

# Combining IMS and Web 2.0 – Assessing the Service Availability

Arlene Pearce [pearce@pvv.ntnu.no](mailto:pearce@pvv.ntnu.no), Terje Jensen and Judith E. Y. Rossebø{[terje.jensen1](mailto:terje.jensen1), [judith.rossebo](mailto:judith.rossebo)}@telenor.com  
Telenor Research and Innovation, NO-1331 Fornebu, Norway

**Abstract**—Combining Web 2.0 and IMS has gained traction for several reasons. Drivers include users' interest in publishing content, innovation volume behind Web solutions and opening channels for commercial interests. In such an environment several components of a service should be orchestrated in order to provide a persistent user experience. Commonly a browser-like interface is then implemented and the interface could be divided into a set of regions. Then, certain parts of the service may work to the complete satisfaction of the user, while other parts are not quite up to the planned behaviour. It is usually also included as the planned behaviour that unauthorized users should not get access to the services or relevant data.

Availability requirements grow backed by increasing commercial interests and growing volume. Accessibility and exclusivity are the two essential characteristics of availability that need to be addressed. This paper presents an enhanced model for service availability intended to capture characteristics of orchestrated services. Application of this model is done by illustrating the combination of external Web 2.0 applications like Facebook and IMS Presence information in the IMS domain.

**Index Terms**—IMS 2.0, Web 2.0, Facebook, Flickr, del.icio.us, Availability concept, service orchestration.

## I. INTRODUCTION

WEB solutions are finding their growing place in the telecom environment. One example of this is the drive behind combining IMS (IP Multimedia Subsystem) and Web 2.0. Two appealing factors related to web solutions include the innovation volume and the modularity concept. This is also fuelled by the commercial and individual users publishing interests. Collating the user experience would imply that a number of service components should be orchestrated. In addition, individual user preferences should be respected. Those preferences would likely be distributed across several systems, within a provider's IMS servers, but also across servers maintained by partners.

Preparing for supporting agile services and overall cost savings, more and more telecom services migrate from dedicated networks to a common IP-based infrastructure. Keeping in mind key commercial and social importance of telecom, the IP-based infrastructure must support services complying with acceptable availability requirements.

Traditionally, the notion of availability has been defined as the probability that a system is working at time  $t$ , and the availability metric has been given by the uptime ratio,

This work has partially been funded by the Research council of Norway project SARDAS (152952/431) and partially within the Eureka project Mobile Fixed Convergence in Multiaccess Environments, Mobicome.

representing the percentage of time that a system is up during its lifetime [1]. Accompanying this interpretation, failure reporting procedures have also been described, e.g. [2] for Public Switched Telephony Network, PSTN. This understanding has served well for describing and analyzing availability of services delivered in dedicated networks such as for voice services in the PSTN/ISDN. However, for describing service availability characteristics and analyzing availability of services in the vastly distributed environment in which IP-based services are deployed, an enhanced notion of availability is required.

Considering the emerging range of IP-based services being delivered in public and private networks today, several challenges follow from the traditional understanding of availability. This paper addresses two challenges. The first challenge is that even with a high mean rate of availability, failure that occurs during peak service request periods will result in high operational loss. One such scenario is a web service with 99.999% average availability that loses connectivity for 5 minutes during peak sales of concert tickets. Such bursty behaviour patterns could be seen for several of the services [3]. The second challenge is that when presented with a set of service components, a user may have different expectations of quality for each component. One example is a buddy list with Presence information fed by various IMS Presence agents.

In the multi-application environment implied by IMS, several features may contribute to the overall user experience. For example, the user interface may collect parts of Presence information, location-dependent data, calendar tasks, streaming video and other service components. Different parts of the user interface may be updated by different servers. Hence, the user experience is collated from different sources. Moreover, the different parts of the user interface may have different weights in the experience depending on the user tasks.

We focus on the problem of considering the variability of the contributions of the different parts involved in the services to the overall availability of a service. Issues are the adequacy of the current availability concept, the definition of availability and the availability management.

The service availability concept model motivated and introduced in [4] is presented and exemplified by a case. This paper, building on [5], presents and elaborates on a conceptual model for service availability providing a case study to demonstrate the applicability of the model to service provisioning in a distributed IMS service environment.

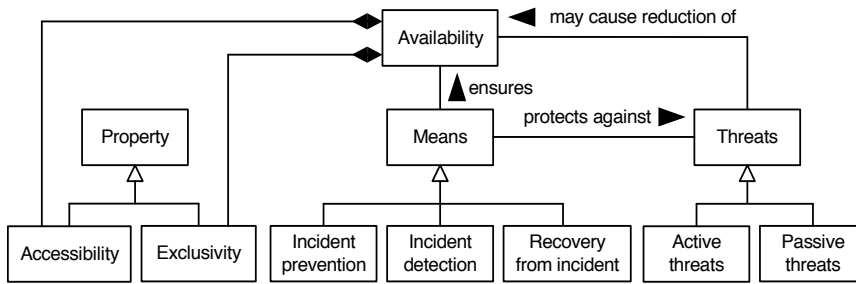


Fig. 1. Conceptual model for service availability

Sect. II provides a brief introduction to the enhanced service availability concept. Sect. III presents an overview of Presence service orchestration in IMS with external Web 2.0 Presence sources. Sect. IV exemplifies how the enhanced service availability concept can be applied for a federated Presence implemented across IMS and external servers.

## II. SERVICE AVAILABILITY

The setting for the enhanced service availability concept is derived from the fields of dependability and security. The definition of availability used as a basis for the enhanced service availability concept is: The property of being accessible and usable on demand by an authorized entity [6], [7]. This definition captures the integral part of securing availability by ensuring access to authorised users while also addressing the aspect of a service being usable in addition to the traditional aspect of readiness for correct service.

The notion of service availability has been further refined using this definition as a basis, to include addressing the *exclusivity* aspect of ensuring that a service is provided to the authorized users *only* [4]. This aspect is important because a system must know how many users are expected to access a service at a given time as well as how long the users are expected to access the service. If the means to ensure that authorized users *only* are accessing a service is too weak, and unauthorized users are able to access a service, the service availability for authorized users may be affected.

The conceptual model of dependability consists of three parts: the attributes of, the threats to and the means by which dependability is attained [8] and provides a basis for the service availability conceptual model as motivated in [9]. In order to classify threats to availability and means to achieve availability in a security setting, we are also motivated by the approach used in the security field of risk analysis as in [10].

This is because incidents resulting in loss of availability do not necessarily escalate into faults and therefore classification of means in terms of faults may become insufficient for availability analysis. An example is the hijacking of user sessions by an attacker or group of attackers, preventing the authorised user or group of users from accessing the service. This incident results in loss of service availability for a set of users, without incurring a fault in the system. An unwanted incident is defined in [11] as an incident such as

loss of confidentiality, integrity and/or availability. A fault is an example of an unwanted incident. The service availability conceptual model therefore classifies the means to achieve availability in terms of countering unwanted incidents.

Services can exist in numerous degraded but operational/usable/functional states between up and down or correct and incorrect. For example, an online newspaper may behave erratically with slow response times for displaying articles browsed without going down or becoming completely unavailable. This means that a more fine grained measure of availability is needed than pure up or down.

The enhanced notion of service availability encompasses both exclusivity, the property of being able to ensure access to authorised users only, and accessibility, the property of being at hand and useable when needed. Exclusivity involves ensuring that unauthorised users cannot interrupt, hijack, or prevent the authorised users from accessing a service. The focus is on preventing the denial of legitimate access to systems and services by prohibiting unauthorised users from interrupting, or preventing authorised users from accessing services. The aim is to ensure access to users while keeping unauthorised users out.

Accessibility is defined as the quality of being at hand and usable when needed. We divide accessibility properties into three major areas: timeliness, correctness and usability. Timeliness is the ability of a service to perform its required functions and provide its required responses within specified time limits. Usability is concerned with the user's perception of the service, and the ease of use of the service. The measure of correctness of a service may differ widely between different kinds of services. These considerations motivate a notion of service degradation [12]. Service degradation can be defined as reduction of service accessibility.

In summary, the overall conceptual model can be depicted as in Fig. 1 (illustrated in UML 2.x format [13]). Availability is affected by means and threats. Means can ensure availability by protecting against threats. Threats may lead to unwanted incidents which may cause reduction of availability.

By means to ensure availability we address protection of the service from incidents leading to a loss of availability. We have categorized the means into i) incident prevention: how to prevent incidents causing loss of availability (e.g. access

control, integrity protection ensuring graceful degradation); ii) incident detection: how to detect incidents leading to loss of availability (e.g. traffic inspection, audit logs); and, iii) recovery from incident: the means to recover after an incident has lead to a loss of availability (e.g. system adaptability, robustness, maintainability, redundancy).

Threats may originate on the inside (inside attackers) or the outside (outside attackers) of the system. The impact of threats varies with the nature of the threats; some threats may result in degradation of the service, others in complete loss of service. For the full motivation and explanation of the model, see [9].

Based on the conceptual model, the availability of a service can be analyzed with respect to exclusivity and accessibility aspects. On an abstract level, a mathematical representation can be given as follows; Let  $A$  denote a service with an availability property for a user group  $U$ , and let  $X$  denote the availability metric for service  $A$ . We represent  $X = (x_1, \dots, x_n)$  as an  $n$ -tuple where  $x_i$  is a measure of an aspect of availability. These include behavioural, preventive and correctness aspects. By this we mean that  $x_i$  describes requirements for a particular availability aspect. The minimum requirement for each  $x_i$  must be satisfied in order to fulfil the total availability requirement  $X$ . Using the conceptual model this idea can be refined as follows: We represent  $X$  as a tuple  $X = (X_1, X_2)$  where  $X_1$  measures the exclusivity properties, and  $X_2$  measures the accessibility properties. Essentially, the aim is to describe the degree of accessibility and exclusivity that is sufficient for the user to be able to activate and use the service. The purpose of service availability metrics is to measure how well service availability requirements have been met.

### III. PRESENCE ORCHESTRATION: IMS AND WEB 2.0

IMS has been promoted by several international bodies as a future platform for providing rich multimedia services. It is access agnostic in the sense that services could be provided over any access type and to any device, as long as the device is capable of supporting the proper client behaviour.

#### A. IMS principles

A layered architecture has been applied for defining IMS, as depicted in Fig. 2. In the core part, we find common session control and common user data. In IMS terms these are referred to as Call Session Control Functions (CSCF) and Home Subscriber Server (HSS), respectively. There are three types of CSCF supporting roaming users, interconnecting between domains and emergency sessions, but these are not depicted individually in Fig. 2.

The HSS stores user identities and user profiles. Each user profile contains at least one service profile with customised information pertaining to which applications are to be invoked, in which order and how services are to be executed. This capability for user-customised service mashups implies that IMS can be considered to be a user-oriented architecture. This is an important factor for combining IMS services with existing 3rd party Web 2.0 services.

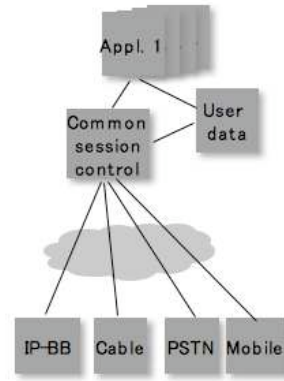


Fig. 2. Layered architecture of IMS (Note that clients are not illustrated)

A range of applications can reside in the common IMS core. Examples of application types are group list managers, Centrex, location, handover support.

#### B. Presence Service

One application that we analyse with respect to the service availability model presented above, is the Presence service. A Presence service is a system that accepts and stores Presence information from information providers, called Presentities and distributes this information to interested parties, called Watchers as illustrated in Fig. 3. Presence sources are nodes reporting Presence information. Examples of Presence sources are clients in handsets, mail/calendar servers, network elements and indications given through web portals. Presence information, as defined by [14], conveys the ability and willingness of a user to communicate across a set of devices. IMS uses SIP-based specifications to provide Presence functionality.

The Presence Server (PS) is located in the presentity's home network. It commonly includes both logic and data storage. Key capabilities include collecting, composing and filtering Presence information.

As depicted in Fig. 3, it is not only user status that can be reported through this mechanism. Other examples include changes in stock prices and programme information from radio stations. The different Presence items may be updated from different sources, involving different service providers (companies and individuals). This implies that the different providers must interact, and cooperate. An immediate case is the incorporation of Facebook news feeds ([www.facebook.com](http://www.facebook.com)), Flickr ([www.flickr.com](http://www.flickr.com)) photo album updates and del.icio.us ([del.icio.us/about/](http://del.icio.us/about/)) bookmark updates into an advanced IMS-based buddy list handler. This scenario would also include Presence items internal to IMS such as buddies with IMS subscriptions. Fig. 5 gives a logical view of the Presence entities in such a system from an IMS perspective. These particular third party services were chosen because they are popular Web 2.0

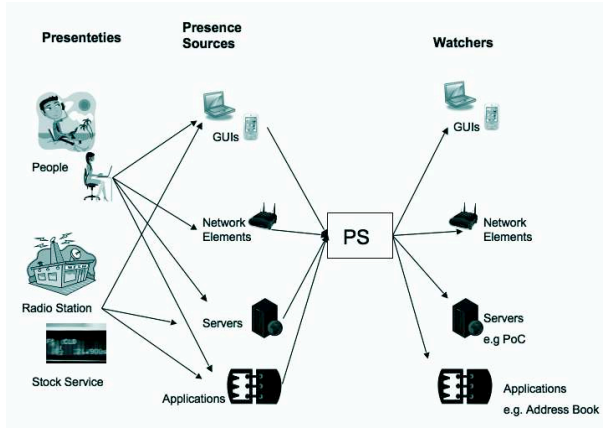


Fig. 3. Presence Server conveying information about users to defined Watchers

services. In addition these services expose HTTP interfaces, making it possible for entities in the IMS service domain to consume them. The various Presence agents presented here are described in 3GPP's Presence specification [15].

Subscription authorization policy is also provided by the PS. This controls which Watchers are authorised to view Presence information. Level of detail may differ for the different Presence items. For example, for certain buddies (for example the user's minor children), the user is allowed to see their locations, while for others (for example business contacts) location may be kept private. Again, these rights may vary over time (day, night, work, vacation etc.)

Both the accessibility and the exclusivity aspects of service availability discussed in Sect. II are related to the Presence item list. For each of the different groups in the item list, the degree of exclusivity and accessibility achieved may differ. For example, during leisure time, colleagues may not want updates on a user's whereabouts. On the other hand, this information becomes relevant during office hours.

A time stamp field in the message format may be used as a rudimentary security mechanism to prevent replay attacks (launched for example to capture user credentials for unauthorized access to a service). For example, tuples whose time stamp are older than the time stamp of the most recently received Presence document should be discarded.

#### IV. CASE – ENHANCED PRESENCE SERVICE AVAILABILITY

The enhanced Presence service based on IMS is assumed to combine a set of different features, such as location information as well as advertisements. Multiple levels of availability are particularly relevant when buddies are subscribers of other domains.

For such a multi-provider case, a block diagram could be constructed for each of the items in the list. For this configuration, certain details regarding implementation of Presence services in other domains would likely not be

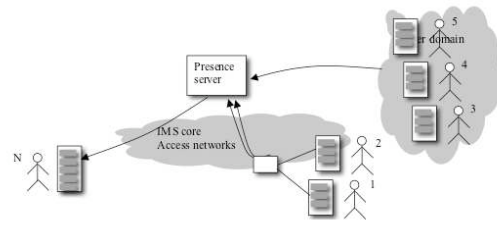


Fig. 4. User with buddies in different provider domains

available. This implies that accessibility levels would then be part of the Service Level Agreements (SLAs) established between providers. It might also be more suitable to state different levels towards user  $N$  for the different user groups in this configuration. However, this will probably depend on the SLA terms. An index  $j$  indicates the group of buddies:

$$A_j = \frac{\sum_i (A_{ij} \cdot \alpha_{ij})}{\sum_i \alpha_{ij}}, \text{ where } \alpha_{ij} \text{ is the level of importance associated with user } i, \text{ and } A_{ij} \text{ is accessibility of Presence information for each user } i \text{ in group } j.$$

Fig 6 illustrates how availability of external Presence services can be modelled from the system depicted in Fig 5 with such a block diagram. Note that only the buddy list service from each provider is modelled Facebook (fb) friends, Flickr (fl) contacts and del.icio.us (dl) networks. Fig 5 indicates that other service modules are offered from these external providers but as they are loosely coupled it is possible to choose only a subset of services at any given time. It is also possible to model the availability of all the services.

In some cases, there may not be any pre-established SLA between the providers. It then becomes a business risk evaluation whether a service provider wants to state any performance levels in the service description to user  $N$ . Potentially, there may be differentiated levels for the different groups,  $j$ .

Level of importance related to a user,  $\alpha_{ij}$ , may vary depending on the role of the user,  $N$ . For example, during working hours, it is more important to follow work colleagues, for example, involved in the same project, than outside working hours. As projects come and go, the colleagues involved will also vary, requiring that this information is easily updated frequently.

##### A. Parameters Included in Enhanced Presence

The parameters given for a Presence item may include nickname, mode, location, as well as others. Typically, location could be given with different levels of granularity. In some cases, e.g. during roaming, the location may also be unknown. On the other hand, there are certain usages where location is considered prioritised.

For user  $N$ , different levels of importance could then be

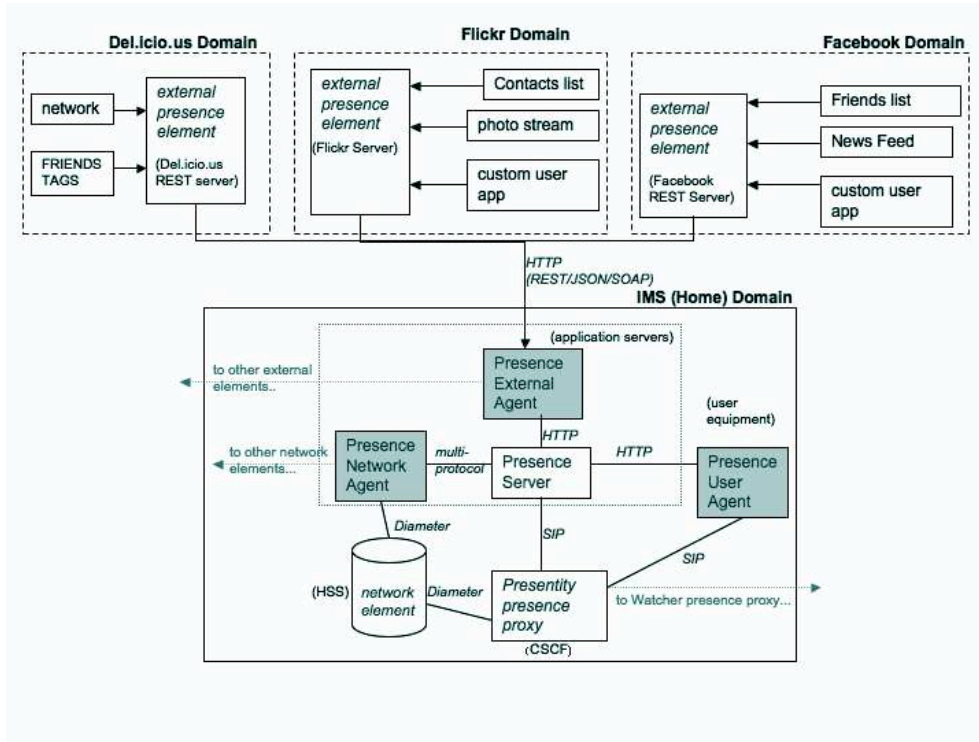


Fig. 5. External Web 2.0 Presence Items

attached to the different parameters. These importance weights may also differ for the different Presence items.

As these parameters may be pushed or pulled from different sources, different response or delivery times would result. In some respects, this is similar to the design of a web page consisting of a set of objects. In order to improve QoS and network performance, Presence parameters should be delivered in appropriate sub-groups. That is, waiting for the last Presence parameters before sending an update to the user's buddy list would likely result in delayed response times and degraded Quality of Experience.

This sub-grouping is particularly a critical aspect when information is collected from vastly different sources, some residing in semi-real-time environments while others within best-effort environment. In effect, this balances the different aspects of availability (correctness, timeliness and usability) as described in Sect. II.

### B. Use of Presence Items for Other Purposes

Having configured means for controlling Presence items on a terminal, one could utilise this for other purposes as well, such as advertising and time-related special offers. One such implementation has been tested in a real user environment in Finland, in the SmartRotuuaari project [16]. The service implemented included highly personalized direct marketing to customer's mobile phones.

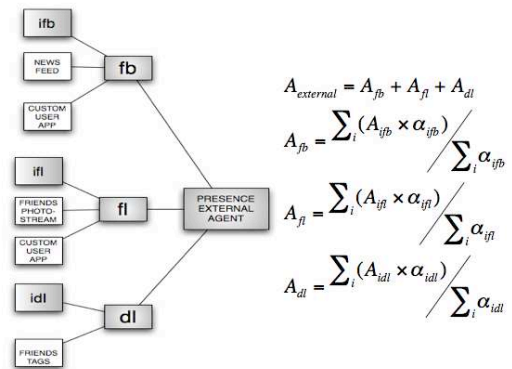


Fig. 6. Block Diagram for External Presence

When users accept such commercial activities, a service provider would then likely have an agreement with other companies for delivering the advertisements/offers. Then, there will also be requirements on accessibility for providing this feature. In the spirit of web 2.0, users could be given the ability to give reviews on advertised products they have tried. Watchers could then choose only to filter which advertisements they accept (for example: "only with a rating above 610" or "only products recommended by my buddies" or "only

products reviewed by more than 500 users”). Again, block diagrams could be made assisting the accessibility evaluation.

### C. Exclusivity

Overall exclusivity could be analysed in a similar manner as for accessibility. That is, either looking at the overall average or looking at each Presence item individually. However, the analysis approaches would likely differ as there are often other types of threats affecting the exclusivity aspect.

The aim with respect to the exclusivity aspect is to ensure that authorised users only should have access to Presence items that they subscribe to. Allowing unauthorised users to access Presence items may have a negative effect on accessibility, but also, may be in violation of privacy directives in the jurisdiction.

Commonly, for the Presence service there will be stricter requirements regarding exclusivity than for accessibility. This is mainly due to privacy aspects, avoiding any third party being able to follow a user’s actions. However, it is also important to ensure that the authorised users are not prevented or interrupted from accessing the Presence items. Activity initiated by unauthorised users can adversely affect the accessibility aspects.

### D. Threats and Corresponding Means

Considering the different network elements, protocols, data, etc. that are involved, various threats would be relevant. Examples of threat agents are malicious users, rogue service providers and unauthorised users masquerading as authorised users. It is important to note here that in the Web 2.0 context, service providers may also be individual users of the system. The vulnerability of the Presence service to distributed denial of service attacks should be evaluated.

Hence, combinations of means to address the threats, covering protection, detection and recovery will be required. Essentially, preventative mechanisms will try to eliminate the possibility of attacks by threat agents or to enable the Presence service nodes to be able to endure attacks without denying service access to authorised users. Detection and recovery will involve detecting attacks on nodes of the service, or against specific users, and responding immediately to restrict impact.

ETSI TISPAN has developed a threat vulnerability and risk assessment (eTVRA) method and tool that may be used to analyse the Presence service [17]. Using the eTVRA method and tool, the threats to availability of the Presence service can be analysed and a set of recommended countermeasures can be identified that when implemented will reduce the overall risk.

## V. CONCLUSION

Recognising the business of delivering dynamic information from widely different sources to individual users, orchestrating services become a key aspect. One example is given in this paper showing how Presence information can be collated from various internet sources and blended in IMS. Considering the federated nature of the Presence service, a range of challenging

aspects need to be addressed, including differentiation of Presence items and parameters for an item while also handling multiple sources of Presence data.

This paper outlines a service availability model, applied for the Presence service in an IMS context. A key point is to include both the accessibility and the exclusivity aspects of the service availability measure. Hence, only the authorized users should be ensured access to the service, and with the proper service levels. So far, it seems that exclusivity is an aspect of availability that has rarely been included in the literature. However, the concept presented here shows where exclusivity fits in with an enhanced notion of service availability.

The enhanced notion seems even more important when considering collaboration between providers, and also between different roles of the same user. Proper service orchestration, then becomes even more important and more challenging.

## REFERENCES

- [1] S. M. Ross, *Introduction to probability models*, 6th ed. Academic Press, 1997.
- [2] P. Enriquez, A. B. Brown, and D. A. Patterson, “Lessons from the PSTN for dependable computing,” Workshop on Self-Healing, Adaptive and self-MANaged Systems (SHAMAN), 2002.
- [3] D. Clark, W. Lehr, and I. Liu, “Provisioning for bursty Internet traffic: Implications for industry and Internet structure,” MIT ITC Workshop on Internet Quality of Service, 1999.
- [4] J. E. Y. Rossebø, M. S. Lund, K. E. Husa, and A. Refsdal, “A conceptual model for service availability,” *Quality of Protection: Security Measurements and Metrics*, vol. 23, 2006.
- [5] J. E. Y. Rossebø, A. Pearce, and T. Jensen, “On understanding availability of services based on ip multimedia subsystem,” in *18th ITC Specialist Seminar – Quality of Experience*, to appear.
- [6] *ISO 7498-2, Information Processing Systems – Interconnection Reference Model – Part 2: Security Architecture*, International Standards Organization, 1989.
- [7] *ISO/IEC 13335, Information technology – Security techniques – Guidelines for the management of IT security*, International Standards Organization, 2001.
- [8] A. Avižienis, J.-C. Laprie, and B. Randell, “Fundamental concepts of dependability,” in *Third Information Survivability Workshop (ISW)*, 2000.
- [9] J. E. Y. Rossebø, M. S. Lund, K. E. Husa, and A. Refsdal, “A conceptual model for service availability,” Research report 337, Department of Informatics, University of Oslo, 2006.
- [10] F. den Braber, M. S. Lund, K. Stølen, and F. Vraalsen, “Integrating security in the development process with UML,” in *Encyclopedia of Information Science and Technology*. Idea Group, 2005, pp. 1560–1566.
- [11] *AS/NZS 4360:1999, Risk Management*, Standards Australia, 1999.
- [12] J. F. Meyer, “Performability evaluations: Where it is and what lies ahead,” in *Proc. of the International Computer Performance and Dependability Symposium*. IEEE Computer Society, 1995, pp. 334–343.
- [13] *UML 2.0 Superstructure Specification, formal/05-07-04*, Object Management Group, 2006.
- [14] J. Rosenberg, *A Presence Event Package for the Session Initiation Protocol SIP*, RFC 3856, 2004.
- [15] *Presence Service; Architecture and functional description, Stage 2*, Third Generation Partnership Project, Technical Specification Universal Mobile Telecommunications System (UMTS), 3GPP, TS 23.141 V 7.3.0 (2007-10), Release 7, 2007.
- [16] T. Ojala, J. Korhonen, M. Aittola, M. Ollila, T. Koivumäki, J. Tähtinen, and H. Karjalainen, “SmartRotuaari context-aware mobile multimedia services,” in *Proc. 2nd International Conference on Mobile and Ubiquitous Multimedia*. Washington, DC, USA: IEEE Computer Society, 2003, pp. 9–18.
- [17] J. E. Y. Rossebø, S. Cadzow, and P. Sijben, “eTVRA, a threat, vulnerability and risk assessment method and tool for eEurope,” in *ARES*. IEEE Computer Society, 2007, pp. 925–933.