reaction to the development. There was Sarnoff's description of the radio music box, Moore's Law and finally the Negroponte Switch. The first two were, indeed, associated with developments that eventually led to the situation described by the latter.

1) *This paper is an edited version of the paper "Media Visionaries: Broadcast Radio, Silicon Chips and the Negroponte Switch" that was presented at the Media Technology and Society conference March 24-25, 2006 in Ann Arbor, Michigan. The original paper will appear in a book of the same name edited by W. Russell Neumann.*

ISSN 0085-7130



http://commons.wikimedia.org/wiki/Image:Nicholas_negroponte.jpg

Nicholas Negroponte

With hindsight, it is possible to say that a phrase was, or was not visionary. At one level, that is beside the point. Phrases or slogans such as the Negroponte Switch are a necessary part of the glue that link techno-political developments with the people who are implementing them. In the heat of the institutional scramble to deal with a changing field, these phrases light the way. It is through the establishment and elaboration of these catch phrases that institutions set their course.

The basic mechanism here is technological or regulatory development. It is not enough, however, that the elements of new mediation forms are in place; there is also the need to organize institutional capacity for these developments. It is here that catch phrases such as the Negroponte Switch or Sarnoff's comment on the radio come into play. These slogans encapsulate a complex technical and policy issue, they come from a legitimate source and they need to be pithy enough that they are engrained in the institutional culture where the developments are taking place. That is, they need to be a rallying cry for the troops who are busy with the development of the technology.

In this paper I will look specifically at the historical context that led up to and coincided with the so-called "Negroponte Switch". In addition, with almost two decades of hindsight, I will look into the fate of the phrase and set it into the broader context of the politics of technology development.

The statement was posited in 1989 by George Gilder. Based on interaction with Nicolas Negroponte he asserted that "What goes over the air (broadcast TV and radio) will go via wire and what goes via wire (telephony) will go over the air." Expanding on this, Negroponte wrote some time later:

> George Gilder and I have shared the podium frequently, and I have learned a lot from him. One of our first encounters occurred about 10 years ago at an executive retreat organized by Northern Telecom (now called Nortel). At this meeting, I showed a slide that depicted wired and wireless information trading places. This idea had been prompted, in part, by some early HDTV discussions, during which I and others questioned whether broadcast TV should get any spectrum at all, since stationary TV sets could be better served by wires (read: fiber).
>
> In contrast, the theory continued, anything that moves needs to be wireless. Phones, largely wired at the time, would go wireless, and TV, largely wireless, would get wired. Gilder called this "the Negroponte Switch". (Negroponte 1997)

The Negroponte Switch was posited in the era of "fiber to the home", the development of High Definition TV and the first rumblings of mobile telephony. This was at the dawn of the popularized internet. The technologies that were on the horizon at that point indicated that perhaps the public would be better served if the signals that had traditionally traveled wirelessly (TV and radio) could be transported by landline techniques while those that had traditionally been wire bound (telephony) could be transported through the ether. The need for capacity to transmit huge amounts of video material and the fact that telephony required much less bandwidth indicated that the switch would be logical.

Today, there are large numbers of mobile phone users as well as the large reliance on cable TV. Thus, we might assume that the prophecy is true. Seen in this light, the idea of the Negroponte Switch is prophetic. However, while there are some general lines of agree-

ISSN 0085-7130©Telenor ASA 2007

ment, there are also many devils in the details. In another sense, the phrase also provides insight into how there is a need to develop institutional ideologies in the sense of Berger and Luckmann (1967). That is, there is the need to mobilize institutions in the implementation of a technical or a regulatory vision. Slogans such as the Negroponte Switch serve this goal.

In this chapter, I will first provide the historical context of the rise of broadcast commercial radio and the changes associated with the Negroponte Switch. Following this, I will look into the degree that the Negroponte Switch was prophetic and then examine it as a type of policy slogan.

The History of Wire and Wireless Mediation

To set the Negroponte Switch into a broader historical context it is necessary to trace the development of electrical and electromagnetic technology as applied to interpersonal communication and the broadcast of entertainment, news and commercial content. It is these two, in their role as a foundation for the mediation of voice, images, entertainment and interpersonal communication, that is the technical core of the Negroponte Switch.

The general line of development is that wired point-to-point – and generally interpersonal – communication developed in the mid 19th century in the form of the telegraph and later the telephone. About the turn of the century radio communication developed and by 1920 was also transmitting audio content, not just Morse code. During the early 1920s in the US there was a brief point of convergence – at least at the institutional level – for these two forms of mediation but basically until the late 1980s after the development of the transistor, wired and wireless communication lived their separate and largely parallel lives. With the rise of high-definition TV and cellular telephony the potential again suggested itself that the given practice was not necessarily set in stone. It is at this point that Gilder and Negroponte suggested the idea of the switch.

Ultimately, the prophecies suggested by the switch have to a degree been achieved, but the picture here is quite muddled. The phrase, however, provides us with good insight into the politics of technology development, particularly when faced with the need to mobilize institutions either for certain types of development or to protect themselves from the assault of new techno-regulatory regimes.

Faraday's Contribution to Electricity and Electromagnetism

To start tracing the development of mediated interpersonal and broadcast communication, it is convenient to go back to the time of Michael Faraday. It is perhaps one of the quirks of history that the fundamental scientific basis for both the generation of electricity and the understanding of the electromagnetic spectrum came from the same person. From relatively simple origins Faraday rose to be Humphry Davy's assistant and eventually a member of the Royal Institute in London.

Looking first at electricity, that is basic to landline telephony, Faraday discovered the method for generating this type of power. Previous to Faraday electricity was basically equivalent to Leyden Jars and lightning. It was Faraday, along with Ampere, Ohm, Volta and Galvani, who worked out the basis for modern electrical technology.[2)] Thus, instead of relying on amber rubbed against felt to generate static electricity, Faraday's work led to the development of reliable production of electricity.[3)] This obviously found its application in terms of electrification, and more interesting for our purposes, telegraphy and telephony.

The same work on electrical generators led to the development of electromagnetic communication, in other words radio. Inspired by the work of the Dane Ørsted he carried out a series of experiments that resulted in the discovery of electromagnetic rotation. Faraday's work on electromagnetism inspired the work of Maxwell in Scotland, Hertz in Germany and finally at the turn of the 20th century, Marconi in the development of radio broadcasting (Winston 1998).

Electricity Applied to Communication

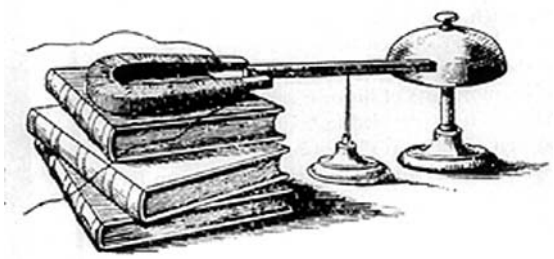
Joseph Henry was the first person to use electricity for the purpose of signaling, that is primitive communication. Henry did this by applying the principles of electricity to various applications in the early 1830s. He developed a system for using magnetism to remotely ring a bell that was the forerunner of telegraphy. In 1837, of course, Samuel Morse received his patent for telegraphy after learning of Faraday's work and that of Henry.

---

2) *Thomas Edison, of course, was a telegraph operator early in his career and went on to make improvements in telegraphy and also to engage in a pitched battle with George Westinghouse as to the benefits of AC and DC power. Foreshadowing the later discussion of Negroponte et al, there was a serious campaign to push the development trajectory of DC power by Edison, and doubtlessly Westinghouse. Each saw the unmet niche of electrification of the home and pressed their case using various forms of what we would call spin.*

3) *Along the way electricity was used for a variety of odd, morbid and even bizarre functions including various forms of therapy, execution and entertainment (Elsenaar and Remko 2002).*

ISSN 0085-7130



Moving to the institutional realm, wire based telegraphy found two niches that immediately assisted in its growth. The first was as a signaling channel for the control of the railroad (Standage 1998) and the second was the transmission of time-sensitive financial information. While there was a period of competition in the industry, the structure of the telegraph industry moved towards monopolization and by the end of the 1800s, Western Union in the US, The British Post office in the UK and Telegrafverket in Norway were the monopolists. Internationally the International Telegraph Union (ITU) was formed in 1865 to develop standards for international interaction.

The roughly parallel but time shifted development of telephony for interpersonal communication followed somewhat the same development as telegraphy. In the US, based on the developments of Alexander Graham Bell, the Bell telephone eventually formed into American Telephone and Telegraph (AT&T). Western Union was offered the possibility of purchasing the patents for the telephone. From this episode we have one of the oft repeated citations that William Ortion, the head of Western Union felt that the telephone was an interesting toy but would not have practical implications (Winston 1998, 54). As with the later Negroponte Switch, the fact that this quip is remembered – and even savored by telephone people – points to the social dynamics of institutional adoption of technology. It is an example of just how much we can get it wrong. Orton was so thoroughly entrenched in the hegemony of Western Union he failed to grasp the fundamental shift presented by the telephone. In this case, we see a phrase being used against the dominant technology of the time.[4)]

Through a series of consolations and the idea of "Universal service" the reach of the telephone expanded. By the start of the 1920s wire based interpersonal and voice based communication was a well established institution in major cities and in many rural areas (Fischer 1992). The telephone and the telegraph co-existed for many years with the telephone gradually taking more and more of the traffic. Western Union had a niche in the area of financial services and the "wiring" of monies that is still essential in many third world countries.

The role of the telephone was, however, not simply for interpersonal communication. While the telephone was not designed to send or receive audio with any fidelity (de Sola Pool 1983), it was nonetheless used for the transmission of various types of entertainment. In the 1870s music was transmitted over the telephone. It seems that there was a positive craze for "telephone concerts" in the late 1870s. On January 29, 1878 one affair was in Warren County Pennsylvania in which singers and musicians performing in Jamestown, NY were heard in the hall through the use of the telephone and included a speech by Thomas Edison introducing his new "phonograph" (Warren County Historical Society 2005). From New Orleans in 1879 we hear of a "telephone concert" given by the telephone company with Miss Minnie Wolf singing the Pizzicato Polka and other pieces (New Orleans Public Library 2005) and in lonely Lake City, Colorado – a town without telephony at the time – in March of 1878 we learn of William Penn Harbottle, the temporary editor of the local Silver World and who, among his other talents, claimed to be a telephone-concert tenor horn soloist (Thompson 1974). Marvin reports on telephone concerts being sent from New York to Rochester and Buffalo in the 1890s (1988; see also Nye 1997). Thus, there was the embryonic idea of broadcasting that, while mediated by the telephone, included the one-to-many structure of the later industry.

On the radio side, Marconi was at work during the latter part of the 19th century to develop a practical method for the transmission of telegraph signals. During this period he continued to push the boundary for the transmission of telegraph signals until transoceanic communication became possible. In December 1901 he had sent a message across the Atlantic from Newfoundland to the UK.

At the dawn of the 20th century, radio was basically unregulated, limited to Morse code, and was largely the realm of hobbyists. However, its ill-fated role in the Titanic disaster and the use of radio in World War I led to the regulation and the eventual commercialization of the airwaves. This along with the technical development of voice modulation led to a genuine radio craze in the 1920s.

4) *Interestingly, this phrase is also recorded for other persons in different countries. According to (Fladby 2003), when the owner of a bank in Drammen, Norway was shown the telephone he was reported to have said something similar. This indicates that either the phrase is an urban legend that is retold in appropriate situations, or that the banker somehow knew of the comments by the Western Union executives.*

ISSN 0085-7130©Telenor ASA 2007

The amateur handling of news regarding the Titanic along with a deep and lurid interest in the fate of so many people turned the public about lassez faire regulation of the radio spectrum (Douglas 1987; Hargittai 2000). There was the convening of the International Radio-Telegraphic Convention in London in 1912 where it was decided to require that ocean going passenger vessels had to have a wireless communication system that was to be staffed 24 hours a day. In addition, there was the passage of the US Federal Radio Act of 1912 that was the first US government involvement in this area. It required the licensing of operators and set aside frequencies for emergency communication.

The movement from radio based Morse code to the modulation of voice continued into the 20th century. The work of Lee de Forest resulted in the vacuum tube. This was essential since it amplified signals and allowed for the wireless transmission of voice. This period also saw some of the first use of radio for the distribution of entertainment. The idea of broadcasting was starting to coalesce. One metaphor that captured the idea of using the technology for the distribution of entertainment.

> The 'Radio Music Box' can be supplied with amplifying tubes and a loudspeaking telephone, all of which can be neatly mounted in one box. The box can be placed on a table in the parlor or living room, the switch set accordingly and the transmitted music received. There should be no difficulty in receiving music perfectly when transmitted within a radius of 25 to 50 miles. Within such a radius there reside hundreds of thousands of families; and as all can simultaneously receive from a single transmitter, there would be no question of obtaining sufficiently loud signals to make the performance enjoyable. (Sarnoff 1920)

In January of 1917 Lee de Forest used his vacuum tube[5)] radio system to broadcast music in a "Concert by Wireless" and a month later broadcast a "Wireless Dance" (QST 1917). Writing in the April 1917 edition of the magazine QST de Forest reported:

> A novel request was one from two gentlemen in Newark, NJ, who asked that on a certain evening we play dance music. This, in order that their guests of that evening, to the number of one hundred, might dance to our Graphonola Orchestra furnished us nightly by the Columbia Graphophone Company. We heard afterwards that this dance was a great success, as was the previous one in Morristown, NJ, for which we also provided the music at Highbridge, NY, thirty odd miles away. (de Forest 1917)

The entry of the US into World War I temporarily halted this development. Amateur radio was suspended during the war for fear of spying and the US Navy took over all radio signaling. Further, the fear of losing control of the radio spectrum prompted the government to regulate it in the years after World War I. This included the organization of Radio Corporation of America (RCA) that resulted from the nationalized Marconi America. The company which was controlled by AT&T and General Electric was given the license to produce radio equipment in the US. Further, the government started the regulation of the radio spectrum and required licensing for those who wished to operate a radio station. (For a discussion of this see: Hazlett 2001; and Moss and Fein 2003.) This meant that it controlled the patents for vacuum tubes which gave it a *de facto* monopoly. In addition, after a short, troubled marriage to AT&T it controlled radio and eventually TV broadcast in the US.

Thus, by the end of World War I the trend towards the canalization of wired and wireless mediation was falling into place. On the one hand point-to-point communication was carried out via wired systems, at this point the somewhat competing systems of telegraph and telephone. In addition, the elements for the development of broadcast radio were on the table and they were starting to be used for what we recognize as broadcast entertainment.

Embedding of the Channels

There were several issues that resulted in the canalization of telephony in the wired world and broadcast in the wireless world. Not to be a technological determinist, but there are good technical reasons for the paths chosen in the early years of telephony and radio. Following Farley:

> As the vacuum tube and the transistor made possible the early telephone network, the wireless revolution began only after low cost microprocessors, miniature circuit boards, and digital switching became available. (Farley 2005)

Thus, as of the 1920s there were not the technical possibilities available for any form of switched interpersonal radio communication such as we now have

5) *De Forest's vacuum tube was a variation of the tube developed by Ambrose Fleming. The similarity between the two led to endless patent disagreements and, until Fleming's patent expired in 1922 these disagreements caused the delay of broadcast radio. (Winston 1998)*

ISSN 0085-7130

with mobile telephony. This is not to say that there was not overlap. Indeed early radio operators sent personal messages to one another and, as we have seen above, the telephone was used for the distribution of entertainment and news.

Looking at radio, there are other considerations. In the early 1920s there was little understanding of whether broadcast technology would attract an audience. Up to that point, what we consider as broadcast radio had been dominated by amateurs who were interested in both sending and receiving transmissions. This metaphor was, to a certain degree, explored by AT&T in a period where they explored the development of radio broadcasting. According to John Brooks, there was a notion that AT&T owning radio stations, starting with WEAF in New York, would allow a telephone subscriber to call into the station and give a radio talk. Thus, radio would be supported by rental of the transmitter. The open time was to be filled with music (Brooks 1976).

Radio broadcasts for the purpose of entertainment or news started with the Westinghouse owned station KDKA and their reporting of the presidential election in 1920. Soon major events such as the Dempsey – Carpentier boxing match and baseball games were becoming a regular feature of radio broadcasts (Ackerman 1945).

The first use of the medium for advertising came in August 1922 when a real estate developer bought 15 minutes to promote a housing development called Hawthorne Court. There were soon others, and eventually radio developed a mixture of entertainment (music, sports, comedy, theatrical performances, etc.) and commercial pitches. The owners of the station worked to manage the boundary between entertainment and commercials. The pattern was nonetheless set. In addition, many types of organizations applied for and received licenses for broadcasting during this period. In addition to major corporation newspapers, there were department stores, YMCA clubs, universities and churches. Following Douglas some of the connections were logical while others reflected the breadth of interest in radio during this period.

> [...] in those euphoric months of early 1922 radio stations were licensed to some very eccentric and inexplicable owners. There was the Yahrling-Rayner Piano Company of Youngstown, Ohio (WAAY); the Palmer School of Chiropractic of Davenport, Iowa (WOC); the C.F. Aldrich Marble and Granite Co. of Colorado Springs, Colorado (KHD); the Omaha Grain Exchange (WAAW); and even the Nushawg Poultry Farm of New Lebanon, Ohio (WPI). (Douglas 1997)

While there was a clear commercial drift in the US, the situation was different elsewhere. Looking to Europe at about the same time, the BBC was being organized in the UK. After a short period of commercial radio, broadcasting in the UK was nationalized and developed programming to be broadcast to its then wide flung colonial empire. The model, that is widely copied, relied on licensing fees in lieu of commercials. Thus, instead of paying for radio when the listener bought toothpaste or shampoo (as in the commercial US system) he or she would pay via an annual license fee. In addition to catering to the desires of the listeners, this model also had the explicit mission of educational and public service programming. The motivation, in the words of its first Director General, John Reith was in typically gendered terms “Making the nation as one man”. Thus, we come again to the interaction between ideological perspective and the development of a techno-institutional organization.

When thinking of the fast coming canalization of wireless broadcast and wire based telephony, this seems to be a particularly plastic moment in history. On the one hand, the soon to be premier radio broadcasting company RCA and the premier telephone company AT&T were indeed in a loose partnership. By 1923 AT&T was able to open a second radio station in Washington that used its telephone network to carry the signal between New York and Washington. Later that year, AT&T even used its telephone system in conjunction with local radio broadcasters to air the first nationwide address by President Harding.

The cooperation between RCA and AT&T was not easy to maintain. AT&T tried to enforce various types of monopolies and set what RCA believed to be inappropriate prices for use of the telephone transport between radio markets. There were lawsuits and various kinds of difficulties. By the middle of 1926 AT&T had sold its radio stations and had agreed to supply RCA with the telephone network in order that they could distribute their radio programs. Thus while there was a wire based network for the transport of what we have come to call content between local radio stations, the final distribution in a city was via ether.

By the mid 1920s the pattern had been cast. While both telephony and radio communication had been used for several decades at this point, it was only when AT&T backed out of the broadcast industry that the pattern upon which Negroponte and Gilder commented became the norm. As we enter into the first two decades of the 20th century, there is the well entrenched wire based transmission of interpersonal communication and there is the nascent broadcasting

ISSN 0085-7130©Telenor ASA 2007

of entertainment, news and not incidentally, commercials via radio.

The Convergence of Wired and Wireless

From the early 1920s until the late 1970s the model of wired interpersonal communication and wireless broadcasting went almost unquestioned. However, the development of first the transistor by Shockley, Bardeen and Brattain in 1947, and the resulting development of the integrated circuit in 1959 by Jack Kilby and independently by Robert Noyce changed the situation. Their development radically reduced the energy and size needed for electronic devices and thus enabled much of the technology development in the latter part of the 20th century (Winston 1998, 220). These developments paved the way for the changes brought by cable TV, the promise of HDTV and the growing development of wireless communication.

Looking first at wireless telephony, in the period after World War II there were several small trials with the intention of introducing a wireless local loop into the "switched" telephone service. These were the first steps associated with the eventual development of mobile telephony. As reported elsewhere (Ling 2004), one of the first such trials was carried out in eastern Colorado near the town of Cheyenne Wells. It was expensive to set up the telephone lines to the wide-spread farmers in the area. As an indication of the distances, some of the "local" farmers used airplanes to commute into town. From the perspective of the telephone company, it was potentially more efficient to connect the farmers into the system via radio. In town a telephone operator could patch the calls into the traditional wire-based system. In this case, the cells were many tens of miles in diameter. The installations at the farmers' homes were stationary. Thus there was no need for systems to deal with "handoff" between cells and indeed the cells were quite spacious when compared with today's. Radio had, however, entered into the realm of switched telephony.

The next advancement in cellular telephony came in the late 1960s with the trials on the New York – Washington Metroliner. In this trial a system was developed by AT&T that allowed the calls from the train traveling between the two cities to be handed off between cells. In order to deal with this system the engineers had to plan frequency use so that adjacent cells were not operating on the same frequency and thus interfering with each other. The development of the transistor as a supercharged version of the vacuum tube also allowed the development of first "luggable" and later quite portable handsets. Thus, in the latter part of the 20th century the development in telephony has seen the rise of smaller and smaller mobile handsets that move away from the geographically fixed landline telephone.

Looking back to the realm of broadcast, TV arose as a popular medium in the 1950s and 1960s (Schwartz 2002). Advances in television including the rise of cable TV, the attempted development of HDTV and digital TV also arose from the development of the integrated circuit (Winston 1998, 140). Cable TV started to become a force in television distribution in the late 1970s with the rise of channels such as HBO and CNN. Since that time it has been a major conduit of information into the home. The major impact has been in terms of broadcast TV, but also internet and telephony have been delivered via the "TV" cable. Interestingly, as these words are being written, television is also being offered via mobile telephone handsets.

Thus, in one way we are moving back to the unity that characterized the first development of electricity and electromagnetism. This is of course an incorrect statement. The ideas of Faraday did not come close to envisioning the two paths of development traced here. On the one hand, electricity was used as a medium for communication via wired channels first in the guise of the telegraph followed by the telephone and later by broadcast communication that also can include telephony. Electromagnetism developed first into radio based Morse telegraphy and later into voice and visual broadcast. With the development of the transistor it was also pressed into service as a form of mediated interpersonal interaction in the form of the mobile phone that is now also becoming a TV terminal.

The process has been described at different points by various persons. Sarnoff's "music box", Moore's prolific transistors and Gilder's formulation are three particularly central examples. Each of them was a prognosis and each of them also functioned to direct the institutional mobilization required for the development of the respective systems.

The Negroponte Switch as a Technical Prophecy

With almost two decades of hindsight we now have a chance to see the value of the prophecy. As noted in the introduction, the widespread adoption of mobile telephony and cable TV seems to indicate; yes, that this switch has indeed happened. In addition, there is the growing use of mobile communication points in the same direction. Point-to-point interpersonal communication had a long life in the wired world and has started to move into the wireless sphere. Thus, we can

ISSN 0085-7130

perhaps assume that Gilder and Negroponte's prophecy has come true.

However, the details are not quite so clear. If we are only thinking of Plain Old Telephony as seems to be the case with Gilder and Negroponte, the slogan holds up. However, taking a few steps back, other issues arise.

Looking at this from the perspective of mobile communication there is undeniably a wireless element. However, there is a lot of interpersonal communication that happens over wired systems. A lot of e-mail and IM is still wired, and significant portions of the population, particularly the elderly still use wire based voice telephony.

Thinking of local radio connections, if we are using a traditional mobile telephone (GSM, CDMA, etc.) or if we use an advanced phone with a Wi-Fi connection and, for example a Skype client, or if we use a so-called Wireless Local Loop system such as Little Smart that is widespread in China, the interaction makes the first part of its journey through the ether (Castells et al. 2004; Sandvig, Young, and Menrath 2004). After that, however, it is back into the wired world. The series of base stations, routers, backbone, etc is all wire based.

If we look at the local situation, Gilder was right; if we look at the broader system, he was not. While in some countries and for some groups, wireless is the dominant form of voice mediation the idea does not hold up in all cases.

Arguing from a slightly different perspective a lot of entertainment has become cable based. This said, there is still a relatively large public for satellite based TV, and increasingly radio. In addition, the traditional terrestrial TV broadcast system is still in place and occupying radio frequency. Thus, there has been the shift that Gilder suggested, but it is partial and has not necessarily resulted in tidying up the resource allocation issues.

There is also a definitional question here. Increasingly, people are using local Wi-Fi (read wireless) connections within their homes to afford them mobility and to avoid some of the "wire spaghetti" that seems to be a part of the PC world. Through these local wireless connections they are working (and engaging in interpersonal communication). They are also surfing the net as a form of entertainment. More to the point, they are downloading music and viewing video that are decidedly entertainment and formerly the turf of the broadcast industry. While the bits that constitute the entertainment flow through different wire based pipes (cable, copper based DSL, etc.), the last critical "local loop" is wireless. To the degree that this is going on, the success of Gilder's prophecy becomes a framing issue. In this case, there is the opposite outcome as when compared to interpersonal communication. If we look at the broader system he was right on. If we look at the immediate user configuration the answer is not so clear.

Negroponte himself has also posed the same question. In 1997 he wrote:

> Was the Negroponte Switch correct after all? ...
>
> A decade later, it seems that this whole switching of places has been contradicted left and right. Satellite TV is doing fine. HDTV just got new spectrum. And the cable business is starting to include telephony. So how should one look at RF [radio frequency] today? (1997; see also Negroponte 2002)

George Gilder, the person who originally posited the name was also in doubt. He wrote, "By 1994 the vision of scarce spectrum behind the Negroponte switch was in a rout" (1994). In a subsequent article he sketched some of the scenario outlined above when he noted:

> In an era of bandwidth abundance, the Negroponte switch – with voice pushed to the air and video onto wires – may well give way to this division between fibersphere and atmosphere. With the fibersphere offering virtually unlimited bandwidth for fixed communications over long distances, the local loop will be the bottleneck, thronged with millions of wireless devices. Under these conditions, a move to high-frequency cellular systems is imperative to carry the increasing floods of digital video overflowing from the fibersphere. (1993)

Others have suggested that economic mechanisms can address some of the spectrum constraints and ease the issues associated with the transition of TV from its analogue era into the coming digital era (Hazlett 2001).

There are clear prophetic elements to the idea of the Negroponte switch and indeed some of the technological changes suggested by it have been realized. However, it would overstate the case to say that reality has followed the plan. The unforeseen rise of Wi-Fi and other technological changes have skewed the picture. In spite of this, for a brief period, the Negroponte Switch was seen as a clear vision of technical development. It crystallized the gist in both the direction of technical developments at the time and pointed to the

ISSN 0085-7130©Telenor ASA 2007

problems being faced by those developments. However, the introduction of other technologies into this mix changed the situation.

The Negroponte Switch as a Policy Slogan

Were the Negroponte Switch simply a technical prophesy, it would soon have been forgotten. It was, however, much else. The reason that this phrase is so well entrenched is that it summarized a complex technical situation, it stated a probable outcome, it came from a very legitimate source in the form of Gilder and Negroponte, and that it also helped to marshal activity in important sectors of the society. In later life it serves as a type of benchmark with regard to the political, technical and social vectors of the time and it is indeed still being debated. Further, the statement came at a time when extra capacity was problematic. There were policy issues and technical futures at play. For these reasons, it gained legs in the minds of various persons who were engaged in the daily work of either developing or marketing technologies where the turf of the "other" group was for some reason desirable.

The Negroponte Switch was a successful slogan. It was used in the mobilization of certain social forces pushing for or alternatively resisting the establishment of a new technological regime. It was a call to arms for those wishing these chances to be made and it was also a warning to those wishing for status quo.

As we have seen, the broad sweep of the phrase has been achieved, or perhaps not, depending on the framing of the data. What is interesting from a sociological perspective, however, is that the phrase crystallized the tensions between significant institutional actors. Being coined at a meeting of landline telephone executives we might well suspect that may have scandalized the meeting, or at least those executives whose jobs it was to maintain the copper based telephone system. It probably also energized the troops associated with the development of HDTV and mobile telephony.

To be sure, it was pithy, quick, it seemed to easily encapsulate broad trends in society and it came from highly credible sources that also had access to publication systems where it could be spread to the far corners of the earth. Thus it is not difficult to imagine that it soon appeared in hundreds, if not thousands of corporate presentations associated with the planning and development.

The cable industry saw it as a summary of how the development of technology would eventually trump the terrestrial broadcasting industry. In a similar way the radio based communication industries, such as mobile telephony, saw it as fitting into their campaigns to gain access to additional radio spectrum.

In this respect, it is far from unique. There are many phrases and slogans that are pressed into service in this way. William Ortion's description of the telephone as a toy, Sarnoff's radio "music box" and Moore's Law have also served a similar function.[6)]

An interesting contrast is seen in the form of Grosch's Law from 1956 that noted "computer performance increases as the square of the cost. If you want to do it twice as cheaply, you have to do it four times slower." This competed with Moore's Law which posited that the characteristics of computers would double every 18 months. The former statement suggested that the direction of development for computers would be for larger and larger machines. Here the politicking was between those advocating a few big computers and many small ones, and developments show that Grosch's side lost in this discussion. Interestingly, however, there is a meta-text associated with each of these two alternatives. The implication with Moore's Law is perhaps associated with the inevitability of the PC revolution. The social use of Grosch's Law, to the degree that it is remembered is that it points to just how bad we – or perhaps poor Grosch – can be when trying to make prognoses.[7)] These other "laws" and slogans have played a similar role to that of the "Negroponte Switch."

---

6) *In a more contemporary example, this time coming from Norway, there is a slogan regarding the increasing reach of internet protocol that states "Alt over IP og IP over alt." (Everything via IP and IP everywhere) as is John Reith's vision for the BBC "Making the nation as one man." As with Moore's Law, this is a statement used in the mobilization of resources.*

7) *As a perhaps flawed, but interesting meter of popularity, there are 136,000 mentions of the Negroponte Switch on the web as of this writing. There is, however, certain confusion as to its application. In most cases it refers to the cable/mobile phone exchange outlined above. In other cases it is more a reference to a physical switch that would re-route these two streams of information and in some cases it refers to Negroponte's idea regarding the replacement of atoms with bits. By this measure Moore's Law is more thoroughly ensconced since it is mentioned on approximately 10.1 million pages while poor Grosch has his law mentioned only about 800 times. Sarnoff's comment on the radio as a "music box" is cited 19,600 times and the quip made by Western Union about the telephone being nothing more than a toy has 43,000 referrals.*

ISSN 0085-7130

Conclusion

The development of technical regimes is a complex process. There are technical developments, regulatory issues and there is the need to mobilize large institutions either in support or in the opposition to the change in technology. In an era of convergence we increasingly meet these issues.

It is into this situation that phrases, such as the Negroponte switch, play a role. These phrases, coined by central people in the development milieu, often summarize a complex system and help others to understand the issues at play.

These catch phrases also have a career. They can become received truth regarding the inevitability of a certain type of development (Moore's Law), a cliché that may even become a straw horse (Western Union's preliminary evaluation of the telephone), a summary that has perhaps a limited shelf life (the Negroponte Switch), or forgotten (Grosch's Law). This is determined by the degree to which they are oversold and by the degree to which they are overtaken by events.

Bibliography

Ackerman, W C. 1945. The dimensions of American broadcasting. *Public Opinion Quarterly*, 9, 1–18.

Berger, P, Luckmann, T. 1967. *The social construction of reality : a treatise in the sociology of knowledge*. New York, Anchor.

Brooks, J. 1976. *Telephone : The first hundred years*. New York, Harper and Row.

Castells, M, Fernandez-Ardevol, M, Qiu, J L, Sey, A. 2004. *The mobile communication society: A cross-cultural analysis of available evidence on the social uses of wireless communication technology*. Los Angeles, Annenberg research network on international communication.

de Forest, L. 1917. *DeForest Wireless Telephone*. October 3, 2007 [online] – URL: http://earlyradiohistory.us/1917df.htm

de Sola Pool, I. 1983. *Technologies of Freedom*. Cambridge, MA, Harvard University Press.

Douglas, G M. 1997. *The early days of radio broadcasting*. Jefferson, NC, McFarland and Co.

Douglas, S. 1987. *Inventing American broadcasting 1899–1922*. Baltimore, Johns Hopkins University Press.

Elsenaar, A, Remko, S. 2002. Electric body manipulation as performance art: A Historical perspective. *Leonardo music journal*, 12, 17–28.

Farley, T. 2005. *Privateline.com : Telephone History*. October 3, 2007 [online] – URL: http://www.privateline.com/TelephoneHistory3A/numbers.html

Fischer, C. 1992. *America calling : a social history of the telephone to 1940*. Berkeley, CA, University of California.

Fladby, R. 2003. *Liers historie*. October 30, 2007 [online] – URL: http://www.lier.kommune.no/Liers-Historie/telegraf_og_telefon.htm

Gilder, G. 1993. The new rule of the wireless. *Forbes ASAP*, March 29.

Gilder, G. 1994. Life after television, revisited. *Forbes ASAP*, February 23.

Hargittai, E. 2000. Radio's lessons for the internet. *Communications of the ACM*, 43.

Hazlett, T W. 2001. *The US digital TV transition : Time to toss the Negroponte switch*. Washington, DC, The Brookings Institute.

Ling, R. 2004. *The Mobile Connection : The cell phone's impact on society*. San Francisco, Morgan Kaufmann.

Marvin, C. 1988. *When old technologies were new : Thinking about electric communication in the late ninteenth century*. New York, Oxford University Press.

Moss, D A, Fein, M R. 2003. Radio regulation revisited : Coase, the FCC, and the public interest. *The journal of policy history*, 15.

Negroponte, N. 1997. Wireless revisited. *Wired*, August.

Negroponte, N. 2002. Being Wireless. *Wired*, 10.

New Orleans Public Library. 2005. *Index to Riders' Digest*. New Orleans, New Orleans Public Library. Available from: http://www.nutrias.org/info/louinfo/ridersdigest/index.htm

Nye, D E. 1997. Shaping communication networks: telegraph, telephone, computer – Technology and the rest of culture. *Social research*, Fall.

QST. 1917. A concert by Wireless. *QST*.

ISSN 0085-7130©Telenor ASA 2007

Sandvig, C, Young, D, Menrath, S. 2004. Hidden interfaces to 'ownerless' networks. In: *The 32nd Conference on Communication, Information, and Internet Policy*, Washington, DC.

Sarnoff, D. 1920. *"Radio music box" memo*. October 3, 2007 [online] – URL: http://earlyradiohistory.us/1916rmb.htm

Schwartz, E I. 2002. Televisionary. *Wired magazine*, 10, 5.

Standage, T. 1998. *The Victorian Internet*. London, Weidenfeld and Nicolson.

Thompson, T G. 1974. *Lake City, Colorado, An early social and cultural history*. Oklahoma City, OK, Metro Press.

Warren County Historical Society. 2005. Warren, P A. *Warren County Historical Society*. Available from: http://www.warrenhistory.org/PHMCapp5.htm

Winston, B. 1998. *Media technology and society*. London, Routledge.

Rich Ling is a sociologist at Telenor R&I. He received his Ph.D. in sociology from the University of Colorado, Boulder in his native US. Upon completion of his doctorate, he taught at the University of Wyoming in Laramie before coming to Norway on a Marshall Foundation grant. Since that time he has worked at the Resource Study Group and has been a partner in the consulting firm Ressurskonsult. Since starting work at Telenor he has been active researching issues associated with new ICT and society. He has led projects in Norway and participated in projects at the European level. Rich Ling has published numerous articles.

email: richard-seyler.ling@telenor.com

ISSN 0085-7130

Call for Papers

– Special issue on "ICT Forecasting"

telektronikk

Telektronikk – Telenor's Journal of Technology

Telenor will support the ISF2008 in Nice, France, which is dedicated to Information and Communication Technology Forecasting, by making a special issue on ICT Forecasting.

For supporting ISF, contributors are asked, if possible, both to present the paper at ISF by submitting the traditional abstract via the ISF website (http://forecasters.org/isf/) and to submit an extended abstract for the journal issue. The following instructions apply only to the journal issue.

Submissions and Selection Criteria

Extended Abstracts of a minimum of *800 words and a maximum of three pages*, including figures and diagrams, are requested by *February 29, 2008*. They will be evaluated by an International Committee of experts based on quality, novelty, and relevance to ICT forecasting issues.
Key words: Long-term telecommunication forecasting, Information and communication technology forecasting, Demand models for new and established telecommunication services, Forecasting and techno-economic modelling, Forecasting models and uncertainties, Risk analysis, Competition, etc.

The full papers based on selected Extended Abstracts will be published in the special issue *"ICT Forecasting"* in the journal *Telektronikk*, vol 104, No 4, 2008.

Please visit the journal website: http://www.telektronikk.com.

Guidelines for Submitting Extended Abstracts

The Extended Abstract must include the following points:

On top of the first page, authors must indicate abstract title, the name(s) of the author(s), mailing address, telephone and fax numbers, and e-mail address.

Authors must indicate clearly how their Extended Abstract will be developed to the full paper.
The Extended Abstracts (typed in 12 pt. font Times New Roman) should be in Portable Document Format (pdf) and submitted by e-mail to: telektronikk@telenor.com. *Please start the e-mail title with the keyword "ISF"*.

In case of any problem in submitting, please contact: Ms Gunhild Luke, gunhild.luke@telenor.com.
For all other questions about the special issue, please contact the Feature Editors: *Dr Kjell Stordahl*, kjell.stordahl@telenor.com, or *Mr Nils Elnegaard*, nils.elnegaard@telenor.com

ISSN 0085-7130©Telenor ASA 2007