# A Model for Enhancing Presence Handling in Instant Messaging

by

**Rudi Victor**

# A Model for Enhancing Presence Handling in Instant Messaging

by

**Rudi Victor**

## Dissertation

submitted in fulfillment
of the requirements
for the degree

## Magister Technologiae

in

## Information Technology

in the

## School of Information and Communication Technology

of the

## Nelson Mandela Metropolitan University

**Promoter: Prof. Reinhardt A. Botha**

**Co-promoter: Mr. Andrew Rutherford**

January 2009

# Declaration

I, Rudi Victor, hereby declare that:

- The work in this dissertation is my own.
- All sources used or referred to have been documented and recognized.
- This dissertation has not previously been submitted in full or partial fulfillment of the requirements for an equivalent or higher qualification at any other recognized education institute.

---

Rudi Victor

# Acknowledgements

I would like to extend my gratitude and appreciation to:

- *God the Father*, who knows all our secrets but never violates our privacy.
- My *family and friends*, for all their support, encouragement and understanding.
- My promoter, *Professor Reinhardt Botha*, and co-promotor, *Andrew Rutherford* for their motivation and guidance throughout. Their contributions are much appreciated.
- *Mrs. Adéle Botha* for the much appreciated proofreading of this dissertation.
- The *NRF*, who through their financial support allowed me to pursue this research.

# Abstract

Instant Messaging (IM) is becoming increasingly popular in social as well as workplace environments. In fact, many employees use the same IM client to communicate with both colleagues and social contacts. Thus, there are valid concerns about the impact of IM on employee productivity.

One of the major advantages of IM over other workplace communication tools such as e-mail and the telephone is the implementation of presence information. In particular, presence awareness is used to determine the availability and willingness of a contact to engage in communication.

A current problem with IM is the one-for-all approach to presence: all contacts receive the same set of presence information. However, presence is rooted in social psychology where it is known that the awareness of another person changes the behavior of oneself. Therefore the identity of a contact affects the availability and willingness directed towards that contact. In order for presence information to be provided to contacts, it must be represented in some type of data format.

The Internet Engineering Task Force (IETF) has done much work in standardizing IM and presence systems. In particular their data format for presence describes a rich set of presence information including, but not limited to, location, activity, awareness, and mood information. Such information may be sensitive and access to it needs to be controlled to ensure privacy. As with access control policies, managing the information as the number of contacts increases becomes cumbersome and complex.

This dissertation draws on the theoretical foundations of presence, current standards in the domain of IM, and lessons from access control to present an enhanced presence handling model for IM. The model is developed in stages, with each stage providing a specific improvement.

The first stage of the model is grounded on the current work of the IETF.

As such it distributes presence on a per-watcher basis.

In the second stage of the model watchers fulfill a specific role and based on this role they receive only the entrusted presence information. In practice, it implies that a "friend" may get more (or less) information than a "colleague".

The third stage of the model introduces the concept of availability profiles by drawing on social awareness principles. Availability profiles add the ability to transform presence and change the presentity's behavior to incoming messages according to the provided presence information.

Finally the dissertation reports on the development of the RoBIM (Role-Based Instant Messenger) prototype. RoBIM is a standards-based IM system that conforms to the IETF SIMPLE protocol and provides various standard IM features. Here, RoBIM serves as a proof-of-concept for the proposed model.

This study contributed to the domain of IM and presence by addressing some of the current presence handling issues. Most importantly, the proposed model takes into account the interpersonal effects of individualizing presence information for different contacts. Thus, the model challenges conventional thought and implementation of presence in IM.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Instant Messaging (IM) is an exciting new computer-mediated communication (CMC) tool which has only been around for about a decade. The rapid increase in Internet-based communication has led to IM becoming very popular in a relatively short amount of time. IM has gained popularity in the workplace as well, and this trend is likely to continue. The information technology research group Gartner stated that by 2010, 90% of employees that use e-mail would use IM as well (Smith & Lundy, 2007).

Currently e-mail is regarded as the *de facto* medium in online communication because it is easy to use and cost-effective. IM is often associated with e-mail because both are text-based CMC tools. However, IM is used very differently than e-mail because of its immediate and responsive nature. A deeper analysis, as done in chapter 2, shows that IM communication executes similar to a traditional face-to-face conversation.

In order to be reachable for traditional face-to-face communication, communicators need to be physically present. However, the Internet has created a way for people to be virtually present. The Internet is often referred to as a virtual community, where people interact and communicate based on whether they are online or offline. In the real-world being online translates to being physically present or reachable. All current IM systems provide a means to signal a person's basic presence state as either online or offline. Information regarding such *awareness* of another person is referred to as presence information. Even the identity of another person can be considered presence information.

The purpose of presence information is to provide an informed view of

another person's current state. This informed view helps the initiator of communication infer when a successful communication attempt is more likely (i.e. when the communication partner's presence state is "online").

One of the main advantages of IM over face-to-face communication is the ability to be much more reachable. By using IM over the Internet there is no geographic constraint as there is with face-to-face communication. Therefore, a person can be online and reachable to numerous people at the same time. However, being present and reachable to all is not ideal as it can lead to unwanted communication and interruption.

IM provides an excellent platform for investigating presence because it normally provides more presence information than a simple online/offline state. Commonly implemented presence states such as *"Busy"* augments the basic state of being online to also imply a degree of availability and willingness to communicate. Next consider some key terms used in this dissertation.

## 1.1 Definition of Terms

In order to follow the discussions with regards to communication and the people who partake in communication, the following terms are defined:

- presentity: an entity or person having presence information;
- watcher: an entity or person having interest in the presence information of a presentity; and
- presence: the current state of a presentity.

"Watcher" and "presentity" are logical roles. As such, they can describe the same IM user depending on whether he/she is the provider or recipient of presence information. For the purposes of this dissertation, the initiator of communication is always referred to as the watcher. The reason being the watcher is the user who receives presence information and this information is used to decide whether to engage in communication. By the same token, the recipient of communication is referred to as the presentity because he/she provides the initiator (i.e. watcher) with presence information.

These terms are defined in more detail in RFC 2778, a basic model for IM and presence, as specified by the Internet Engineering Task Force (IETF)

(Day, Rosenberg, & Sugano, 2000). The model will be presented in more detail in chapter 4.

## 1.2 Motivation For This Study

The research undertaken for this study was motivated by the following realizations.

### The Popularity of IM

IM is an increasingly popular communication medium and can be used for many purposes. Additionally, the popularity of IM has followed e-mail into the workplace. There are conflicting opinions as to the affect of IM on employee productivity. Some believe it increases productivity (Licari, 2005), while others suggest that IM increases interruptions in an unfair way (Czerwinski, Cutrell, & Horvitz, 2000; Cameron & Webster, 2005). Thus business managers are sceptical about the benefits of IM versus the risk of reduced productivity. The popularity of IM in the workplace can largely be attributed to employees themselves installing IM clients at work (Vos, Hofte, & Poot, 2004).

On the social front IM is very popular as a tool for networking, especially among teens (Grinter & Palen, 2002). Informal conversations and the scheduling of "get-togethers" are common social uses for IM. Overall the increasing popularity of IM increases the number of contacts per user. The realization of the increasing popularity of IM, especially since it employs presence awareness information, is an important motivator for this study.

### Privacy Concerns

IM is used for personal as well as business purposes and these two application contexts can interfere and clash concerning the controls and degree of intrusion, privacy and reachability. The basic desire in terms of online privacy can be defined as having the ability to be visible to those you choose and invisible to the rest (Palen & Dourish, 2003).

The Internet allows people to be very public and reachable. If an e-mail address unwillingly falls into the hands of an advertisement mailing list it could lead to significant amounts of unwanted e-mails. Consequently

communication tools need to specifically provide mechanisms to preserve user privacy while allowing public reachability.

The rich set of presence information specified by the IETF also contributes to a greater need for administering and enforcing control over presence information. As the specifications of the IETF become more widely adopted, the prevailing question is "who gets to see which presence information, and when?" (Sugano et al., 2004).

#### The Current Ineffective Presence Handling Model in IM

People want to represent themselves in different ways to their IM contacts within a particular context. Currently if a presentity wants to project a status of "Busy" to friends, the presentity automatically projects a status of "Busy" to colleagues as well. This can lead to confusion and misunderstandings as stated in the following section.

## 1.3 Problem Statement

The use of presence in IM is very basic and not particularly effective. The first problem is that all watchers are treated identically and are provided with the same presence information. Situations can easily exist where the presentity wants to treat watchers differently. A simple example would be to present a colleague with an *Available* presence state while presenting a presence state of *Busy* to a friend.

Secondly, presence information needs to be limited according to the trust awarded to the watcher. Presence information can be of sensitive nature which might only be appropriate for a limited set of watchers. A colleague might need to know in detail where a meeting is taking place, while a friend might only need to know that his contact is at work. This issue becomes more prominent if watchers are clients whom the presentity is not familiar with outside of cyberspace. A presentity may want to treat such watchers with very limited trust.

Even if the first two problems are resolved a third problem is anticipated. This problem has to do with the management of presence information which is mostly done by the presentity. The cost-of-effort to keep presence information

tailored for each watcher becomes overwhelming, including keeping all that information consistent based on the presentity's most current context.

To summarize the problems in a single succinct statement: *The current handling of presence information in IM is ineffective and has various limitations that need to be addressed.* These limitations can be addressed by meeting the objectives discussed next.

## 1.4 Objectives

The primary objective of this research is to create an enhanced presence handling model applicable to IM. The model will allow presentities to control how they present their presence information to watchers. In order to achieve the primary objective a number of secondary objectives need to be addressed. These objectives are to:

- allow fine-grained control over what presence information is given to which watcher and when;
- manage presence information in an efficient and scalable manner; and
- provide watchers with individualized presence awareness and allow the presentity to demonstrate awareness accordingly.

This research will set out to achieve these objectives in a methodical way.

## 1.5 Methodology

This dissertation reports on research conducted in a systematic manner to achieve the stated objectives. It is worthwhile to note that this research forms part of a larger research initiative within which multi-channel communication and presence is being investigated (Rutherford, 2008). Nevertheless the research presented here indeed form a coherent goal-oriented unit.

This research aims to use appropriate techniques to address the issues of the research of problem understanding, model development and evaluation. As the research sets out to develop a *model*, design-science will form the primary research paradigm used. Hevner, March, Park, and Ram (2004) describe design science as follows: "in the design science paradigm, knowledge

and understanding of a problem and its solution are achieved in the building and application of the design artifact". They further propose seven guidelines for design science. The next paragraphs describe why design science is deemed appropriate by stating how each guideline applies to this research.

- *Design as an Artifact.* Design science must produce a viable artifact in the form of a construct, a model, a method, or an instantiation. In this research the aim is to produce a model for presence handling in IM.

- *Problem Relevance.* The objective of design science is to develop technology-based solutions to important and relevant business problems. It is argued that the proposed artifact would aid in making communication inside (and outside) business more efficient by reducing the interruptions to workers. Since the adage "time is money" survived, it is deemed that such productivity-oriented measures are indeed relevant to business.

- *Design Evaluation.* The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods. While it is not the intention of this research to provide empirical proof of performance improvements, proof-of-concept prototyping and scenario-based reasoning does demonstrate the appropriateness of the model.

- *Research Contributions.* Effective design science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies. In this research the model developed provides a clear and verifiable contribution. In addition the proof-of-concept prototype provides an extendable test-bed for future research.

- *Research Rigor.* Design science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact. This research aims to minimize ambiguity by formalizing the model where applicable. In addition various graphical representations will aim to enhance clarity. Furthermore, the development of the prototype feeds back into the development of the model and as such enhances the quality of the resultant model.

- *Design as a Search Process.* The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in

the problem environment. The model arrived at in this research is the culmination of research from several domains.

- *Communication of Research.* Design science research must be presented effectively both to technology-oriented as well as management-oriented audiences. The production of this dissertation in itself contributes to this objective. In addition it is foreseen that one or more relevant publications will result.

In the conclusion chapter these guidelines will be revisited to confirm adherence.

## 1.6 Layout of the Dissertation

A diagram depicting the layout of this dissertation is shown in figure 1.1.

The first five chapters provide important background information regarding IM, presence and the implementation of presence in IM. The problems from the problem statement are explored with greater detail and focus as the chapters progress. Chapter 2 focuses on IM and how it is currently used. Chapter 3 considers presence from its roots in social psychology while chapter 4 is more concerned with standards and specifications of presence. Access control is studied in chapter 5 in line with the presence processing specifications in chapter 4.

Chapter 6 then introduces an enhanced presence handling model for IM. Development of the model will be done in three stages, each progression building on the previous.

Reporting on the implementation of a software prototype, based on the proposed model, is done in chapter 7. The prototype will be used to demonstrate the feasibility of the model and will be used in scenario-based evaluation.

Chapter 8 will end the dissertation with a synopsis of the research. It will briefly restate the research problem, the proposed solution and how this study has conformed to the intended research paradigm of design-science.

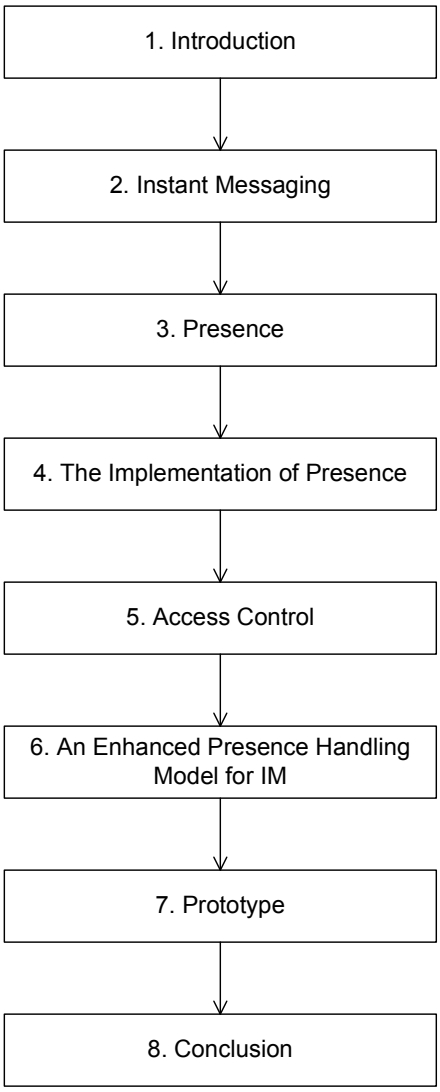


Figure 1.1: Layout of the dissertation

# Chapter 2

# Instant Messaging

The appearance of the Internet has enabled us to connect and communicate across barriers of time and space. The Internet is often referred to as the global online community and therefore a place where people can be *present*. As in a real-world community, members desire the ability to feel connected and to interact with one another. Computer-mediated communication (CMC) systems such as electronic conferencing systems, e-mail and voice mail systems enable and encourage interactions within the online community. However, traditionally none of these systems provide a way to be aware of the *presence* of others. It is in this space that another CMC system, namely Instant Messaging (IM), has flourished.

IM is essentially a text-based medium that allows a conversational style of communication. These conversations execute similar to real-life face-to-face interactions and are usually between two people. A key feature of IM is its use of presence awareness. This allows people to be aware of one another for various purposes. In its most basic form presence awareness conveys whether a potential communication partner is connected or present. However, presence awareness can be extended to include various types of social and contextual information. Presence awareness is not limited to IM and can therefore be implemented in other communication technologies as well. Due to the importance of presence with regards to this dissertation, presence will be covered in more detail in chapter 3.

Despite the increasing popularity of IM, there are still many issues and tensions surrounding IM. The penetration of IM into the workplace has accentuated issues such as standardization, interoperability, security and privacy.

The remainder of this chapter will explore the concept of IM and investigate its use of presence information.

## 2.1 Instant Messaging Defined

There are several definitions of IM in use. Although the various definitions largely overlap, there are some distinct differences. Therefore, it is best to keep in mind that IM is an evolving technology and as such no single definition should solely be depended upon.

Greene and O'Mahony (2004) define IM as "synchronous electronic communication where two or more users exchange textual or voice messages via electronic devices". The authors further state that "IM combines bi-directional communication with information about the status of each user on the network". This definition is very general and includes text-based communication and voice messages. It should be noted that this definition includes status information *about* users, indicating the importance of presence information.

Hung, Huang, Yen, and Chang (2007) refer to IM as "a synchronous computer-mediated messaging system defined as a type of communication service that enables the creation of a private chat room with other users". The term *private chat room* provides a good analogy of IM communication as an isolated room where users can *chat* to each other.

Avrahami and Hudson (2006) state that "IM facilitates one-on-one communication between a user and his/her list of contacts in a text based-format". Furthermore, Avrahami and Hudson (2006) describe presence as one of the most important *features* of IM. For the purposes of this dissertation, presence lies outside the definition of IM but forms an integral component or feature. Thus with respect to IM, presence is the awareness of other users or contacts on the IM system (Greene & O'Mahony, 2004).

All mentioned definitions have the following in common: *text-based near real-time communication between two or more users.* However, some include presence awareness in the definition while others only describe presence awareness as an important feature of IM.

In this dissertation presence awareness is viewed as the awareness of other users within a communication network. Presence can therefore be used with


Figure 2.1: General AIM user interface

other communication technologies and is not limited to IM. The content and scope of presence awareness information differs widely from implementation to implementation and will not be explored in this chapter.

For the purposes of this dissertation the author will subscribe to the definition of IM provided by Avrahami and Hudson (2006): *"Instant messaging (IM) is a near-synchronous communication technology that facilitates text-based conversation between two or more participants"*. In the next section the current form of IM will be examined by means of a popular IM client. Together with presence awareness, the definition of IM will be illustrated through an overview of the core features of a typical IM system.

## 2.2 The Current Form of IM

In this section AOL's Instant Messenger (AIM) will be used to illustrate the basic features of an IM client as it is presented to users. AIM has the largest user base of all current IM clients worldwide. Also, AOL has acquired a patent for the term "Instant Messenger" which inspires confidence to consider AOL's AIM as an IM client archetype (Kontzer, 2002). Although implementations differ, there are three essential features common to all IM clients:

- a contact list;
- presence information; and
- text-based conversation.

The general user interface of AIM is shown in figure 2.1. It illustrates two of the three essential IM client features, namely a *contact list* and *presence information*. Additionally, the general user interface serves as a platform from which the third essential feature, *text-based conversation*, is initiated.

### 2.2.1 A Contact List

An IM user initiates communication with other users by using a contact list, also called a "buddy list". It is often the case that IM users know their contacts personally, as the word "buddy" implies (Smale & Greenberg, 2005). Each user manages his/her own list of contacts by adding, removing or even blocking contacts.

To preserve privacy, consent is normally requested to add a contact to a contact list. The potential contact then accepts or denies the request. Only after obtaining approval can a user engage in communication with a particular contact. The storage of users' contact lists is normally centralized by keeping them on IM servers. This allows the user to access a single contact list from different instances of an IM client. Therefore an IM user will have access to the same contact list from home and from work.

A typical contact list is shown in figure 2.1. It contains the name of a contact, Jill Jane Hill, as well as her customized presence status information below her name. Visual categories for organizing contacts can be seen in the collapsible "Family", "Co-workers", and "Offline" groupings.


Figure 2.2: AIM status messages

### 2.2.2 Presence Information

In section 2.1 presence awareness was defined as information regarding the awareness of other users of a communication network. In IM this is often implemented as a single keyword or phrase which is associated with an icon. Figure 2.2 shows the different presence status messages provided by AIM. Additionally, a custom presence status message can be defined.

#### The Meaning of Presence States

The implied degree of availability and willingness through the use of commonly implemented presence states are largely subjective. However the actual keywords used to describe presence states provide insight into their meaning. "Available" conveys that a contact is available and ready for conversation of any degree of importance. A presence state of "Busy" means that a contact only wants to communicate when the content of the conversation is important or of a high priority. An "Off-line" presence state simply conveys that the contact is not available or reachable at that time. "Away" does not say much about the willingness of a contact. Instead, it implies that the contact is currently inactive with regards to using the IM client. Consequently, the contact should not be expected to respond immediately.

Every IM user is responsible for maintaining his/her own presence state. This can easily lead to an incorrect presence state due to negligence or forgetfulness of the IM user. It has been shown that IM states are often set incorrectly (Avrahami & Hudson, 2004). To remedy this situation some IM

clients, like AIM, automatically show a presence state of "Away" after a specified period of inactivity. This is done by monitoring the keyboard usage. However, presence information, as it is currently implemented, does not provide any benefit if it is ignored by the user initiating communication.

#### Operational Implications of Presence States

Presence information does not actively prevent communication. It is merely an aid to the IM contact to infer the likelihood of a successful communication attempt. Therefore a contact is just as reachable if his presence state is set to "Available" as when it is set to "Busy".

The usefulness of presence information relies on the initiator of the communication to correctly assess the conveyed presence information. For instance, if a contact has a presence state of "Busy" it may signify that the contact is unwilling to engage in a social or low-importance conversation. Therefore, if this information is ignored or misunderstood by the communication initiator it could result in an unsuccessful communication attempt. On the contact's side it could result in unwanted interruption. This is because all incoming messages are shown in a salient manner on the contacts side, usually in the form of a pop-up chat dialog window.

### 2.2.3 Text-Based Conversation

When a user double-clicks on a contact in the contact list, a dialog window opens. When the user sends an instant message, an identical dialog window appears on the contact's side as well. A dialog window keeps track of all comments made within an IM session. It does so by displaying all instant messages sent by both participants during the IM session. An IM session is similar to a face-to-face conversation, consisting of numerous interactions and having a start and an end. By keeping track of all comments within the IM session, both participants have the ability to reflect on earlier comments.

The dialog window is split into two parts as shown in figure 2.3. The main part displays all instant messages for that IM session or conversation. The second part, on the bottom, is where the user crafts a message before it is sent. Thus, messages are not sent as they are typed which affords the user the ability to compose and edit a message before it is sent.


Figure 2.3: Simple IM conversation using AIM

Most current IM systems, like AIM, allow off-line messages to be sent when a buddy is not connected to the IM system (i.e. presence state of "Off-line"). This is a feature that overlaps with other off-line messaging mediums such as voice-mail and e-mail.

A common feature in IM is the use of emoticons. The word "emoticon" is essentially a merge of "emotion" and "icon". Therefore an emoticon can be defined as an icon or small picture that portrays an emotion or feeling. Because it is difficult and cumbersome to express emotions and feelings in a text-based form, emoticons were introduced to improve the emotional expressiveness of IM (Hung et al., 2007).

### 2.2.4 Extra Features and Services

There are many extra features provided by current IM systems. These include but are not limited to file-transfer, voice-calls, video-conferencing, e-mail integration and current news bulletins. Many of these extra features are supported by AIM as shown at the bottom of the dialog window in figure

2.3.

Most current IM systems provide integrated communication. Voice calls and e-mail are often provided as additional services. Strictly speaking these are not IM services, but IM is increasingly used as a platform for communication convergence. When to use which communication medium is an important question. In the following section IM will be compared to other communication mediums. Each communication medium will be discussed in turn and the chapter will conclude by describing the unique application context of IM.

## 2.3 Instant Messaging and Other Communication Mediums

There are many reasons for choosing a particular communication medium. Such reasons include the message content, medium ease-of-use, expected responsiveness and desired speed of communication. The capabilities of the medium itself is also a major influence.

It can be said that the choice of the medium actually affects the message and therefore becomes part of it (Mcluhan, 1964). In this section the nature of IM communication will be investigated in comparison to other communication mediums.

A commonly used theory for classifying communication mediums is Media Richness Theory (MRT). MRT is defined as a measure of how quickly information can change a recipient's understanding (Trevino, Lengel, & Daft, 1987). Immediate feedback and cues such as gestures and voice tone increase the speed of understanding (Daft & Lengel, 1986). Face-to-face interaction rates very high in terms of information richness and is often regarded as the benchmark against which other communication mediums are measured.

According to studies done by Newberry (2001), text-based communication such as IM and e-mail are ranked on the lower end of the scale of media richness. In the middle range of richness is telephone communication. Previous research has indicated that the richness of IM communication lies somewhere between that of telephone and e-mail communication (Segerstad & Ljungstrand, 2002). The remainder of this section will further investigate

the differences among these communication mediums.

### 2.3.1 Face-to-face

Face-to-face interaction utilizes many types of nonverbal cues to enhance communication. Body language, hand gestures and voice tone contribute to processes such as turn-taking, emphasis of key words and the conveying of emotions (Trevino et al., 1987). The purpose of non-verbal communication is to help the recipient understand and interpret the exchanged information correctly.

Face-to-face communication also provides instantaneous feedback during communication. Therefore, reactions to a matter of high priority or emotional depth can be immediately observed and responded to. Consequently, face-to-face is the preferred medium in most instances as it incorporates non-verbal communication to a high and effective degree.

One of the main drawbacks of face-to-face interaction is the limitation of co-location. Participants of a face-to-face interaction are required to be within a close proximity to one another. This has created a need for alternative communication mediums such as the telephone, e-mail and IM. Among these three, telephone communication is the highest ranked according to richness (Connell, Mendelsohn, Robins, & Canny, 2001).

### 2.3.2 Telephone

Telephone communication share many of the characteristics of face-to-face communication. Both are primarily verbal, synchronous, and execute in a conversational style. Although telephone communication is less rich than face-to-face interaction, it has several advantages.

Unlike face-to-face communication, telephone communication does not suffer from the limitation of co-location. It allows people to be much more reachable than is possible with face-to-face communication alone. Also, in potentially uncomfortable or intimidating situations, telephone communication can be a more appropriate choice. For instance, the possible intimidation associated with addressing a superior at work can be avoided by conducting a telephonic conversation instead of a face-to-face conversation.

The ability to be reached immediately, anywhere is a prolific advantage

of telephone communication. Unfortunately, it also increases interruption of ongoing activities. Furthermore, the lack of cues makes it difficult to politely convey unwillingness to an engaging communication partner. One resolve is to simply ignore incoming calls to avoid interruption. The problem with such a solution is that potentially important calls could be missed. Despite the problems associated with interruption, the telephone has gained significant popularity.

### 2.3.3 E-mail

E-mail is very different from telephone and face-to-face communication. The most notable difference is the use of text as the medium for communication. This affects the style of e-mail communication considerably. E-mail messages are only sent once the sender is satisfied with the message as a whole. This affords the sender the ability to carefully plan and edit a message before it is sent. Therefore, e-mail normally exhibits a much more structured and formal style of communication than the telephone.

E-mail messages are sent using a store-and-forward mechanism which is prone to delays. Messages can take minutes or even hours to reach the recipient depending on server congestion. Therefore, the speed of e-mail communication is much slower than either face-to-face or telephone communication. This can cause frustration when trying to use e-mail in a synchronous manner. However, when a message has to be carefully structured, e-mail is a better choice.

Traditionally, the main advantage of e-mail over face-to-face and telephone communication is that messages can persist long after they have been received. This allows a recipient to reflect on specific information from a previous e-mail message. However, much of everyday communication is of ephemeral nature and would not benefit from being recorded.

Another key advantage is that e-mail, in contrast with face-to-face and telephone communication, does not immediately require attention. Therefore, e-mail communication is considered to be considerably less interruptive to ongoing tasks (O'Kane & Hargie, 2007). The inconspicuous nature of e-mail contributes to its asynchronous style of communication.

The lack of non-verbal cues make e-mail messages much more likely to exhibit ambiguity. E-mail is therefore not suited to situations where ambiguous

content or emotions are involved.

### 2.3.4 Instant Messaging

The definition of Instant Messaging (IM) as used in this dissertation states that IM is text-based and executes in a near-synchronous manner. Although e-mail is also text-based, the nature of IM interaction is much more instant. The speed of IM communication facilitates quick-response interactions similar to face-to-face and telephone communication. Therefore, it would appear that IM fits in somewhere between face-to-face, telephone and e-mail communication.

Similar to e-mail, the text-based nature of IM makes it much more likely to cause misunderstanding and ambiguity (Hung et al., 2007). This is primarily because IM lacks the useful non-verbal communication cues that face-to-face communication employs. It is therefore much harder to express feelings and emotions through IM.

The nature of the content that is being communicated also impacts the communication medium. For instance, if a bank account number is communicated verbally through a face-to-face conversation there is a valid possibility of the recipient making an error. In contrast, IM like all text-based interaction can convey this information much more accurately than voice-based mediums.

Like face-to-face and telephone communication, content is often, but not always, of ephemeral nature and IM conversations generally do not persist for later reflection. The exception is the workplace environment where messages may have to be persisted for auditing requirements. However, this depends greatly on the actual IM system used due to the security implications of persisting such information. In section 2.3.1 it was stated that communicators are limited in face-to-face communication because they have to be in relatively close physical proximation of one another to communicate. IM, like all mediated communication, does not suffer from such a limitation.

IM shares the immediacy and salience of face-to-face and telephone communication. An incoming IM message is immediately shown on the recipient's screen in a pop up dialog. However, the expected responsiveness of IM is less than the expected responsiveness of face-to-face and telephone interactions. Responsiveness can be described as *demonstrated availability* because the

recipient has a choice of how responsive he or she wants to appear (Avrahami & Hudson, 2006). High responsiveness implies a response is expected immediately and *vice versa*.

With IM, the sender of the message is not notified when a message is read by a recipient. This is also known as plausible deniability. In IM, plausible deniability can act as a shield against unwanted conversations. IM messages are normally devoid of the normal formalities associated with telephone calls and get right to the point (Nardi, Whittaker, & Bradner, 2000). Therefore the initial IM message can help the recipient decide whether to reply and engage in textual conversation with the sender. If an IM message is ignored it is not considered as rude as ignoring a face-to-face conversation. The salient recption of IM messages leads to a drawback shared with face-to-face and telephone communication, namely interruption.

As mentioned before, telephone interruption can be avoided by simply not answering the phone. However, IM provides a much more informed solution. Interruption is minimized in IM communication through the use of presence awareness. Presence awareness, as mentioned in section 2.1, is having information about the current state of another user. It is considered to be a distinct advantage over face-to-face and telephone communication. By knowing the state of the contact, an IM user can calculate the likelihood of successful communication *before* the actual attempt is made. If a contact has a presence state of busy the initiator may heed to this indication and not bother the contact for arbitrary chat. The result is less unwanted interruptions.

In comparison to telephone communication, IM is much more cost-effective. One reason being that text-based communication is more efficient to transport across communication infrastructure. Also, most popular IM clients are proprietary freeware which means there is no startup cost. Combined with a better strategy to minimize interruption, it seems that IM is a viable alternative for telephone communication.

### 2.3.5 Comparative Discussion

Every communication technology discussed so far has a unique combination of communication characteristics. Table 2.1 lists some of the characteristics of each communication technology mentioned in this section. It can be seen

Table 2.1: Comparison of communication mediums

| Medium | Style | Timing | Nature |
|---|---|---|---|
| Face-to-face | Rich | Synchronous | Conversational |
| Telephone | Focused | Synchronous | Conversational |
| Instant Messaging | Flexible | Near-Synchronous | Conversational |
| e-Mail | Formal | Asynchronous | Calculated |

from the table that the application contexts of these communication technologies overlap. Therefore it depends on the communicator to choose the most appropriate channel of communication based on the communication needs within a specific situation.

IM has shown that it can be an efficient alternative for quick-response conversations normally executed using face-to-face and telephone communication. It positively adds multi-tasking ability and cost-effectiveness to conversations but brings with it an increase in ambiguity. As mentioned in the previous section, message ambiguity is increased by the limited amount of nonverbal cues available in IM.

In situations that require immediate response or conversational interaction, IM would be a better choice to e-mail because it supports such communication better. However, external factors such as reachability also come into play when a decision is made on which communication technology to use.

Critical mass theory specifies that a technology is only useful once a specific amount of people are using it (Soe and Markus (1993) as cited by Cameron and Webster (2005)). For instance, if IM is deployed within a company, it can only be considered useful if everyone with which that employee has to correspond uses IM. Instant messaging has not yet reached its critical mass while e-mail has established itself as a very popular medium. Furthermore, liaisons with related companies' contacts may also be considered important and it is here that IM still needs to grow. The lack of a standard IM protocol, unlike e-mail, remains a stumbling block towards IM reaching its critical mass (To, Liao, Chiang, Shih, & Chang, 2008).

Instant messaging provides the unique benefit of the embedded use of presence awareness. Unlike most implementations of the other communication technologies discussed, IM allows the initiator of a communication to infer the likelihood of a successful communication attempt before such an

attempt is made. Presence awareness will be covered in greater detail in the next chapter.

In this section IM was shown to be a flexible and conversational technology which can be used as an alternative to other communication technologies. It was also shown to be uniquely applicable in certain situations. The workplace provides many different communication scenarios and the growing use of IM in such scenarios is covered in the next section.

## 2.4 Instant Messaging in the Workplace

Instant Messaging began as a social tool for communicating with friends across the internet but has since found its way into the workplace (Vos et al., 2004). The social beginnings of IM has caused concern among management that productivity could decrease due to non-work related conversations. Furthermore most IM implementations within the workplace rely on free IM clients such as AOL's Instant Messenger (AIM) and this raises more concerns regarding security and privacy. However, IM has been found to support workplace communications and all these claims will be further explored in this section.

In 2007 it was predicted that IM use in the workplace would increase by 20% annually until 2010 (Smith & Lundy, 2007). Furthermore, 90% of workers that have e-mail addresses would also have an IT-controlled IM account by the same time (Smith & Lundy, 2007). These statistics make it very clear that the use of IM is growing rapidly, especially in the workplace. Therefore, the use of IM can either positively or negatively impact workplace productivity as discussed next.

### Workplace Productivity

A study by Czerwinski et al. (2000) has shown that IM can lead to lowered productivity through its interruptive nature. The same conclusion was made by Cameron and Webster (2005) who classified this interruptive nature as unfair towards the recipient of the message. The reasoning is that the conversation is always at an appropriate time for the initiator while the recipient may be engaged in other tasks.

Table 2.2: Influence of communication systems on productivity

| Medium | Caller | Callee |
|---|---|---|
| Voice calls | No decrease of productivity because the caller (it is assumed) initiates communication to acquire work-related information | There is high probability of lowered productivity for the callee because communication can be disrupting to ongoing work-related activities |
| Instant Messaging | No decrease because the caller (it is assumed) initiates communication to acquire work-related information | There is a moderate probability of reduced productivity because the presence service can be used to minimize interruption |
| e-Mail | Reduction in productivity due to waiting for a response | Slight possibility of causing interruption (and therefore reduce productivity) because the receival of e-mail messages is controlled by the callee |

Existing technologies such as telephone and face-to-face conversations can also be interruptive and IM should therefore not be singled out in that context. In fact, other studies have found that IM is less-interruptive than other workplace communication mediums such as the telephone and face-to-face conversations (Nardi et al., 2000).

An IM message can be replied to at a more convenient time without incurring a feeling of unresponsiveness. This asynchronous characteristic is not shared by telephone and face-to-face conversations both of which incur much more pressure to respond quickly. Table 2.2 shows the effect of different communication technologies on productivity. Despite possible interruptions, IM provides many benefits to the workplace as discussed next.

## IM Usefulness in the Workplace

IM has been shown to support informal communication which is prevalent in the working environment (Isaacs, Walendowski, Whittaker, Schiano, & Kamm, 2002). This is mostly due to its light-weight synchronous nature that allows messages to be replied to quickly. In contrast, e-mail is an asynchronous medium and messages can take hours to be delivered depending on server congestion. Furthermore, IM allows a sense of presence awareness that signals whether a contact is online and willing to engage in conversation. The knowledge of a contact's presence can reduce unwanted communication and thus provides IM with a distinct advantage over telephone and face-to-

face conversations. Therefore, geographically dispersed co-workers can better schedule communication by making an informed decision based on a contact's presence information. It has even been found that situations exist where users prefer IM to other communication mediums such as the telephone, email and face-to-face communication (Cameron & Webster, 2005).

A common perception regarding IM is that it is primarily used for personal communication with family and friends when used in the workplace. However, both Isaacs, Walendowski, Whittaker, et al. (2002) and Handel and Herbsleb (2002) dismissed this notion and found that only a small percentage of IM conversations are non-work related. Only 13% of conversations analyzed by Isaacs, Walendowski, Whittaker, et al. (2002) included personal content and only 6.4% were explicitly personal. The impact of using free IM clients in the workplace will now be outlined.

## Free IM Clients

In the workplace, most IM implementations use free IM clients instead of corporate IM solutions (Vos et al., 2004). This is mostly due to users installing IM clients, which they use at home, on workplace computers. However, free IM clients are normally based on proprietary protocols and these do not take into account the security considerations and risk factors of the workplace. According to Stone and Merrion (2004) online security risk is evaluated according to the concepts of *confidentiality*, *integrity*, and *availability*. Confidentiality is compromised when an attacker gains access to a user's private data. Without proper end-to-end encryption an attacker can eavesdrop on communication and consequently breach confidentiality. Integrity relates to keeping data accurate by protecting it against unwanted modification. The use of some form of hashing algorithm ensures that no outsider can modify data along the way towards its destination without detection. Availability is provided by making sure that authorized users have access to information and services when they require them. A hacker can potentially target a service and use denial-of-service attacks to prevent legitimate use by other users. Most free IM clients do not implement proper security mechanisms as prescribed in various IETF specifications relating to IM.

### Interoperability

One of the biggest issues concerning technology adoption is interoperability (Hildebrand, 2003). Proprietary protocols contribute to another practical problem where multiple IM clients are required because existing free clients are generally not capable of interacting. The use of open standard protocols is one of the most prominent reasons for the success of e-mail. Consequently, e-mail has grown to be an integral part of today's organizational environment. An e-mail can be sent to anyone across the world because the common protocols are globally accepted open standards. Organizations have a need to interact, and for that to happen on a global scale interoperable protocols and standards are required.

IM lacks a one-for-all protocol, but the Internet Engineering Task Force (IETF) has done much work to develop such a protocol. The IETF formed the Instant Messaging and Presence Protocol (IMPP) in 1998 which in turn has proposed two such protocols: SIMPLE (SIP Instant Messaging and Presence Leveraging Extensions) and XMPP (Extensible Messaging and Presence Protocol). Of the two, only XMPP has been implemented by a major free IM client, Google Talk (Gtalk) (Strom, 2006). However, Microsoft has put considerable weight behind its support for SIMPLE (Hildebrand, 2003).

A reason for the slow adoption of open standards for IM systems is largely due to the current model of advertisement-based free IM networks. The companies behind these networks risk losing their user bases in the event of open standard IM adoption. Additionally, they stand to lose advertisement-generated revenue in the event that users opt for advertisement-free IM systems (Alvestrand, 2002). The development of this situation is discussed in the following section on the history of IM.

## 2.5 A Brief History of Instant Messaging

The history of IM has mostly been shaped by a few large Internet companies. The most influential being America Online (AOL), Microsoft and Yahoo. However, none of the aforementioned companies were responsible for *creating* IM. This honor may well be attributed to the developers of a UNIX tool called 'talk' (Greene & O'Mahony, 2004).

## UNIX talk

The UNIX 'talk' tool was released in 1973 and it allowed a user to communicate with others using text characters. Contrary to current IM systems, each character was sent as it was typed which made it difficult to identify turns within a conversation. Transmission of a character was almost instant, facilitating near-synchronous communication. Thus, the UNIX 'talk' tool qualifies as being the original IM application in view of the definition of IM in this dissertation.

The UNIX 'talk' tool lacked many of the features commonly found in current IM clients. Most notably, it lacked any form of presence awareness. The application provided no indication of whether the intended recipient was available before a message was sent or if the recipient even received the message at all (Greene & O'Mahony, 2004). To circumvent this problem 'talk' was often used together with a tool called 'finger'. The UNIX 'finger' tool was able to determine if a user was available. When used together, the 'talk' and 'finger' tools could be considered a legitimate presence-and-IM system. However, the two tools were loosely connected and several years passed before IM, in its current form, was introduced.

## ICQ

In 1996, a small Israeli company called Mirabilis developed an IM system called ICQ (pronounced 'I seek you'). ICQ introduced many IM features at that time, such as a buddy list and presence subscriptions (Salin, 2004). ICQ quickly became popular and gained a user base of over 850,000 within six months of its release. Its user-friendly GUI and free use largely contributed to its success.

Currently ICQ has over 400 million registered users, of which 15 million are active on the network. Most ICQ users are located in Europe and Asia.

Major Internet businesses such as America Online (AOL), Microsoft and Yahoo realized the success of ICQ and subsequently entered the IM market. AOL introduced its own IM client, AOL Instant Messenger (AIM), and shortly thereafter acquired both Mirabilis and ICQ. Microsoft and Yahoo soon followed AOL into the IM market and introduced their own solutions, MSN Messenger and Yahoo Messenger, respectively. More recently, Google

Table 2.3: Active user base per instant messaging system

| Messaging System | Active Users |
|---|---|
| AIM | 51.5 million |
| ICQ | 15 million |
| MSN / WLM | 27.3 million |
| Yahoo | 21.9 million |
| GTalk | <1 million |

introduced an IM client called Google Talk (GTalk). Table 2.3 shows the amount of active users for the different major IM systems within the United States in 2005 (Mills, 2005).

## AIM

America Online (AOL) already provided popular online communication services such as electronic bulletin boards during the early 1990's (Salin, 2004). In May 1997 AOL released the AOL Instant Messenger (AIM) to compliment its existing services. AIM quickly gained popularity, especially in America. It became the major competitor of ICQ until AOL acquired the company behind ICQ in 1998. Its user base continued to grow steadily and with the acquisition of Mirabilis, AOL covered about 90% of the IM market at that time. The introduction of several other competitors significantly reduced AOL's market domination. Currently, AIM is used by approximately 51.5 million people and is still the most popular IM client worldwide (Mills, 2005).

## MSN Messenger

Microsoft introduced its IM solution, MSN Messenger, in 1999. By coupling their IM system to other services such as e-mail and the Windows computer operating system, they quickly built a large user base (Salin, 2004).

Currently Microsoft has two IM clients, Windows Live Messenger (which used to be MSN Messenger) and Windows Messenger. The latter is an IM client released with the Microsoft Windows XP operating system. Further development on Windows Messenger was halted in 2007 in favor of Windows Live Messenger (WLM). Collectively, Microsoft has approximately 27.3 million users.

### Yahoo Messenger

During the same year that Microsoft introduced its IM solution, Yahoo released Yahoo Messenger. Due to the brand being well known on the Internet, Yahoo easily attained a large user base. Currently Yahoo Messenger has a user base of around 21.9 million. Since July 2006, Yahoo Messenger and Windows Messenger/Windows Live Messenger users have been able to communicate across the respective IM client boundaries (Mills, 2005).

### Recent History

In recent years the popularity of IM has increased significantly. However, IM use has been restrained by the independent development of the major IM systems. Traditionally none of the companies behind the major IM systems have developed their solutions with collaboration in mind (Alvestrand, 2002). They have all been built on proprietary protocols. The most probable reason for this is the fear of losing market share. This has caused a lot of frustration among users because they have had to resort to using multiple IM clients to communicate across IM clients.

### MXIT

Within the mobile community IM is becoming more popular as well. MXIT[1] is a java IM client for mobile phones with a registered user base of over 3 million, most of which reside in South Africa ("MXIT Reaches the Three Million Mark", 2005). The company responsible, MXIT Lifestyle, is currently venturing into the European market. Recently, a desktop beta version of the MXIT client was released which runs on the Windows platform.

## 2.6 Conclusion

This chapter has demonstrated and discussed the most important concepts regarding Instant Messaging. A definition of IM was established for the purposes of this dissertation: "*Instant messaging (IM) is a near-synchronous*

[1] http://www.mxit.co.za

*communication technology that facilitates text-based conversation between two or more participants*" (Avrahami & Hudson, 2006).

Common IM features such as a contact list and text-based communication were depicted by using the popular IM client AIM. It was shown how presence information is embedded within contact lists. Such presence information albeit very simple conveys more meaning than simply stating whether another user is reachable. Short keywords like "Busy" are provided to contacts to impart awareness of the IM user's presence.

The current state of Instant Messaging was shown to be very proprietary. IM vendors such as AOL and Microsoft are generating revenue through the embedding of advertisements in their IM clients. In order to protect their revenue, these vendors are lethargic in working towards interoperability and adopting open standards.

The Internet Engineering Task Force (IETF) has been investing a lot of effort in developing specifications concerning interoperability and standardization. Furthermore, the IETF has proposed two protocols, XMPP and SIMPLE, as candidates for the role of being the single standard IM protocol. The SIMPLE protocol is still in development but has already gained support from Microsoft and IBM. This is partly because SIMPLE is an extension to SIP, which is itself a popular signalling protocol used in Internet telephony. Currently, XMPP is implemented by Google's IM client, GTalk which was first released in 2005. Therefore, of the two, only XMPP is currently implemented on a large scale with a significant user base.

IM was compared to various other communication mediums and it was said that presence awareness sets IM apart from the rest. Presence awareness was defined as information regarding the status of another user on the communication network. Presence awareness can be used to calculate an appropriate time for engaging in conversation. In the following chapter presence will be explored in more detail.

# Chapter 3

# Presence

In the previous chapter the various strengths and weaknesses of IM were discussed and it was suggested that the current implementation of presence is still very basic. However, section 2.3.5 clearly stated that IM distinguishes itself from other mediated communication technologies through its implementation of presence awareness information. Presence awareness enhances the ability of IM to facilitate satisfying and effective communication. Studies have noted that IM conveys a sense of "being together" or "feeling connected", even without exchanging any messages (Nardi et al., 2000; Hwang & Lombard, 2006). Therefore, presence in IM appears to convey social psychological effects on participants as implied by the term "social presence".

Although face-to-face communication generates the most social presence, it is not always feasible to interact in such a manner. The alternative is mediated communication where participants do not communicate directly in a face-to-face interaction style. The goal of mediated communication is not just to mediate "being there" but also to mediate "being with other people" (Biocca, Kim, & Levy, 1995; Czitrom, 1982; Mcluhan, 1964; Lombard & Ditton, 1997). The latter implies a social connection which Rettie (2003) describes as *connectedness*. Humans are social beings that desire to be socially connected to others and even physical presence can be considered an effort to increase the level of social presence (Biocca, Burgoon, Harms, & Stoner, 2001).

The growth of the Internet has led to a growth in the amount of online mediated interactions. Also, the increasing affordability of being "online" allows people to stay connected for much longer. All these interactions exe-

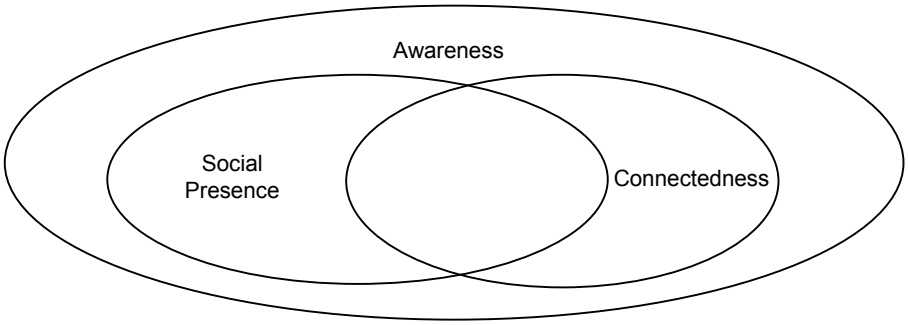


Figure 3.1: The relationship between connectedness, social presence and awareness

cute without physical face-to-face confrontations and this makes presence an important concept to explore (Rheingold, 1993).

In this chapter the concept of presence as it relates to mediated communication is considered. The roots of presence in social psychology will be explored, and the difference between presence awareness, social presence, and telepresence expressed. Next consider the concept of telepresence.

## 3.1 Telepresence

Presence can be categorized into *telepresence* which relates to "being there" and *social presence* which relates to the sense of "being together" (Biocca, 1997; Biocca et al., 1995; Heeter, 1992). Mediated communication allows one to be virtually present without actually being physically present. Therefore telepresence can also be described as virtual presence.

Most mediated communication technologies are used to gain access to people and not places (Fischer, 1988). In order to reach other people one needs to have access to them via a shared place. In mediated interactions sharing of a physical presence is not possible but through telepresence a virtual place can be shared.

As mentioned previously, humans are social beings and therefore the most prominent reason for physical presence is to increase a sense of social presence (Biocca et al., 2001). Telepresence only provides the platform through which social presence can be experienced in a mediated environment.

Rettie (2003) explains the relationship between awareness, social pres-

ence and connectedness as interrelated, but argues that the experience of connectedness is different from the sense of social presence. Furthermore, the experience of connectedness is more fundamental than social presence and simple awareness because it is a basic human need. Although it is through awareness that mediated communication attain sufficient levels of social presence, connectedness is the underlying factor that drives the need for awareness systems. Connectedness can therefore be described as telepresence because it provides the underlying "shared space" from which social presence and awareness can be realized. Figure 3.1 shows the relationship between connectedness, social presence and awareness.

Both To et al. (2008) and Hwang and Lombard (2006) report the positive role that the experience of connectedness (i.e. telepresence) performs in IM communication. It can be difficult to differentiate between social presence, connectedness and awareness. Social presence always implies awareness of another person and is seen as a property of the communication medium (Short, Williams, & Christie, 1976). For instance, a mobile text message provides almost no presence of the sender yet it may create a rich experience of connectedness (Rettie, 2003). Similarly, awareness is possible without social presence. For instance, one could be aware of an object, which would exclude the domain of social psychology in general, and social presence in particular (Rettie, 2003). Like simple awareness, telepresence is not concerned with the social interpersonal relationship of communicators. Telepresence is realized by simply knowing that someone is sharing the same virtual space.

## 3.2 Social Presence

Presence within the context of mediated communication is rooted in social psychological studies. Classic social psychologist, George Herbert Mead's work on "the other" can be seen as the earliest form of social presence (Biocca et al., 2001). However, the term *social presence* was introduced by Short et al. (1976) and defined as *"the salience of another person within mediated communication and the salience of their subsequent interactions"*. In general terms, social presence is a measure of how aware two communication partners are of one another within a mediated environment (Biocca, 1997; Rettie, 2003). Gunawardena (1995) defines social presence as *"the degree to which a*

*person is perceived as a 'real person' in mediated communication".*

The definitions mentioned above highlight two important characteristics of social presence. The first is its measurable level indicating that it can differ and fluctuate which has been investigated by many studies (Gunawardena, 1995; Rice, 1993; Walter, 1992; Walther & Burgoon, 1992). The second is its subjective nature which is based on interpersonal perceptions. The latter is noted as a "subjective judgement of the medium" to facilitate social presence (Short et al., 1976).

Biocca et al. (2001) split social presence into three components:

- *Co-presence* which is a sensory awareness of the person being communicated with. Social presence through mediated communication is not simply a binary "is there or not" concept. In mediated communication the senses of a user are extended through the communication technology and represented to the other in a limited manner (Biocca et al., 2001). The other is commonly represented in some kind of avatar, agent or simple construct (Biocca & Nowak, 2001).

  Each sensory channel is a medium for experiencing social presence where small environmental variables make a difference. These sensory channels exist at some level of mutual awareness (Goffman (1959) as cited by Biocca et al. (2001)). Co-presence can be said to be mutual awareness when the focus of the interaction is on the sensory awareness of each other. In simple terms it can be described as simply "being together" (Greef & IJsselsteijn, 2000; Ho, Basdogan, Slater, Durlach, & Srinivasan, 1998).

- *Psychological involvement* in that social presence can exist only through access to another intelligence. Simply being present does not necessarily imply social presence. For instance, a corpse may be physically present, but not socially present (Biocca et al., 2001).

  Short et al. (1976) imply attentional requirements and define social presence as a "salience of the interpersonal relationship". Rice (1993) supports this notion of psychological involvement in studies of media appropriateness and relates them to two other social concepts, intimacy and immediacy.

Savicki and Kelley (2000) define social presence with the emphasis on the ability to project a sense of self through a limited communication medium. Projecting a sense of self requires to some degree psychological engagement.

- *Behavioral engagement* which means that there needs to be a reciprocal exchange of communication behaviors. Such communication behaviors include eye contact, non-verbal mirroring, and turn-taking. (Biocca et al., 2001).

  Interactive virtual environments and multiplayer online computer games have opened many new avenues for behavioral interaction (Biocca et al., 2001).

  Palmer (1995) defines social presence as "the effective negotiation of a relationship through an interdependent, multi-channel exchange of behaviors". Such an exchange of behaviors implies *behavioral engagement*.

To understand social presence we need to understand the behaviors used to simulate and infer the content of other minds (Biocca & Nowak, 2001). The much referenced work of Short et al. (1976) draws more directly from social psychological work on nonverbal communication in interpersonal interactions (Weiner & Mehrabian, 1968). Nonverbal communication is split into the concepts of intimacy and immediacy as discussed next.

### 3.2.1 Intimacy and Immediacy

Although social presence is specifically studied in mediated interactions, the underlying concepts of face-to-face interaction such as intimacy and immediacy are at the core of social presence. Intimacy behaviors such as eye-contact and physical proximity is an integral part of natural conversation (Argyle & Dean, 1965). Immediacy behaviors are also prevalent in natural communication and particularly relate to the synchronous behaviors that help maintain a connection between communication partners (Weiner & Mehrabian, 1968). Verbal cues such as timing, pause, inflection and nonverbal cues such as facial expression are properties of face-to-face conversation that allows the interaction to execute smoothly.

Intimacy is a variable of *psychological distance* which is kept in equilibrium by two interacting participants (Argyle & Dean, 1965). In a normal face-to-face encounter, intimacy is kept at a comfortable level through immediacy behaviors of eye-contact and the physical proximity of the participants. Greater levels of intimacy lead to greater levels of social presence (Short et al., 1976). Text-based communication mediums like IM and e-mail are severely limited in their relative levels of intimacy.

If a communicator is able to adjust the level of intimacy easily it leads to a satisfying and effective conversation. Also, if there are more immediacy behaviors available it contributes to a smoothly executing interaction. Interactions with maintained eye contact, close proximity, body leaning forward, and smiling increases intimacy (Burgoon, Buller, Hale, & deTurck, 1984).

It has been suggested that an increase in immediacy behaviors to be used, allows more control over the intimacy within a mediated conversation (Short et al., 1976). Consequently, the social presence is increased and the result is more efficient and satisfying interaction. Connell et al. (2001) conducted an empirical study within which impression management, media richness and presence were investigated. Their questionnaire focused on self-awareness and communicators' satisfaction with how they projected themselves. The surprising result was that the telephone came out as favored over face-to-face communication. The suggested reason is that the telephone provides an optimum balance between awareness and comfortable intimacy.

The second psychological concept involved in social presence is immediacy. Tu (2002b) defines immediacy as the psychological distance that a communicator places between himself and the recipient of the communication. The distance is a function of eye contact, smiling, vocal expressiveness, physical proximity, appropriate touching, leaning toward a person, gesturing, using overall body movements, being relaxed and spending time with someone.

Immediacy is difficult to deliver in mediated communication because of the lack of nonverbal cues. However, the importance and existence of online immediacy is not negated by this shortcoming. Instead Tu (2002b) notes that immediacy becomes even more important due to its scarcity and the basic human need for social contact.

If intimacy and immediacy are functions and actions that are continuously

changing, it could be said that social presence is not static and that there may be varying degrees as stated in the next section.

### 3.2.2 The Degree of Social Presence

The degree of social presence is not concrete per communication medium and is considered to be a subjective property of the medium as perceived by the communicator (Short et al., 1976). Social presence can also be defined as the degree to which physical presence can be simulated within a communication medium (Connell et al., 2001). Thus, social presence is a phenomenon that varies with different mediated communication channels. Social presence studies show that a higher level of social presence is preferred because it allows more robust immediacy behaviors and dynamic intimacy behaviors.

The degree of social presence is a subjective factor and may be affected by simple issues such as computer literacy (Tu, 2002b). For instance, if someone is not used to typing, messages in a text-based medium such as IM may be composed slower. Consequently, it may lower the judgement of the medium's social presence by the recipient of the message because of frustration due to the long wait.

Short et al. (1976) wrote a very influential book concerning social presence. They measured social presence by a "subjective quality of the communications medium". Their approach used semantic differential scales that capture some of the social and emotional capabilities of the medium. It should be noted that the users were asked to judge the medium directly. Biocca et al. (2001) suggests that this may not be the best way as the phenomena of presence is more about how real the experience was than how real the medium is.

Short et al. (1976) refer to three features of interpersonal communication: involvement, immediacy and intimacy. These features were never actually measured by Short et al. (1976), but by others such as Burgoon and Hale (1987) and Nowak (2000).

If one considers all social presence to be dynamic, then studies of face-to-face communication are relevant for mediated communication (Biocca et al., 2001). Gunawardena and Zittle (1997) measure intimacy by combining the semantic differential scales as used by (Short et al., 1976) with a focus on intimacy.

Nowak (2000) distances himself from the judgement of the medium to a measure of *social attraction* called homophily (McCroskey, Richmond, & Daly, 1975). Through the concept of homophily the user's perception of avatars and agents can be measured in mediated environments.

Some verbal markers or nonverbal indicators such as facial expression may be indicative of social presence (Biocca et al., 2001). Nonverbal behaviors such as proximity to the other person are used as dependent or independent variables in studies of social interaction (Biocca et al., 2001).

Social Presence is a dynamic variable and dependent on both the medium and the user's perceptions (Tu, 2002b). Short et al. (1976) measured social presence through a semantic differential technique using a seven-point assessment. They suggested that a higher level of presence in a medium convey higher levels of being more sociable, more personal, more sensitive, and warmer.

In online learning environments it has been shown that social presence can be increased by training learners to use CMC technologies (Tu, 2002b).

Within the context of online learning Tu and McIsaac (2002) split social presence into different components than Biocca et al. (2001) as was given in section 3.2. The interpretation of social presence by Tu and McIsaac (2002) is given next.

### 3.2.3 The Three Dimensions of Social Presence

Tu and McIsaac (2002) split social presence into *social context*, *online communication* and *interactivity*. Each will be considered in turn.

#### Social Context

Social context is constructed from CMC users' characteristics and their perception of the CMC environment (Tu, 2002b). If communication partners are unfamiliar with each other and the conversation is task orientated and more public, the degree of social presence will degrade (Tu, 2002b; Walter, 1992). Different social processes, settings and purposes are components of social context and affect social presence (Walter, 1992).

### Online Communication

Online communication is concerned with the attributes of the language used online and the applications of online language, such as the characteristics of CMC, computer literacy skills, online immediacy and online language skills.

The text-based format of CMC communication requires some level of computer literacy (i.e. typing, using the mouse, and reading from the screen). Without these skills of communication people develop communication anxiety (Gunawardena, 1991). Garramone, Harris, and Anderson (1986) and Perse, Burton, Kovner, Lears, and Sen (1992) established that the degree of social presence among computer bulletin board users were higher for those who were more active than those who were not. Essentially Perse et al. (1992) found a positive correlation between social presence and users' perception of their own computer expertise. Similarly, Tu (2002a) found that users who make use of emoticons experienced higher levels of social presence.

### Interactivity

Interactivity includes the activities in which CMC users engage and the communication styles they use. Response time, task types, topics and group size are examples of communication styles and activities of CMC users (Tu, 2002b). The potential for feedback during communication contributes to the degree of salience of the other person in the interaction (Tu, 2002b).

Immediate response is another component of interactivity. If an immediate response is expected and not received, a feeling of low interactivity is created which negatively affects the quality of the interaction (Tu, 2002b). Gunawardena (1995) differentiates between interactivity and social presence, arguing that interactivity is an awareness service which leads to social presence if it is realized. Awareness services convey presence through the characteristics of the communication medium. Awareness provides information about another person as outlined next.

## 3.3 Awareness

Awareness is defined as *"the understanding of the activities of others, which provide context for your own"* (Dourish & Bly, 1992).

In a paper by Tang and Begole (2003), the future of "awareness" services are explored by discussing current research prototypes. The three prototypes discussed are Awarenex (Tang et al., 2001), Rhythm Awareness (Begole, Tang, Smith, & Yankelovich, 2002) and Lilsys (Begole, Matsakis, & Tang, 2004). These prototypes all look beyond the current implementation of presence in IM to better accomplish online collaboration among groups. Tang and Begole (2003) suggest that future collaboration tools should not only include current presence state information but also include *future reachability*, *availability* and *context* information.

Tran, Yang, and Raikundalia (2005) distinguish among four categories of awareness in IM: *turn-taking awareness*, *contextual awareness*, *emotional awareness* and *presence awareness*.

### 3.3.1 Turn-taking Awareness

Turn-taking is a fundamental process of normal human interaction. It is hard to facilitate turn-taking in mediated environments because it is based on various verbal and nonverbal cues which do not exist explicitly in mediated communication.

MSN Messenger uses the text-indicator "...is typing" to provide awareness that the other user is busy typing a message. In this way, the person reading the indication can wait for the message before responding. Many IM clients like GTalk and MSN Messenger use an *idle* status message to convey inactivity on the other side of the mediated channel. Inactivity implies an "open floor" and a very small likelihood of "talking at the same time". The Hubbub prototype also shows whether the shared text window is the main focused window on the other user's PC (Isaacs, Walendowski, & Ranganthan, 2002).

Most implementations of mediated turn-taking awareness are still lacking when compared to the social norms of turn-taking in natural conversation. Nevertheless, turn-taking awareness contributes to improve the *behavioral engagement* component of social presence which leads to a more satisfying mediated interaction.

### 3.3.2 Contextual Awareness

Contextual awareness raises the level of social presence by providing desired information *about* the interaction between communicators. The most popular example of contextual awareness is the use of a "shared window". In section 2.2.3 the shared window containing the complete conversation dialogue was introduced. This shared conversation window provides awareness of the flow of messages between the two participants. Such awareness makes it easy to reference messages exchanged earlier in the conversation. Also, multi-threaded conversations can be dealt with more efficiently by viewing both sides of the conversation.

Identity awareness is another form of contextual awareness. The knowledge of whom a user is communicating with influences the content of messages as well as the behavior of the user (Tran et al., 2005). It was shown by Tu (2000) that in mediated learning environments the familiarity of the person on the other side improves social presence.

People use nicknames, avatars, colored circles (Viégas & Donath, 1999), and even "sound ID's" to portray themselves within a virtual mediated environment. It has also been shown that people use multiple identities online to project themselves differently (Tran, Yang, & Raikundalia, 2004). The ability to express yourself through a medium as intended, contributes to social presence (Short et al., 1976).

### 3.3.3 Emotional Awareness

The social desire for perceiving and projecting emotions is an important and fundamental human need (Hancock, Landrigan, & Silver, 2007). The correct interpretation of another person's affective state decreases the likelihood of misunderstandings. Kato, Kato, and Akahori (2007) found that less emotional cues in mediated communication increased the perceived level of negative emotions more than positive emotions.

Both a person's own emotional state as well as the perception of the emotional state of others affect communication processes and conversation content (Damasio, 1994). However, it is difficult and cumbersome to express emotions in text-format (Sánchez, Hernández, Penagos, & Ostróvskaya, 2006). Emoticons such as ':-)' have been introduced in IM to express emotions

through simple symbols made up of a short combinations of text characters. Emoticons are often used to strengthen the emotional nuance of messages similarly to how facial expressions help to convey a message in face-to-face interactions (Derks, Bos, & Grumbkow, 2008). Many IM clients such as Yahoo Messenger and AIM provide a menu from which emoticons can be chosen.

Emotional awareness helps to provide a more accurate impression of the other communicator's emotional state. Therefore emotional awareness mechanisms help to heighten social presence with regards to *psychological engagement*.

### 3.3.4 Presence Awareness

Presence awareness lies at the heart of social presence in a mediated environment. The awareness of another's presence in a mediated environment raises the level of *co-presence* and *psychological engagement* even without *behavioral engagement* (To et al., 2008).

McClea, Yen, and Huang (2004) define presence awareness as the connectivity and availability of the users in a mediated evironment. To et al. (2008) go further and describe presence awareness as the ability to know the availability of a potential communication partner before a communication attempt is made. This knowledge allows users to indicate the most appropriate time and means of being contacted (Licari, 2005).

Traditionally, IM is the only communication tool that implements presence awareness natively. It is this ability that separates and distinguishes IM from other communication tools. However, presence awareness is not limited to IM and can possibly be implemented in any communication tool. In IM it is often implemented as a basic state and a descriptive keyword. Keywords such as *online*, *available*, and *busy* all help to portray with greater accuracy, the most current state of another user. These status keyword were introduced and discussed in section 2.2.2.

To et al. (2008) found that presence awareness was the single biggest positive factor encouraging workplace IM adoption. This is in accordance to what Hwang and Lombard (2006) found within the user base of college students. The only difference is To et al. (2008) refer to presence awareness by the more general term of *social presence*.

## 3.4 Privacy

Privacy, like social presence, is rooted in social psychology but has been found useful in information technology research. Social psychologist Irwin Altman defines privacy as the "selective control of access to self" as cited by Palen and Dourish (2003). Presence was defined in chapter 2 as information that describes the status of a user. Therefore, according to the definition, a user should have selective control over access to his own presence information.

Most current IM systems do provide users with mechanisms to control access to their presence information. Also, it is common for IM systems to require authorization to simply communicate with another user on the network. Such authorization is permitted by the adding of contacts to one's contact list.

The increase of IM use has led to the fusion of different usage contexts (e.g. work, family, friends). In other words, people are increasingly using IM clients to communicate with colleagues at work, family at home and friends. The implications and concerns with privacy amongst these spheres will be shown using the privacy model introduced by Palen and Dourish (2003).

Based on the privacy model by Irwin Altman, Palen and Dourish (2003) illustrated the tensions associated with privacy control. One of the prominent observations is that privacy is dynamic and changes with context. It is therefore not feasible to introduce hard and fast rules about if and how much privacy should be relinquished to a particular person. The three tensions associated with privacy regulation are *the tension of privacy and publicity*, *the tension of identity and audience* and *the tension of temporality*.

### 3.4.1 The Tension of Privacy and Publicity

The problem with privacy is that it is not simply a matter of withholding information. From a social perspective we give out certain information to make ourselves available and even display individuality. However, we do keep certain information private and can therefore be said to have both a private and a public life (Palen & Dourish, 2003). To manage privacy we need to look at both of these contrasting virtues.

Within a mediated communication system, a user needs to disclose some information to inform his/her contacts of his/her existence in the virtual

community. The use of a personalized avatar or status message in an IM client allows a user to exhibit individualistic qualities.

A problem arises when a contact is added with whom a user wants to limit certain types of information. An IM user may not want to share his cell phone number with colleagues, but only with friends and family. Therefore, a tension exists between remaining private by not disclosing presence information and being public by sharing presence information.

### 3.4.2 The Tension of Identity and Audience

People construct their actions according to the observers of their actions. In other words, the relationship and context of the other person will affect how a person presents himself. Also, when acting as part of a bigger group, such as a business, a user may be compelled to behave in a certain manner (Palen & Dourish, 2003).

Mediated environments make it noticeably more difficult to regulate how a person presents himself. The possible persistence of an interaction further increases the need to communicate and represent oneself as intended. Therefore, a tension exists between the way that a person is represented in relation to the associated observer of such a representation.

### 3.4.3 The Tension of Temporality

Communication usually occurs between people who share some kind of relationship with each other. Every interaction between them builds on the previous encounter. The regulation of privacy needs to be seen within the context of the concepts of past, present and future.

An example of this tension would be the amount of information that is shared to a particular person. If this amount is changed it could be perceived as an act of degenerated trust. Another issue is the dynamic quality of privacy. At different times of the day, privacy requirements may be different and need to be handled differently.

The presence processing model proposed by Rosenberg (2005) prescribes the way the IETF's SIP and SIMPLE protocols describe the process of filtering presence information. The whole approach to presence-handling by the IETF is discussed in the next chapter.

## 3.5 Conclusion

In this chapter presence has been explored and considered in detail. The roots of presence were shown to be in social psychology. Presence can be split into *telepresence* and *social presence*. Telepresence is the concept of "being there" while social presence is the concept of "being with other people".

Social presence can be experienced through mediated communication by means of awareness services. These awareness services enhance the interaction and the following four were mentioned: turn-taking awareness, contextual awareness, emotional awareness, and presence awareness. Of these awareness services, presence awareness was shown to be very useful as it has a practical benefit of helping communicators estimate the likelihood of a successful interaction attempt.

Currently, the implementation of presence in IM awards full trust to all watchers, providing them with all presence information. However, the IETF has specified several RFCs relating to presence information that differs from what is currently implemented. In the next chapter, the efforts of the IETF with regards to presence and the implementation of presence in IM will be considered in detail.

# Chapter 4

# The Implementation of Presence

In the previous chapter *presence* was considered with a view to its social and psychological roots. It was argued that awareness services, such as presence awareness, can greatly enhance the quality of mediated communication.

Since this research focus is on Instant Messaging, which by definition operates in a technological milieu, this chapter considers how the concept of presence is implemented. Specifically, the the efforts of the Internet Engineering Task Force (IETF) are investigated. Presence-related specifications will receive particular focus.

## 4.1 Efforts of the Internet Engineering Task Force

In chapter 2 it was said that most IM and presence systems operate on proprietary, non-interoperable protocols. The Internet Engineering Task Force (IETF), a global organization developing Internet standards, realized this and have done much work in standardizing IM and presence.

The IETF formed the Instant Messaging and Presence Protocol (IMPP) working group in 1998. Originally the IMPP working group was chartered to define the basic requirements and data formats that could be used to develop a global interoperable IM standard. The results of this endeavor are documented in RFC 2778 and RFC 2779. Figure 4.1 shows a list of the

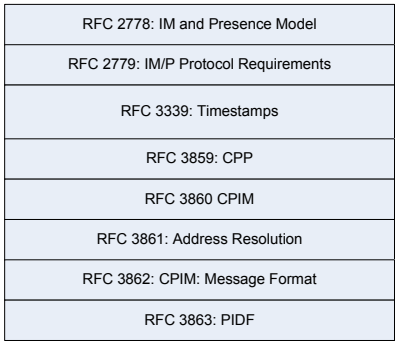


Figure 4.1: IMPP specifications

specifications presented by the IMPP working group.

The IMPP working group was since disbanded and several working groups were chartered to continue its work. Each of these working groups specified potential protocols based on the IMPP. Before the termination of the IMPP working group, the IMPP focused on work involving interoperability between instant messaging systems. These RFC's include:

- RFC 3863: a common extensible Instant Message format (message/cpim) (Sugano et al., 2004)
- RFC 3863: a common extensible presence information data format (application/pidf+xml) (Sugano et al., 2004)
- RFC 3860: a common profile for Instant Messaging (CPIM) (Peterson, 2004a)
- RFC 3859: a common profile for presence (CPP) (Peterson, 2004b)

The completion of the IMPP's goals and its consequent disbanding led to to the SIMPLE and XMPP working groups being instantiated. These two working groups aimed at providing a standard interoperable IM and presence protocol which could be deployed globally. The XMPP working group is currently closed and XMPP has become a standard. In contrast, development on SIMPLE continues and it is the prime candidate for the position of single standard IM protocol (Smith, 2007).

However, both the SIMPLE and XMPP protocols are based on the IMPP model for IM and presence which is considered next.

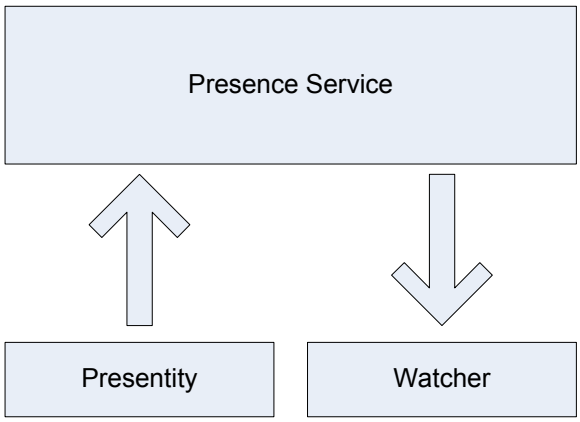


Figure 4.2: IMPP presence model

## 4.2 A Basic Model for IM and Presence

One of the main contributions by the IMPP working group was the specification of a basic model for IM and presence. The model specified in RFC 2778 is abstract enough to depict most current implementations of IM and presence systems (Day et al., 2000). It is seen as a platform on which further specifications of IM and presence should be based.

The IMPP presence-specific part of the model is depicted in figure 4.2 and the IM-specific part in figure 4.3. The model also defines a basic nomenclature for IM and presence. For the purposes of this dissertation, the two most important terms presented in RFC 2778 are *watcher* and *presentity*. A presentity is an entity that has presence information. In contrast, a watcher

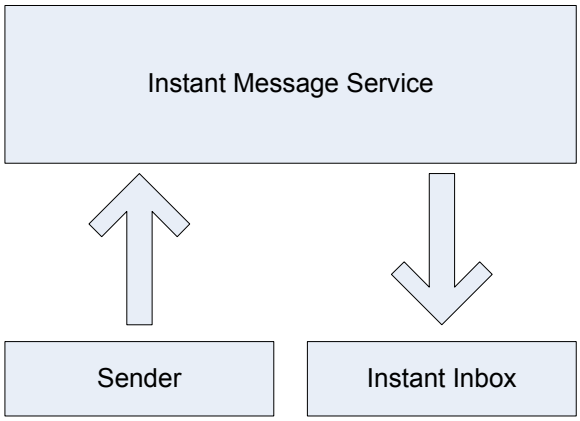


Figure 4.3: IMPP Instant Messaging model

is an entity that is interested in the presence information of a presentity.

Presence-handling within the IMPP model can be described as follows. A watcher has to ask permission to subscribe to a presentity's presence information. Once a presentity permits the watcher to do so, the presence service makes the presentity's presence information available to the watcher. The watcher will have access to a presentity's presence information as long as the subscription is valid.

There are three different types of watchers: pollers, fetchers, and subscribers. A fetcher simply fetches presence information whenever the need arises. A poller is a fetcher that fetches periodically. Lastly, a subscriber does not actively retrieve presence information but gets notified with presence information by the presence service when appropriate. The following section answers the question of "how is presence information represented?"

## 4.3 A Data Model for Presence

The Presence Information Data Format (PIDF) is specified by the IMPP in RFC 3863 and describes a document format for grouping presence information (Sugano et al., 2004). The PIDF document format was specified by the IMPP working group with the purpose of becoming the *de facto* specification for presence information.

Within IM and presence architecture, the presence server keeps record of a presentity's most current presence information. Such presence information is kept in a PIDF document to be presented to watchers. According to the watcher-subscriptions, the PIDF document is then presented to the watchers.

An example of a PIDF document is given in figure 4.4. The PIDF was designed with multi-channel communication in mind of which IM is one such channel. Communication channels are represented by tuples with an associated priority. Two tuples are shown in figure 4.4, the first starts with the xml tag *<tuple id="bs35r9">*, the second starting with *<tuple id="eg92n8" textgreater*. Channels are either *open* or *closed* but may contain an extra descriptive keyword such as "busy" or "available". In the example document depicted in figure 4.4, both channels are "open", indicating that the presentity is reachable through both. Similarly, a "closed" channel indicates that a presentity is off-line or not reachable.

```
<?xml version="1.0" encoding="UTF-8"?>
   <presence xmlns="urn:ietf:params:xml:ns:pidf"
        xmlns:im="urn:ietf:params:xml:ns:pidf:im"
        xmlns:myex="http://id.example.com/presence/"
        entity="pres:someone@example.com">
     <tuple id="bs35r9">
       <status>
         <basic>open</basic>
         <im:im>busy</im:im>
         <myex:location>home</myex:location>
       </status>
       <contact priority="0.8">im:someone@mobilecarrier.net</contact>
       <note xml:lang="en">Don't Disturb Please!</note>
       <note xml:lang="fr">Ne derangez pas, s'il vous plait</note>
       <timestamp>2001-10-27T16:49:29Z</timestamp>
     </tuple>
     <tuple id="eg92n8">
       <status>
         <basic>open</basic>
       </status>
       <contact priority="1.0">mailto:someone@example.com</contact>
     </tuple>
     <note>I'll be in Tokyo next week</note>
   </presence>
```

Figure 4.4: Example PIDF presence document

In RFC 4480, the Rich Presence Extensions to PIDF (RPID) are specified (Schulzrinne, Gurbani, Kyzivat, & Rosenberg, 2006). As the name implies, it provides richer presence information by extending the PIDF specification. The extensions include *what the presentity is doing*, *the presentity's mood*, and *the type of place the presentity is in*. Overall, the extensions are implemented in such a way that much of the information can be derived automatically.

## 4.4 A Model for Processing Presence

The Internet Draft "A Processing Model for Presence" provides a model that describes and defines the processing operations used by presence agents in processing presence information in a SIP and SIMPLE environment (Rosenberg, 2005). The model is depicted by figure 4.5.

Watchers whose subscriptions have been accepted receive presence information notifications. To fulfill notifications the presence server must generate

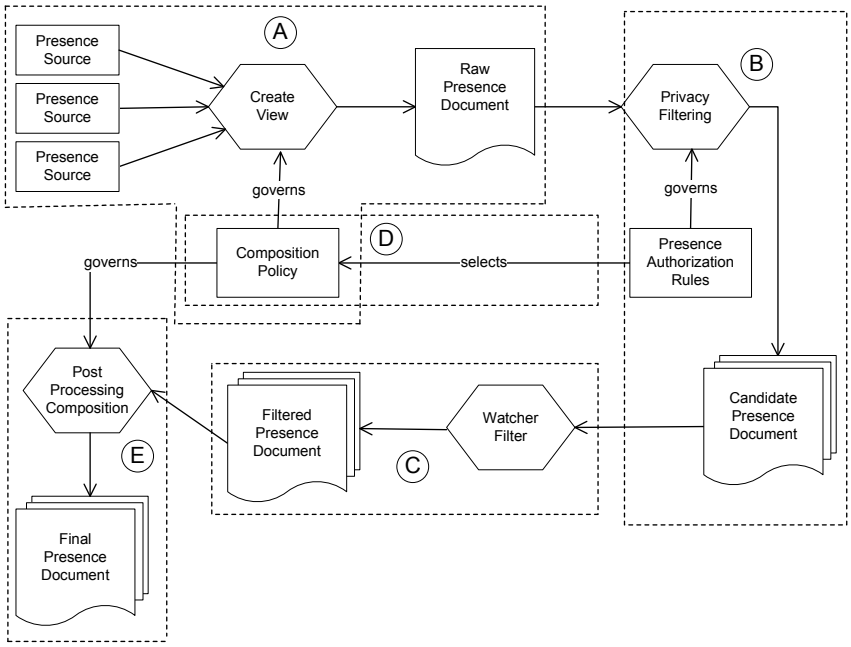


Figure 4.5: IETF presence processing model (Rosenberg, 2005)

a presence document for each watcher. This presence document generation process, detailed by Rosenberg (2005), shall be discussed in brief.

The presence processing model depicted by Figure 4.5 has been partitioned into five logical groupings labeled A, B, C, D and E. Grouping A represents the collection and composition phases which result in the creation of the raw presence document. A default composition policy is used to generate this raw presence document.

Grouping B represents the privacy filtering which must take place on a watcher-by-watcher basis. Presence authorization rules (Rosenberg, 2005) are applied to ensure that each watcher receives only that subset of presence information to which he/she is entitled i.e. the candidate presence document. It is possible at this point that the authorization policy can select a composition policy other then the default to generate the presence document sent to the watcher. This optional process is indicated by D.

Grouping C illustrates the ability of the watcher to optionally further filter the presence information he/she will receive, resulting in a presence document filtered according to his/her requirements.

Grouping E shows the optional application of further composition rules prior to the generation of the final presence document.

The most prominent difference between current implementations of presence processing and the SIP/SIMPLE processing model is the level at which presence information is filtered. In current implementations there is no filtering and all authorized watchers receive the same presence information. In the model shown here presence information is filtered for each watcher. This process of filtering would be done by the enforcement of a presence policy as discussed next.

## 4.5 Presence Policy

A presence policy is the set of authorization rules that governs the presence filtering process as described in the previous section. A presence policy is specified in RFC 5025 as a contract between the presentity (who specifies the policy), and the presence server (which correctly enforces the policy). According to RFC 5025, the presence policy describes "what presence information can be given to which watchers, and when" (Rosenberg, 2007).

A presence policy is expressed as a set of rules in an XML document. Figure 4.6 depicts a ruleset containing a single rule which starts with the *<cr:rule id="a">* XML tag.

RFC 5025 extends the more generic document format for expressing privacy as defined in RFC 4745 (Schulzrinne et al., 2007). The presence rules, shown in figure 4.6, illustrate that every presence rule in a ruleset consists of three distinct sections, namely:

- conditions: under which the presence rule applies,
- actions: specifying what action the presence server takes, and
- transformations: transforming the presence document to be presented to a watcher.

### Conditions

The *conditions* component describes the circumstances under which a presence rule will be applied. Most importantly, the *Identity* element specifies the identity of the watcher to whom the presence rule applies. In figure 4.6, the presence rule applies to a watcher called "John" as seen in the conditions element.

```
<?xml version="1.0" encoding="UTF-8"?>
   <cr:ruleset xmlns="urn:ietf:params:xml:ns:pres-rules"
    xmlns:pr="urn:ietf:params:xml:ns:pres-rules"
    xmlns:cr="urn:ietf:params:xml:ns:common-policy">
    <cr:rule id="a">
     <cr:conditions>
      <cr:identity>
       <cr:one id="John"/>
      </cr:identity>
     </cr:conditions>
     <cr:actions>
      <pr:sub-handling>allow</pr:sub-handling>
     </cr:actions>
     <cr:transformations>
      <pr:provide-services>
        <pr:service-uri-scheme>sip</pr:service-uri-scheme>
        <pr:service-uri-scheme>mailto</pr:service-uri-scheme>
      </pr:provide-services>
      <pr:provide-persons>
        <pr:all-persons/>
      </pr:provide-persons>
      <pr:provide-activities>true</pr:provide-activities>
      <pr:provide-user-input>bare</pr:provide-user-input>
       <pr:provide-unknown-attribute
        ns="urn:vendor-specific:foo-namespace"
        name="foo">true</pr:provide-unknown-attribute>
     </cr:transformations>
    </cr:rule>
   </cr:ruleset>
```

Figure 4.6: Example presence authorization rule

RFC 5025 does not extend the conditions specification in RFC 4745. Therefore, presence information cannot be used as conditions without extending RFC 4745. The exception is *sphere* which is also defined in the RPID extensions to PIDF. Sphere is similar to location, allowing the presentity to use a *mode of operation* to limit the visibility of a watcher to presence information. Examples of sphere values are work, social, and meeting. In order to provide a watcher with a different set of presence information in the work sphere and the social sphere, two rules will then be defined for that watcher where the conditions are as follows:

- rule 1: identity condition equal to the watcher, and sphere condition equal to “work”, and
- rule 2: identity condition equal to the watcher, and sphere condition equal to “social”.

### Actions

The *actions* component specifies what actions the presence server should take. Subscription handling is the only type of action currently specified in RFC 5025. The subscription can be handled by the following actions:

- block: the subscription is rejected, which is also the default action;
- confirm: the server places the subscription in a pending state for the presentity to confirm;
- polite-block: the subscription is placed in the active state, but a generated presence document specifies that the presentity is unavailable; and
- allow: the subscription is placed in the active state, allowing the presence document to be provided according to the transformations.

### Transformations

The *transformations* component aims to control a watcher's visibility to certain parts of a presence document. For instance, transformations can limit a presence document to only person, device, or service associated elements. However, authorization can be controlled in a more fine-grained fashion by defining transformations that allow a specific element of the presence document to be presented to the watcher.

The XML schema defined in RFC 5025 specifies the acceptable values for presence rule document elements. Within the schema, transformation elements are expected to have boolean (yes or no) values. The exception is the *provide-user-input* transformation which controls access to the *user-input* element in the PIDF document according to predefined values. It should be noted that transformations describe how the presence document is transformed, not the transformation of presence document values. In other words, it specifies that you can view presence attribute values but does not allow "transformation" in the sense of changing presence attribute values.

Both the PIDF and RPID documents are encouraged to be extended by their respective specifications. Therefore, a *provide unknown attribute* is defined in the presence rule XML schema to make provision for authorizing

presence information not currently described in the schema. However, this transformation is also limited to a boolean type.

Every *action* or *transformation* is also known as a permission, seen as a positive grant of presence information to the watcher. In accordance with chapter 3, RFC 5025 states that presence information is sensitive and the authorization of presence information is a critical function of presence systems.

## 4.6 Conclusion

The current implementation of presence in IM can be described as a watcher subscribing to the presence information of a presentity. The IETF has captured this in their basic model for IM and presence. Additionally, the IETF is working towards a single Internet standard for IM and presence. The two candidate protocols are SIMPLE and XMPP.

The PIDF and RPID presence document formats aim to include a rich set of presence information not available in current implementations of presence in IM. The motivating factor is the increasing pervasiveness of computing and the growing connectedness and communication facilities afforded by the Internet.

In order to accommodate presentity privacy concerns and workplace security policy, a presence processing model is defined by Rosenberg (2005). The model describes the various processes used in the distribution of presence information. Of particular interest is the privacy filtering via presence authorization rules.

A presence policy consists of presence authorization rules and describes a contract between the presentity and the presence server. The presentity defines the rules and the presence server implements the ruleset in accordance. The purpose of a presence policy is to define "what presence information can be given to which watchers, and when". Such authorization is similar to an access control system that limits access to a resource based on a set of permissions. Prior to proposing an enhanced presence-handling model, the concepts of access control should be considered next.

# Chapter 5

# Access Control

In chapter 4 the IETF presence processing model proposed by Rosenberg (2005) was discussed. The model specifies the processes regarding the filtering of presence documents according to a presence policy. In section 4.5 presence policy was defined as a collection of several presence rules, each of which applies to the identity of a single watcher.

A presence document contains detailed information concerning the presentity. The purpose of the IETF presence processing model is to only provide a subset of the complete presence document to each watcher. In other words, each watcher only receives the intended presence information according to the presence rules.

The limiting of presence information based on the watcher's identity is similar to access control principles which specify access rights to privileged resources. In order to better understand the control of access to privileged resources, access control principles will be considered in this chapter.

## 5.1 Defining Access Control

Without context, access control is a vague concept and can even describe physical access control to a building or a house. In this chapter access control is considered within the context of computer and information security. Access control can be defined as the protection of access to resources by unauthorized users (Zhang, Li, & Nalla, 2005).

Access control essentially answers the question of "who has which type of access to what?" A good example is a computer user attempting to open

a file on a computer network. If the user has sufficient *access rights*, the user will be *authorized* to open the file. Authorization is thus the positive result of an access control decision. In literature, an access right is also referred to as a privilege or a permission (Sandhu et al., 1996). Access control mechanisms enforce an access control policy which is defined at a more abstract level. The access control policy specifies the authorizations that the access control mechanisms enforce.

An access control authorization can be formalized by the triple: $(s, o, p)$ where $s$ is a subject, $o$ is an object, and $p$ is the permission for the object (Bertino, Ferrari, Buccafurri, & Rullo, 1999). Within the discourse of access control, the entity requesting authorization is referred to as the *subject*. Similarly, the resource to which the *subject* requests authorization, is referred to as the *object*. The type of access right for an object may be specific to the object. Therefore, the permission is required as part of the authorization. In the aforegoing example, the computer user is the subject, the file is the object, and the permission is to *read* the file.

If the file contained sensitive information, such as credit card numbers, the user could potentially misuse such information. In order to prevent this from happening, the "read" access right to the file might not be awarded to the user. The controlling of access rights normally falls into the following two paradigms:

- Discretionary Access Control, and
- Mandatory Access Control.

### 5.1.1 Discretionary Access Control

Discretionary Access Control (DAC) enables the owner of an object to specify the access control policy. In other words, the object owner has the ability to award authorizations to other subjects at his or her discretion (Ferraiolo & Kuhn, 1992). If the computer user in the example above has ownership of the file, then according to DAC-based access control, the user should be able to give another user access to that file.

One of the biggest problems with DAC-based access control systems is the decentralized management of access rights. In a purely discretionary access control system it is impossible to enforce enterprise-level policy because each

owner independently manages his/her owned objects. In contrast, Mandatory Access Control prescribes the opposite.

### 5.1.2 Mandatory Access Control

The Mandatory Access Control (MAC) approach to access control specifies that the system owner or system administrator manages the authorization of all access to all objects by all subjects. This approach is most notably implemented in military systems where confidentiality is of the utmost importance.

A key problem with the MAC approach is that each object within the system needs to be labeled with the authorization level required to access the object. This labeling process is very difficult to maintain and administrate. The most prominent MAC-based access control model is the Bell-La Padula model developed for the U.S. Department of Defence.

Both MAC and DAC conform to *the principle of least privilege* which states that subjects should only be awarded the access rights required to perform authorized tasks. All other access rights are denied by default. Within the context of the aforegoing example, the computer user will not have access to read the file unless such an authorization is added.

The management of capability lists is also troublesome because the introduction of new objects or object access rights are not reflected to authorized subjects.

Different paradigms can be implemented in many ways. For the purposes of this dissertation there is no reason to cover these comprehensively. However, the next three types of models are representative of current business system implementations.

## 5.2 Access Matrix Models

The Access Matrix Model defined by Harrison, Ruzzo, and Ullman (1976) was one of the first and most straight-forward access control models to be used in computer security. The model is also referred to as HRU according to the initials of the authors. Access rights of an HRU-based access control system can be depicted by a two-dimensional array as shown in table 5.1.

Table 5.1: An example of an Access Matrix

| | File1 | File2 | Program1 | Program2 | Alice | Bruce | Carol |
|---|---|---|---|---|---|---|---|
| Alice | RW | | E | | | | |
| Bruce | R | RW | RWE | E | | | |
| Carol | | R | | E | | | |

A row entry exists for each subject within the system and a column entry exists for each object. Furthermore, subjects are considered as objects and have both a row and a column entry within an Access Matrix. Each cell or row-column intersection specifies an access right.

The example in table 5.1 shows the permissions for three users. The user "Alice" has two permissions:

- read/write (RW) access to the file "File1", and
- execute (E) rights to the program "Program1".

If Alice attempted to access the "File1" file the Access Matrix would be traversed for the cell where the row with subject called Alice and the row of object "File1" intersect. According to the matrix given in table 5.1 the relevant cell contains the "RW" access right which means that Alice will be authorized to read from and write to the file.

The HRU model has several limitations. For instance, there is no way to define ownership of objects. A further problem is the inability to apply constraints to the assignment of rights. For instance, it may be that access should only be granted at certain times during the day. Access Control Lists (ACLs) and Capability-lists (C-lists) are derivative implementations of the Access Matrix Model and will be considered next.

### Access Control Lists

An access control list (ACL) is associated with an object and essentially contains a list of permissions. In a DAC-based access control system, each subject has an ACL which is managed by the owner of the subject. Within the Access Matrix shown in table 5.1, every column can be seen as an ACL where the object (e.g. "File1") has a list of permissions (e.g. RW and R) associated with subjects (e.g. Alice and Bruce).

### Capability Lists

A Capability-list (C-list) is similar to an ACL but is located on the subject. It is essentially a list of authorizations about what access rights the subject has to which objects. In table 5.1 every row can be likened to a C-list where every subject (e.g Alice) maintains a list of permissions (e.g. RW and E) to respective objects (e.g. "File1" and "Program1"). Next, lattice-based access control is considered.

## 5.3 Lattice-based Models

Lattice-based access control models aim to model the flow of information (Sandhu, 1993). The Bell-La Padula model as well as various military and multi-level access control models are examples hereof.

In lattice-based models each subject and object are typically assigned a label. This label represents a level of trust. For an object this represents the minimum required trust level of the subject. If a subject meets this minimum trust level, information can "flow" from the object to the subject (i.e. read rights). For information to "flow" to an object (i.e. write rights) the object has to have the same or higher label.

Practically the model has several drawbacks. Not only is it necessary to explicitly label every subject and object in the system, but the labels tend to be very coarse-grained. A subject with a "secret" label does not necessarily need access to all secret documents. Similarly objects created by a subject with "secret" clearance does not necessarily contain "secret" information and need not necessarily be classified as such.

While some of these issues have been addressed by various implementations and additional constructs (such as compartmentalization), lattice-based models remain mostly used in military environments and do not receive much attention in commercial settings.

Next, role-based access control models will be considered, which unlike lattice-based models are popular in commercial settings.

## 5.4 Role-Based Access Control Models

A family of Role-Based Access Control models was proposed by Sandhu et al. (1996). However, the idea of role-based security was first formalized by Ferraiolo and Kuhn (1992). It was found that most enterprizes have unique security requirements that are not met by either the MAC or DAC paradigms.

There are essentially two popularly referenced specifications for RBAC. The older of the two is the family of RBAC models formalized by Sandhu et al. (1996) and commonly referred to as RBAC96. Secondly, the US National Institute of Standards and Technology (NIST) have formalized RBAC which is commonly referred to as the NIST/RBAC standard (Ferraiolo, Sandhu, Gavrila, Kuhn, & Chandramouli, 2001). Although RBAC96 is extended and enhanced in various ways by the NIST/RBAC model, RBAC96 remains popular and robust. Thus, for the purposes of this dissertation, the RBAC96 specification is sufficient.

The main motivation for Role-Based Access Control (RBAC) is the ability to easily enforce enterprise security policies and to simplify security management (Ferraiolo, Cugini, & Kuhn, 1995). Users generally do not have discretionary access to enterprise-level objects. RBAC provides a natural translation of security policy into access control enforcement by means of roles. The concept of implementing roles stems from the realization that within an enterprize, control is governed by employee function rather than ownership (Ferraiolo & Kuhn, 1992).

At a high level of abstraction, a role is a collection of functions that a person within that role can perform. The ease-of-management associated with RBAC is due to the extra layer of roles between objects and users. Moreover, the process of managing users-to-roles and the process of managing roles-to-permissions is done independently.

The family of RBAC reference models introduced by Sandhu et al. (1996) is split into four models, namely:

- $RBAC_0$ : the base model,
- $RBAC_1$ : the base model with role-inheritance,
- $RBAC_2$ : the base model with constraints, and

- $RBAC_3$ : the consolidated model containing both role-inheritance and constraints.

The base model ($RBAC_0$) includes all the essential components of an RBAC system. $RBAC_1$ and $RBAC_2$ both inherit from $RBAC_0$ and cannot be compared directly (Sandhu et al., 1996). However, it is possible to implement the role-inheritance of $RBAC_1$ through constraints. Therefore, the $RBAC_1$ model is not strictly necessary but the concept of role-inheritance has sufficient semantical meaning to be treated in its own. In order for any system to claim adherence to the RBAC model, the complete base model needs to be implemented. The key components of the base model will now be discussed.

### 5.4.1 The Basic Components of RBAC

The RBAC family of models presented by Sandhu et al. (1996) specify the following semantical concepts:

- user: a human being or an intelligent autonomous agent;
- role: a semantic description of authority and responsibility to members of the role;
- permission: an approval of a particular type of access to an object;
- user assignment (UA): a many-to-many relation between users and roles;
- permission assignment (PA): a many-to-many relation between roles and permissions; and
- session: a temporal activation of a subset of roles for a particular user.

At the start of the chapter it was said that a user is mapped to the concept of a subject. However, in RBAC a session is the subject that requires authorization. A user may have many concurrent sessions but each session is only mapped to a single user as shown in figure 5.1. It can therefore be argued that the user is still the subject, albeit through a intermediary session.

A role is an authority or responsibility associated with some semantic description. A user may be mapped to many roles and each role may have

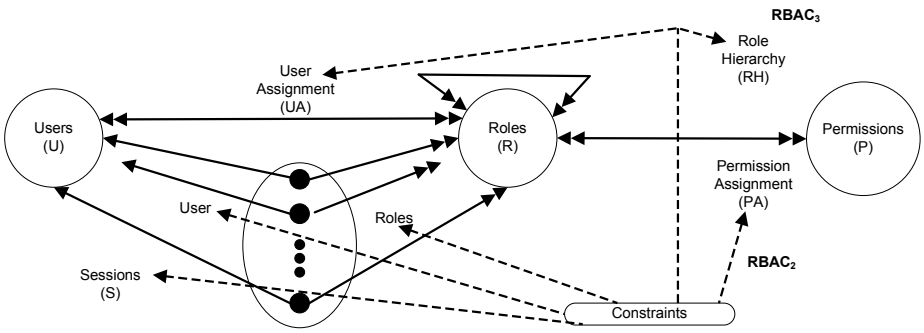


Figure 5.1: Role-Based Access Control model (Sandhu et al., 1996)

many users. A session is created when a user activates one or more roles with which that user is associated.

Permissions were previously defined as a type of access to an object. However, Sandhu et al. (1996) specify that a permission may apply to multiple objects. Within the context of RBAC, permissions are assigned to roles. This is in contrast to the traditional permission assignment where permissions are assigned to subjects.

### 5.4.2 $RBAC_0$: The Basic Model

The following mathematical representation of RBAC formalizes its basic use (Sandhu et al., 1996):

- The $U$, $R$, $P$, and $S$ sets (users, roles, permissions, and sessions respectively),
- $PA \subseteq P \times R$, a many-to-many permission to role assignment relation,
- $UA \subseteq U \times R$, a many-to-many user to role assignment relation,
- $user : S \rightarrow U$, a function mapping each session $s_i$ to the single user $user(s_i)$ (constant for the session's lifetime), and
- $roles : S \rightarrow 2^R$, a function mapping each session $s_i$ to a set of roles $roles(s_i) \subseteq \{r \mid (user(s_i), r) \in UA\}$ (which change with time) and session $s_i$ has the permissions $\cup_{r \in roles(s_i)}\{p \mid (p, r) \in PA\}$

It is expected by Sandhu et al. (1996) that each user will be assigned to at least one role and each role assigned at least one permission. A further

assumption is that only one system owner can change the $U$, $R$, $P$, and $S$ sets and the $PA$ and $UA$ relations. To change these, special permissions called *administrative permissions* are required.

### 5.4.3 $RBAC_2$: RBAC with Constraints

The $RBAC_2$ model introduces the concept of constraints as an extension to the base model. A constraint can be defined as a rule that states whether a specific value to change an $RBAC_0$ component is acceptable (Sandhu et al., 1996).

An example of using constraints is to enforce the principle of "separation of duty". Separation of duty prevents the provisioning of mutually exclusive roles to user. For instance, a user may be not be both a purchasing manager and an accounts payable manager. The combination of these two roles opens an opportunity to commit fraud.

Constraints can be applied to all $RBAC_0$ components including sessions and user assignments. In general, constraints allow the enforcement of higher-level enterprise policies at the access control point. Sandhu et al. (1996) describe constraints in a very informal view, claiming that constraints are very specific to their implementation. However, it is specified that constraints may apply to user assignments, permission assignments, users, or roles.

### 5.4.4 RBAC Versus DAC and MAC

Sandhu et al. (1996) claim that RBAC is sufficiently generic in its handling of authorization to allow it to be configured for use as either DAC or MAC based access control. This was shown to be possible through mathematical formalization (Osborn, Sandhu, & Munawer, 2000). Therefore, in this dissertation MAC and DAC are seen as paradigms for access control and RBAC is seen as an actual model that may or may not be configured to adhere to these paradigms.

## 5.5 Access Control, Privacy and Presence

RBAC has been shown to be applicable to various unintended environments. One such implementation is the use of RBAC to control access to XML

documents (Crampton, 2004). Essentially, encryption is applied to various sections of an XML document with different encryption keys. The encryption is done sequentially to allow a subject hierarchical access to the contents. In the nature of RBAC, the subject is mapped to a role and the role contains permissions (decryption keys) to access the content of the document.

RBAC has also been applied to the domain of privacy. A family of models based on the models formalized by Sandhu et al. (1996) is specified by Ni, Trombetta, Bertino, and Lobo (2007). These models, collectively called P-RBAC are focused on protecting the privacy of customer data with regards to how such information is handled in an enterprize. The goal is to enforce enterprise-level privacy policies within data processing systems.

Access control for presence services has been investigated by Dersingh, Liscano, and Jost (2005). Presence services are seen as any service (e.g. web service) that is defined for a presentity. They then use context information, specifically location information, to authorize access to all services based on the location. An example would be a presentity being in a meeting room. The presentity's current location is then mapped to the object (the meeting room). The presentity is then provided authorization to all presence-based services associated with that meeting room.

The above studies show that access control, such as RBAC, has proven useful in the domains of privacy and presence. Additionally, an implementation of RBAC is given that shows how access within an XML document can be controlled. In chapter 4, it was shown that the IETF presence data format is an XML document. Therefore, it appears that RBAC may be useful in controlling access to information provided to watchers. In contrast to RBAC, both MAC- and DAC-based access control systems are difficult to manage. The simplicity and transparency of managing access control with RBAC may prove to be important for feasibility in controlling access to presence information in IM.

## 5.6 Conclusion

In this chapter access control was defined as the protection of a resource against unauthorized access. Access was shown to be controlled by either the owner of the object (Discretionary Access Control) or the system owner

(Mandatory Access Control). However, both approaches have limitations as shown in the relative sections.

Role-Based Access Control (RBAC) is an access control model that allows easy management of access rights. It adds the notion of roles which are defined as a set of permissions with semantic meaning. Users are associated with roles and roles with permissions. Consequently there is an extra layer of abstraction between users and permissions. Therefore, RBAC maps more directly to the job description and associated access rights as found in an enterprize.

Access control, and RBAC in particular, were shown to be useful in unintended situations. The use of RBAC to control access to various parts of an XML document can be likened to the IETF presence processing model which limits a presence document based on the watcher.

The next chapter, therefore, proposes an enhanced model for presence-handling that considers the current standards for presence handling (chapter 4), the social psychological aspects of presence information (chapter 3), and lessons learned from access control in the context of Instant Messaging.

# Chapter 6

# An Enhanced Presence Handling Model for IM

This dissertation has so far considered the concepts of IM and presence from various perspectives. In chapter 2 the current state of IM was shown. Chapter 3 considered presence from its social psychological roots and how these social factors find their way into IM systems. The IETF standards concerning presence were provided in chapter 4 as the basis for a more formal view of the status quo. In chapter 5 access control principles were related to the authorization of presence information shown in chapter 4.

The purpose of this chapter is to propose an enhanced presence handling model for IM. For the purposes of this dissertation, presence *handling* is not only seen as the *processing* of presence but also the *administration* thereof. Before delving into the details of the model, consider an overview.

## 6.1 Overview

The enhanced presence handling model proposed in this chapter will be developed in progressive stages as shown in table 6.1. Initially, $PH_0$ will be constructed by inheriting the key functionality of the presence handling model of Rosenberg (2005). $PH_0$ will be formalized and considered as a watcher-based presence handling model. The purpose and problems of $PH_0$ will also be discussed.

The next step will be to modify $PH_0$ with the addition of role-based concepts as presented in the RBAC specification by Sandhu et al. (1996).

Table 6.1: Overview of the PH-model development

| Model Progression | Administration | Processing |
|---|---|---|
| $PH_0$ | watcher-based presence | presence filtering |
| $PH_1$ | role-based presence | watcher-to-role mapping |
| $PH_2$ | availability profiles | availability filtering |

This second step will elevate $PH_0$ to become a role-based presence handling model ($PH_1$). In addition to adding roles during the administration of presence rules, the processing model must be adapted to enforce such administration. A formalization of $PH_1$ will be presented and the manner in which $PH_1$ addresses some of the problems of $PH_0$ will be discussed.

Lastly the final model, $PH_2$, will be formulated by adding the concept of an *availability profile* to $PH_1$. The purpose of $PH_2$ is to improve the presentity's ability to better handle incoming messages from watchers. The $PH_2$ model will be conceptualized and discussed as well.

Only the $PH_2$ model references the concept of presentity directly. However, all three stages in model development considers presence-handling from the perspective of the presentity. Therefore, in the first two models, the presentity is not featured in model formalization. The watcher-based presence handling model ($PH_0$) will now be considered.

## 6.2 $PH_0$: Watcher-based Presence

The presence handling model, $PH_0$, now proposed, is strongly based on the model proposed by Rosenberg (2005). For reading convenience, the presence handling model proposed by Rosenberg (2005) will be referred to as the Rosenberg-model. The Rosenberg-model describes presence handling in the SIP and SIMPLE protocols and provides much more detail than is relevant for the purposes of the $PH_0$ model. Therefore only the essential components and processes have been retained to form $PH_0$.

In order to formalize the $PH_0$ model, the following sections provide a specification of the environment and a description of the processing of presence information in the context of $PH_0$. This format will be followed throughout the chapter, as the model progresses.

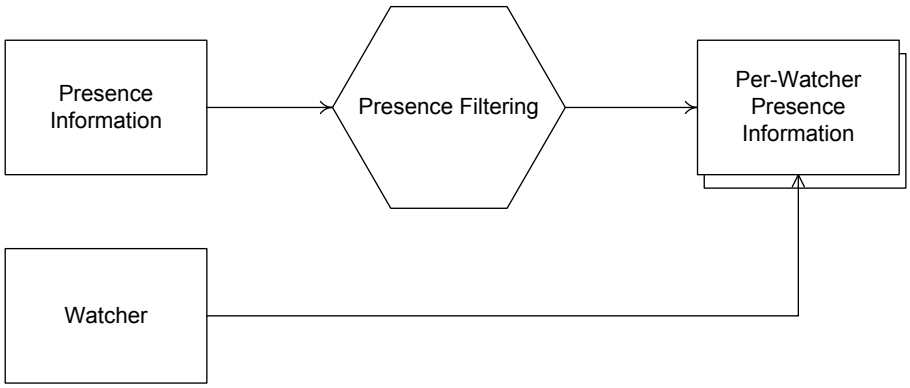


Figure 6.1: Architectural view of $PH_0$, adapted from the Rosenberg-model

### 6.2.1 Administration

The Rosenberg-model was introduced in chapter 4. It was shown that the Rosenberg-model filters presence within the context of the IETF's data formats for presence information (the PIDF and RPID document formats). The PIDF and RPID document formats were shown to include a vast set of possible presence information. Such presence information was shown in chapter 3 to be of a sensitive nature, leading to privacy concerns. The Rosenberg-model allows presence to be filtered on a per-watcher basis, which is a major benefit of the model over most major IM implementations. While not widely implemented, the Rosenberg-model describes the IETF's presence handling with regards to the SIP and SIMPLE protocol. Thus it is considered the status quo.

Figure 6.1 shows an architectural view of the $PH_0$ model, adapted from the more detailed model as represented in figure 4.5 on page 52. However, figure 6.1 describes the most crucial part of the Rosenberg-model, namely presence filtering on a per-watcher basis.

There are three basic concepts in the $PH_0$ model namely, watchers, presence, and subscriptions. These can be formalized as follows:

- $W$, $P$, and $S$ representing watchers, presence attributes, and subscriptions; and
- $PA \subseteq P \times W$ representing presence-watcher assignments.

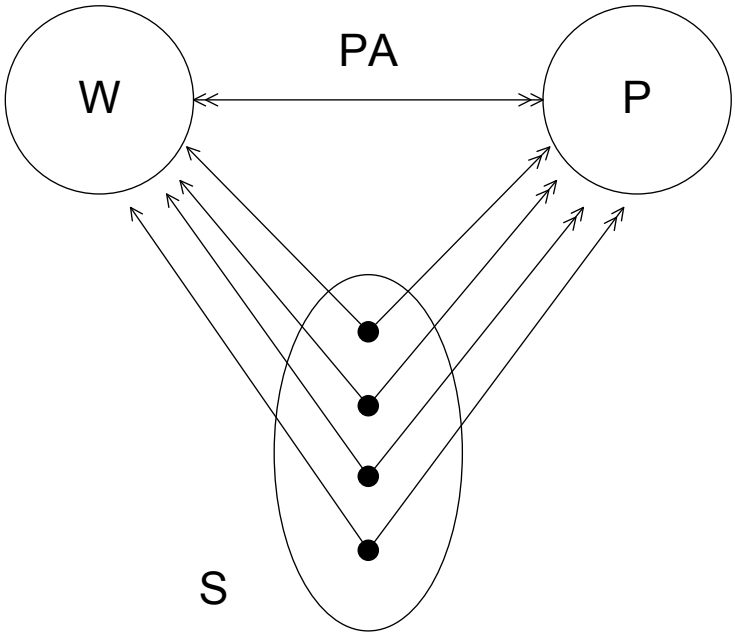


Figure 6.2: Graphical view of $PH_0$

### 6.2.2 Processing

A graphical view of $PH_0$ is given in figure 6.2 and shows that presence is directly associated with watchers in a many-to-many relationship. In figure 6.2 this relationship is represented by a double-headed arrow between the watchers and presence entities. In other words, a watcher can have access to various presence attributes, and every presence attribute can be provided to many watchers. The formalization of this processing can be represented as follows:

- $watcher : S \rightarrow W$ a function mapping each subscription $s_i$ to a single watcher $watcher(s_i)$; and
- *presence attribute*: $S \rightarrow 2^P$ where each subscription $s_i$ is mapped to a set of presence attributes where *presence attributes*$(s_i) \subseteq \{p \mid (p, watcher(s_i)) \in PA\}$.

A watcher obtains a subscription for a presentity's presence information after authorization by the presentity. Thus, every presence subscription is associated with only one watcher. For the duration of the subscription, the watcher is mapped to a set of presence attributes. Every presence attribute

to be provided to a watcher needs to be authorized in the presence policy. Access to all other attributes is implicitly denied.

### 6.2.3 Discussion

The key difference between the processing of the $PH_0$ model and most current implementations of presence handling is its per-watcher filtering of presence information. This difference addresses an important issue within current IM systems, namely the one-for-all approach to handling watchers. However, there are some fundamental flaws in the model. These flaws can be summarized as follows: An undue burden is placed on (a) the presence server in the generation of presence documents and on (b) the presentity in authorization rule creation for controlling access to his presence information.

Server burden relates specifically to the apparent need to generate a presence document for each subscribed watcher. This may not be a big issue for a small number of watchers, but the processing burden will surely increase as the watcher base grows. This will definitely be a reality for companies with employee and customer bases in the hundreds or thousands.

The number of potentially subscribed watchers can also place a burden on the presentity. Maintaining control over who gains access to presence information and the amount of presence information provided is of prime concern to a presentity. However, as a presentity's watcher base grows so too does the need for additional authorization rules to ensure control over presence information.

In chapter 4 it was shown that the "sphere" condition can be used to provide different presence information to the same watcher. However, presence rules for each sphere will have to be defined for each watcher, creating an unmanageable presence policy. In the future there may be other such conditions further complicating management of a presence policy.

In chapter 5 the benefits of the RBAC model for access control systems were mentioned. It was said that RBAC simplifies administration and provides a more natural mapping between security policy and enforcement. It was also found that a presence policy is very similar to a security policy in that both control access through authorizations. Therefore, in the following stage, RBAC concepts will be applied to $PH_0$, resulting in $PH_1$.

Table 6.2: Mapping of RBAC-to-$PH_1$ concepts

| RBAC | $PH_1$ Equivalent |
|---|---|
| User | Watcher |
| Permission | Presence |
| Session | Subscription |
| Permission-Assignment | Presence-Assignment |
| Role | Role |
| User-Assignment | Watcher-Assignment |

## 6.3 $PH_1$: Role-based Presence

In the Role-Based Access Control (RBAC) model, users are associated with roles, and roles with permissions. A detailed description of these concepts was given in chapter 5. The concepts of $PH_0$, namely watchers, presence, subscriptions, and presence-assignments can be mapped to their semantic equivalents in the RBAC specification. The mapping of RBAC concepts to $PH_1$ is shown in table 6.2. Therefore, the addition of role-based concepts to $PH_0$ can be formalized by drawing on the formalization of the RBAC model.

### 6.3.1 Administration

The specification of $PH_1$ extends $PH_0$ by adding roles. The architectural view of $PH_1$ is given in figure 6.3. The diagram shows that instead of a watcher directly accessing presence information, the watcher is first mapped to a role. The $PH_1$ concepts can be formalized as follows:

- $W$, $P$, and $S$ are unmodified from $PH_0$;
- $R$ and $WA$ are added, representing roles and watcher-to-role assignments respectively;
- $WA \subseteq W \times R$;
- $PA$ is modified from $PH_0$ to contain presence attribute-to-role mappings where $PA \subseteq P \times R$.

The most notable change in $PH_1$ from $PH_0$ is the introduction of roles (R). Roles are defined by Sandhu et al. (1996) as a set of permissions. Similarly, a role in $PH_1$ is defined as a set of presence attributes. In $PH_0$ such a presence attribute is assigned directly to a watcher with a subscription.

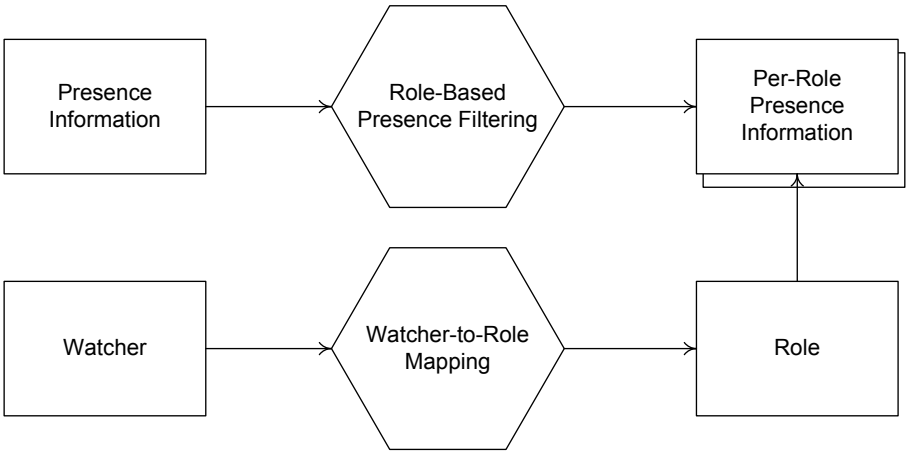


Figure 6.3: Architectural view of $PH_1$

In contrast, a presence attribute is assigned to a role in $PH_1$. The processing details are provided next.

### 6.3.2 Processing

Figure 6.4 graphically depicts the specifics of the $PH_1$ model. The figure shows that a watcher can be assigned multiple roles, and that a role may be assigned many watchers. If a watcher activates several roles during the same subscription, scenarios are likely where multiple roles can cause conflicts in the sets of presence information approved for a watcher. The same issue exists on per-watcher filtering of presence such as in $PH_0$. This issue has been addressed in the IETF specification on *expressing privacy policy* (RFC 4745) but the resolution is implementation specific. Therefore, the issue need not be addressed by $PH_1$.

The processing specific details of $PH_1$ can be formalized as follows:

- $roles : S \rightarrow 2^R$ where
- each subscription $s_i$ is mapped to a set of roles $roles(s_i) \subseteq \{r \mid (watcher(s_i), r) \in WA\}$ and
- subscription $s_i$ has the set of presence attributes $\cup_{r \in roles(s_i)} \{p \in (p, r) \in PA\}$.

Similar to $PH_0$, $PH_1$ processes presence information based on the subscription with which a particular watcher is associated. It should be noted

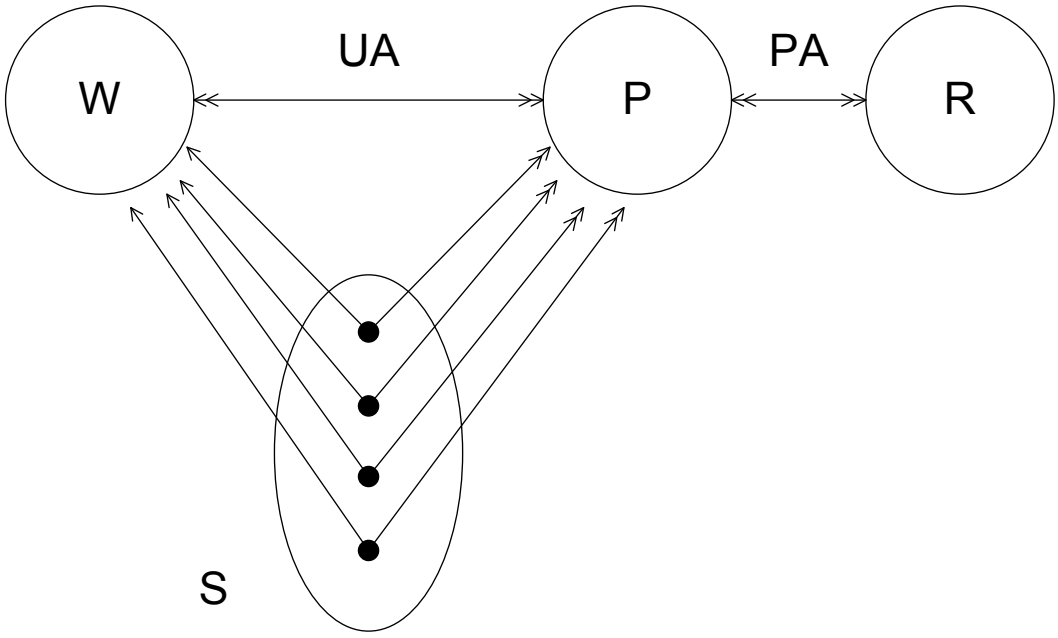


Figure 6.4: Graphical view of $PH_1$

that a presentity's presence information is in a state of flux, continuously changing and updated. In contrast, a subscription does not change as often. An important distinction should be made between the presence information authorized within a subscription and the actual received presence information. The latter may change during the lifetime of a subscription.

### 6.3.3 Discussion

Ferraiolo and Kuhn (1992) state that the concept of roles stems from the realization that in an enterprise control is governed by an employee's role and function. In chapter 3 it was said that the identity awareness of another affects the behavior of oneself. Thus, a presentity behaves differently, based on the identity of the watcher. In $PH_0$ the control of presence information is aligned with this behavior. However, as the number of watchers increase, it is not feasible for the presentity to maintain distinct per-watcher handling. Therefore, $PH_1$ introduces roles to group watchers according to their "function and role" (Ferraiolo & Kuhn, 1992) with regards to the presentity. In other words, a watcher is organized according to the relationship that the

presentity has with that watcher. In chapter 2 it was said that IM users (presentities) and their contacts (watchers) normally know one another outside of the Internet. Hence, a relationship normally exists between the presentity and the watcher.

The application of roles to $PH_0$ results in an improved, role-based presence handling model. It provides a presentity significant advantages in scalability and management of watchers. Also, the presence service only needs to process presence information for the set of roles which alleviates system processing resources. However, the ability to filter presence information for a watcher, albeit through a set of roles, does not take into account the needs of the presentity. The following step considers these.

## 6.4 $PH_2$: Availability Profiles

The interruptive potential of IM communication was stated in chapter 2. Furthermore, the low cost of sending a message combined with the Internet facilitating constant connectedness increases the chances of being disturbed, especially in the workplace.

It was pointed out in both chapters 2 and 3 that although IM was interruptive, it was less so than telephone and face-to-face communication due to its implementation of presence information. If a presentity conveys his/her presence state as "busy", a watcher may interpret that it is not currently the best time to initiate an interaction and spare the presentity the possible interruption.

The problem with providing a presence state such as "busy" to watchers, is the total reliance on the watchers' interpretation or consideration of such information. Although the presentity implies availability and willingness with a presence state, it is not enforced on the presentity's side of the interaction. The concept of *availability profiles (AP)* provides such a mechanism.

### 6.4.1 Administration

An architectural view of $PH_2$ can be seen in figure 6.5. The diagram shows that $PH_1$ has been extended with the addition of an availability filter connecting the watcher (W) and presentity (Pr). The input to the availability filter from the watcher-side is the presence information provided to the

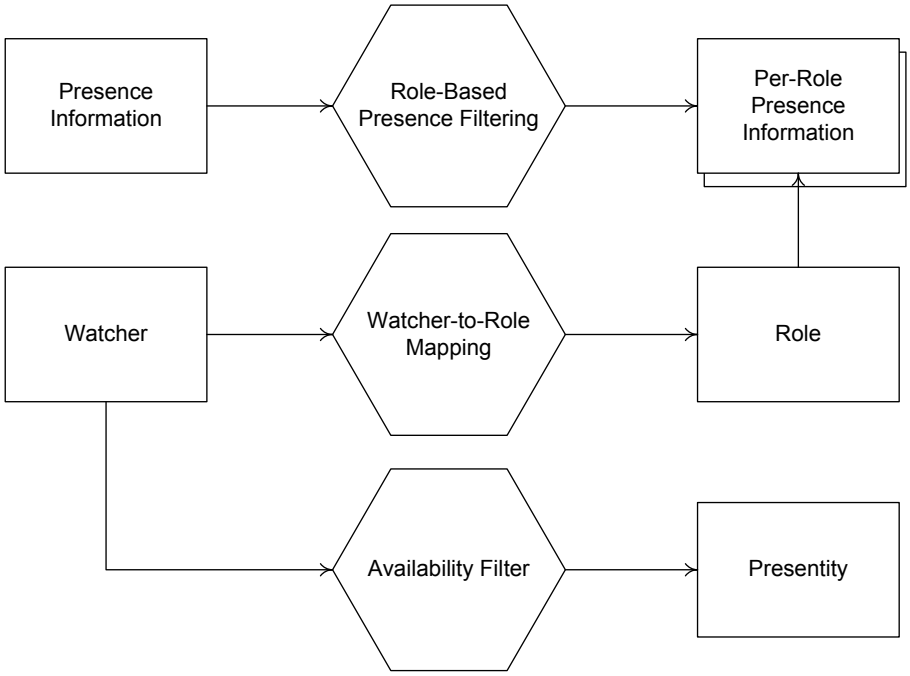


Figure 6.5: Architectural view of $PH_2$

watcher. In particular, the presented availability information, as embedded in the presence state, is of concern. The availability filter is used to provide the presentity with an availability profile with which incoming messages from a particular watcher is handled.

An availability profile can be defined as the availability and responsiveness with which any incoming IM message is handled. The purpose of an availability profile is to help the presentity handle an incoming message with the appropriate amount of responsiveness. Responsiveness was defined in chapter 2 as *demonstrated availability*.

The $PH_2$ model can be defined as follows:

- $PH_2$ is unmodified from $PH_1$ except for the addition of availability profiles (AP).
- An availability profile defines how to handle an incoming watcher-message based on the availability-related presence information provided to that watcher.

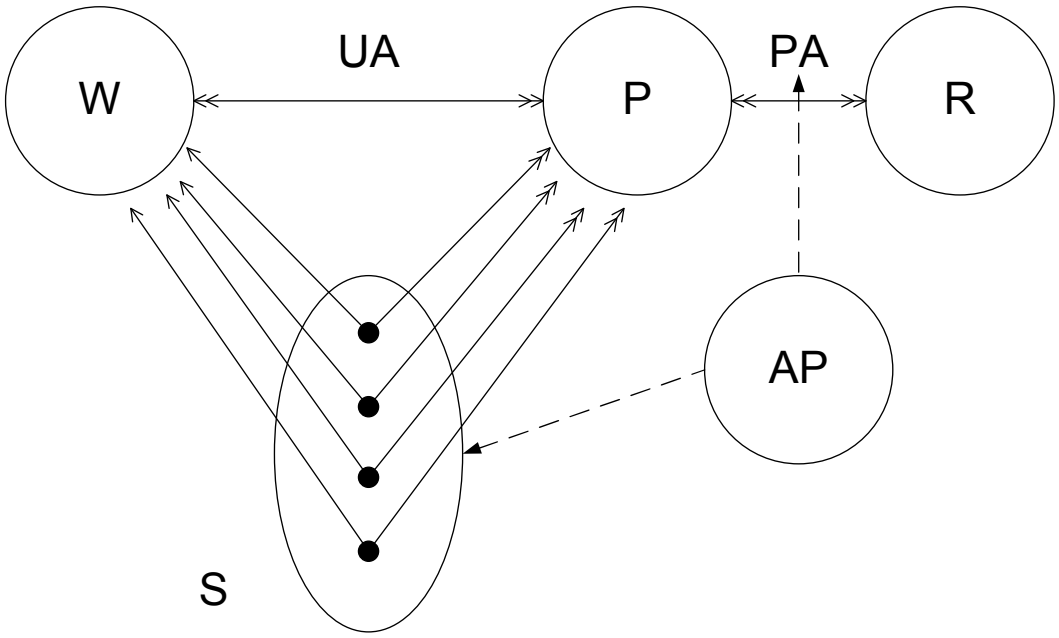


Figure 6.6: Graphical view of $PH_2$

### 6.4.2 Processing

The processing of $PH_2$ extends $PH_1$ as shown in figure 6.6. The model does not specify how an availability profile is to be implemented in an IM system. However, it can be likened to a real-life working environment where person A is working in his office. If person B enters the office unannounced, it may cause interruption even without communication. However, if person B leaves a note in front of the office, person A will be spared the interruption.

In the example, it can be said that person A presented a presence state of "busy" to all. However, the door was not locked, leaving potential for handling emergencies as well as the potential of unwanted interruptions. An availability profile can be likened to the process of locking the door and providing the key to a specific set of people. The people that have keys, will be able to enter person A's office, implying high availability. Similarly, to the people without keys, person A can be seen as busy but can be reached by leaving a message at the door. An availability profile can thus help person A to demonstrate the appropriate amount of availability.

The availability filter uses the set of presence information as authorized by the presence-assignments of the watcher within a subscription to produce

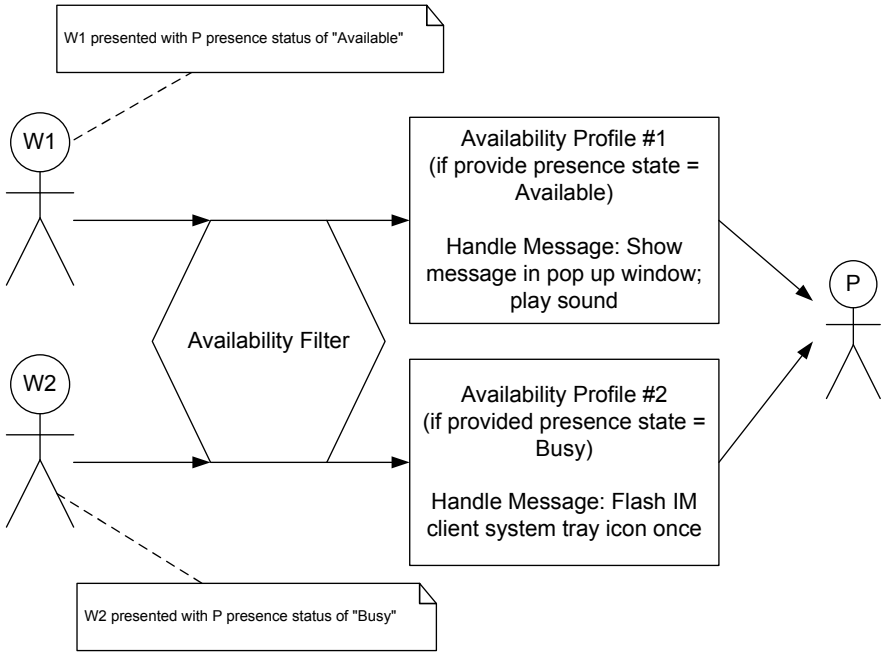


Figure 6.7: Example of $PH_2$ potential implementation

an availability profile as mentioned in the specification section. Furthermore, the availability profile may modify presence information regarding availability before it is presented to the watcher. This association between the availability profile and presence-assignments is indicated with a broken-line arrow in figure 6.6. Availability profiles can be likened to constraints in RBAC in that both can limit values. However, availability profiles can modify actual presence attributes and also describe watcher message handling behavior on the presentity's side, none of which can be achieved with RBAC constraints.

### 6.4.3 Discussion

The QnA prototype presented by Avrahami and Hudson (2004) can be described in the context of availability profiles. QnA is based on a model for predicting responsiveness in IM. Every incoming message is processed to see if it contains a question. If a question is received, the message is displayed more saliently than if it is not. This saliency can be considered an availability profile where questions are treated with a different availability profile than non-questions.

In the $PH_2$ model an availability profile is associated with a subscription,

and with a watcher through implication. In more general terms, as long as a watcher is provided presence information within a certain combination of authorizations, messages from the watcher will be treated in a certain way. One such implementation possibility is to associate the availability profile with the presented presence state. Furthermore, the availability profile can then use the presented presence state to handle incoming messages with a certain salience. Figure 6.7 illustrates such an implementation example. The background to figure 6.7 is that watcher W1, and watcher W2 are communicating with presentity P. The figure shows that different presence states have been presented to each watcher, i.e. "Available" to W1 and "Busy" to W2. The availability profiles (AP1 and AP2) handle incoming messages based on the presented presence state of P. AP1 allows a watcher's message to be displayed saliently in the form of a pop up window and a sound alert if the presented presence state is "Available" (indicating high availability). Similarly, AP2 handles incoming messages to watchers that were presented with a presence status of "Busy". Therefore messages from W2 will be handled inconspicuously by only flashing the IM client icon in the system tray once. The end result is that the presentity is helped in demonstrating the right amount of availability through the level of salience with which the message is displayed.

## 6.5 Conclusion

In this chapter an enhanced presence handling model for presence in IM was proposed. The model was developed in three stages, the first an adaptation from the current presence handling model ($PH_0$).

The second progression of the model was named $PH_1$, and extends $PH_0$ through the application of Role-Based Access Control principles. The addition of roles simplifies the management of presence policy for the presentity and improves processing efficiency on the server side.

In $PH_2$, $PH_1$ is enhanced with the addition of availability profiles. An availability profile helps the presentity to demonstrate the amount of responsiveness that was conveyed to the watcher at the time of an incoming message. Availability profiles can, for example, be implemented to display incoming messages with different levels of salience.

All three models were described in a set-based manner, similar to the specification of the family of RBAC models (Sandhu et al., 1996). Additionally, the broader architectural view of each model was given to support the processing view. In the next chapter a prototype demonstrates the features of the model.

# Chapter 7

# Prototype

RoBIM is short for Role-Based Instant Messenger and essentially provides a standards-based implementation of the role-based presence handling model defined in chapter 6. Additionally it implements several other features commonly found in current IM systems.

The RoBIM prototype was developed to accomplish three goals:

- to impart a better understanding of the problem domain, the research problem and potential solutions;
- to show the feasibility of the role-based presence model defined in chapter 6; and
- to be used in future related research within the discourse of multi-channel communication and presence management.

RoBIM has undergone three iterations which have collectively supported the goals of the prototype. The first generation prototype ties in with the first goal in that it was used to explore IM and presence and the current limitations of presence-handling. The second generation RoBIM was focused on illustrating how the presence-handling problem can be solved by implementing the presented role-based presence model. Lastly, the third generation RoBIM prototype was a more polished and concise improvement of the second generation prototype. The last two prototypes were both built using the SIMPLE and SIP specifications. However, the third generation prototype provides a more extensible and re-usable system architecture to support the third goal mentioned above. Unless specified otherwise, reference to the RoBIM prototype will imply reference to the third generation prototype.

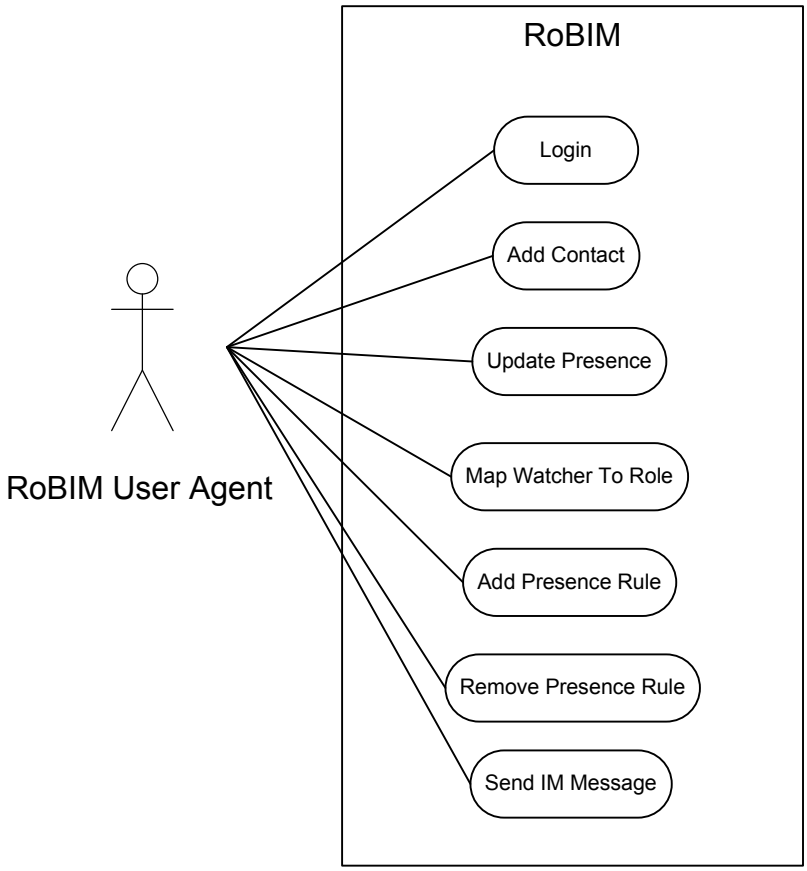


Figure 7.1: RoBIM user agent use case diagram

The rest of this chapter will look at the design decisions, especially with regard to the aspects that received focus. The intentionally neglected aspects will also be mentioned. However, the most attention will be given to the implementation of the role-based presence model and its translation into the SIMPLE-based IM system. To illustrate the feasibility of the implemented model a scenario-based discussion will be conducted. This chapter commences with an overview of the RoBIM prototype.

## 7.1 Prototype Overview

The prototype can be split into two entities at the systems view level: the user agent which is the IM client, and the stateful proxy which is the SIP server with SIMPLE extensions. Although the server-side is important for the more technical details, the IM client provides a well-organized perspective

on the functionality of the prototype. Figure 7.1 shows the use cases for the user agent.

The use cases are organized into the following RoBIM client components:

- login (the "Login" use case);
- contact List (the "Add Contact" and "Send IM Message" use cases);
- presence manager ("Update Presence" use case);
- watcher manager ("Map Watcher to Role" use case); and
- presence rule manager (the "Add Presence Rule" and "Remove Presence Rule" use cases).

Every RoBIM client components will now be considered in turn, with the purpose of describing the use cases related to each.

### 7.1.1 Login

The login screen can be seen in figure 7.2 and allows a user to register with a specified SIMPLE server. The implementation of the RoBIM prototype did not extend to the complexities of using DNS for cross-Internet or network domain communication. The domain name provided is only used as an identifier by the SIMPLE server to ascertain whether the incoming registration request is within the *domain-of-influence* for that SIMPLE server.

A SIP or SIMPLE server that handles registrations is called a registrar. In the case of the RoBIM prototype all logical SIP and SIMPLE roles were implemented within the same server. However, at the lower level these were implemented in separate extensible components.

### 7.1.2 Contact List

The contact list contains most of the commonly found functionality within an IM client. It covers three of the UA use cases shown in figure 7.1: adding contacts, observing presence, and sending instant messages. The contact list is depicted in figure 7.3.

An user acts as a watcher when adding contacts and observing presence. In accordance with the IMPP model defined in RFC 2778, a watcher subscribes to the presence of the presentity (Day et al., 2000). For the purposes

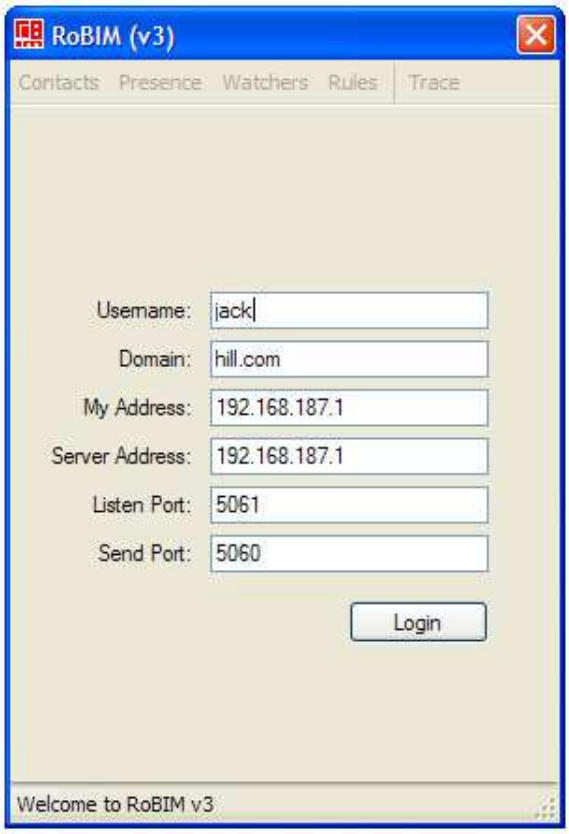


Figure 7.2: RoBIM IM client login screen

of the prototype, subscriptions are always approved automatically without obtaining consent from the presentity. Subscription approval is followed with an immediate update of presence information. Furthermore, the server will notify the watcher whenever a presentity's presence information changes.

Instant messages are sent from a chat dialog window which is opened when a contact is double-clicked on the contact list. The pop-up chat dialog allows the user to type text-based messages and send them to the contact.

### 7.1.3 Presence Manager

The presence manager is used to update the presentity's presence information on the presence server. In the case of RoBIM, the SIMPLE server implements the logical role of presence server. The interface to manage presence is shown in figure 7.4. Within the bigger picture at least some of your presence information could be inferred from other areas. As discussed in chapter 2, some IM systems do this by showing "Away" if their was inactivity on the

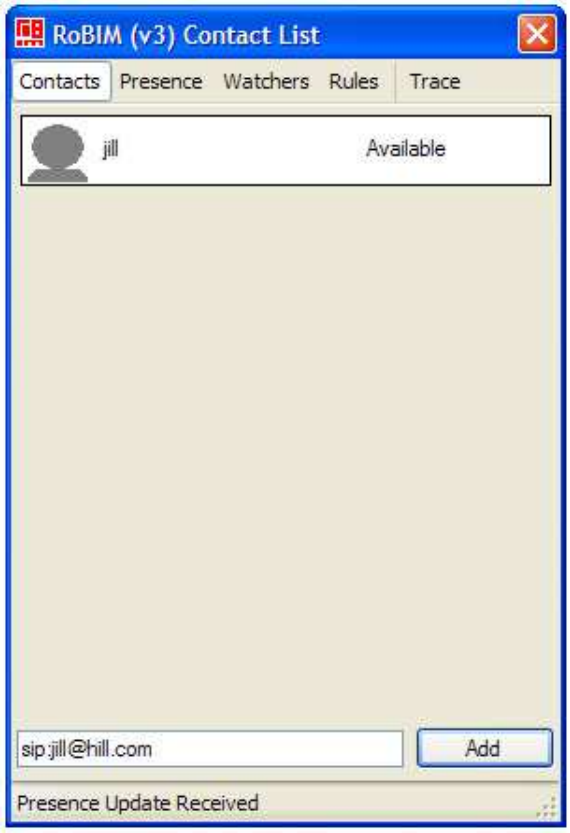


Figure 7.3: RoBIM contact list

computer for a specific time.

As a design decision, a presentity's presence information is limited to *basic status*, *location*, and *activity*. All fields have system-defined states and these are all shown in table 7.1.

Presence information is communicated between presentities, presence servers, and watchers via the IETF IMPP working group's Presence Information Data Format (PIDF) defined in RFC 3863 (Sugano et al., 2004). Both *location* and *activity* are specified in the PIDF specification.

Table 7.1: The implemented presence fields and corresponding states

| Presence Field | Values | | | |
|---|---|---|---|---|
| Basic State | Available | Away | Busy | Do Not Disturb (DND) |
| Location | Home | Work | On Holiday | |
| Activity | Working | Idle | On Telephone | |

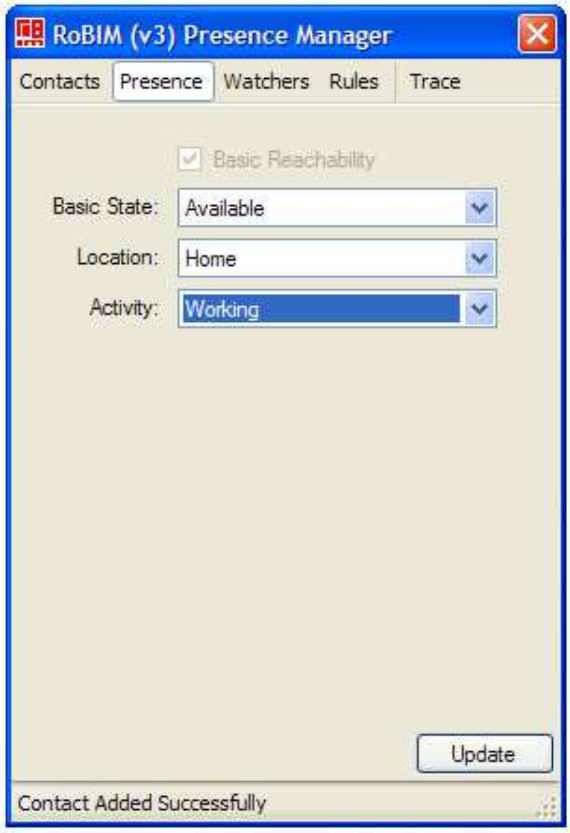


Figure 7.4: RoBIM presence manager

### 7.1.4 Watcher Manager

The watcher manager allows a presentity to map watchers to roles. The interface to manage watchers is shown in figure 7.5. The ability to manage or even observe watchers is a feature that is not commonly found in IM systems. However, it was essential for the RoBIM prototype and is supported by the SIMPLE specifications (Rosenberg, 2004b).

In figure 7.5 the implemented roles are shown as *friend*, *family*, and *colleague*. The chosen roles represent the virtual boundary of work and social spheres. The implementation of the role-based presence model itself does not limit the number of roles. The translation of the role-based presence model will be covered in detail in a later section.

It should be noted that the management of watcher-lists does not execute within the SIP infrastructure but via a protocol called XML Configuration Access Protocol (XCAP) which allows documents to be manipulated over HTTP.

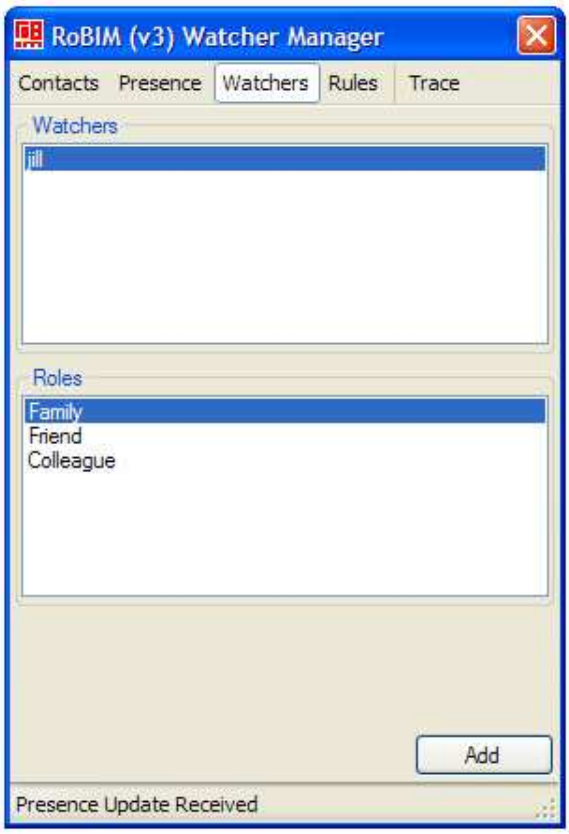


Figure 7.5: RoBIM watcher manager

### 7.1.5 Presence Rule Manager

The presence rule manager is the primary focus of the RoBIM prototype. It shows the current presence rules and initiates the presence rule designer interface. The presence rule manager is shown in figure 7.6. As the diagram illustrates presence rules are defined on a per-role basis.

The specification of presence rules in RFC 5025 does not currently address roles or groups and was carefully extended to facilitate role-support as specified in $PH_1$.

Another RoBIM enhancement to the presence rules specification is the ability to not only control permissions but also modify presence information. The RoBIM prototype allows the presentity to specify a single *basic status* which is then forwarded and modified based on watcher-role mappings.

The main strength of role-based presence handling is that the presentity does not need to continually manage a set of presence information for each watcher. However, to maintain presence information sets for each role also

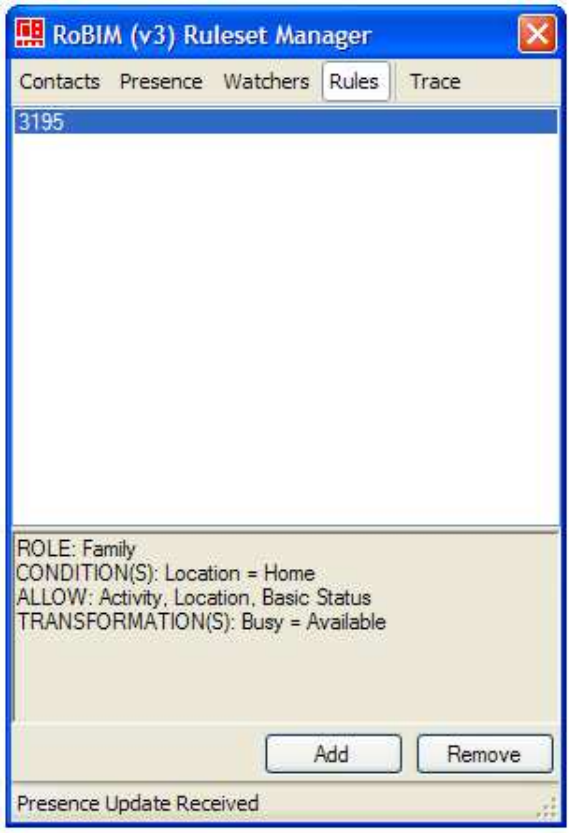


Figure 7.6: RoBIM presence rule management screen

leads to an unsustainable amount of effort. Therefore, the RoBIM prototype increases the flexibility of rules by allowing the separate and complete transformation of presence information (as specified in the $PH_2$ model) on a per-role basis (as specified by $PH_2$).

## 7.2 Standards-based Approach

The design decision to support the IETF's SIMPLE specifications was made for various reasons. A standards-based approach supports all three prototype goals in the following ways:

- The investigation of IETF specifications has fostered a comprehensive understanding of IM and presence.
- The SIMPLE protocol already includes specifications that are core to the role-based presence handling model described in this dissertation.

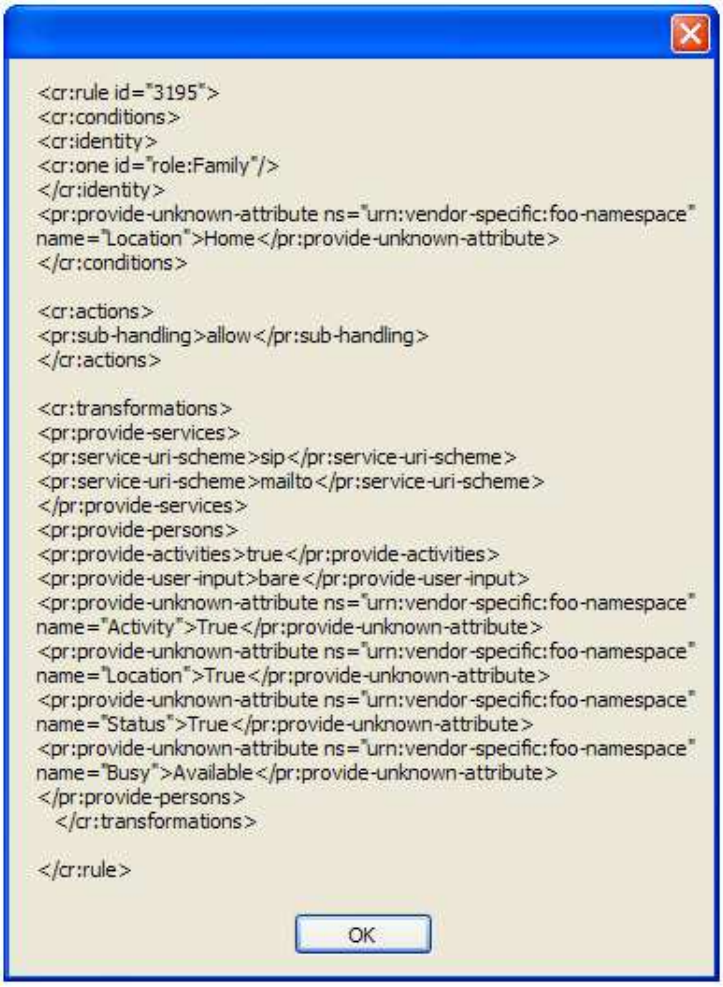


Figure 7.7: RoBIM presence rule

- The potential future usefulness of the prototype is greatly enhanced by its authenticity in terms of supporting IETF specifications.

Although the SIMPLE protocol is the main focus, a thorough understanding of its relationship to the IMPP working group and the SIP protocol is required. The interdependence of the IMPP working group, SIP, SIMPLE, and the prototype was discussed in chapter 4. In the following section the implementation of the SIP protocol in RoBIM is discussed.

### 7.2.1 The Session Initiation Protocol

The SIP protocol was briefly discussed in chapter 4. SIP is essentially a session-establishing protocol which facilitates multi-channel communication. Although there are third-party SIP components available, it was decided to implement SIP from the ground up into the RoBIM prototype. One of the reasons was to abstain from dependencies on third-party code libraries which could hinder and complicate future prototype development.

Most of the core SIP protocol specification was implemented to some extent. The major exception being the *dialog*-related sections. The SIMPLE protocol does not require SIP dialogs to be created for IM communication when IM messages are sent in *page-mode*. However, the foundation for adding dialog-based IM communication does exist in the current prototype for future extension.

#### Request-Response Model

SIP uses a request-response model for communication between SIP elements. In other words, a SIP message is either a request from a User Agent Client (UAC) or a response from a User Agent Server (UAS).

There are six types of requests specified in RFC 3261: REGISTER, INVITE, ACK, OPTIONS, CANCEL, and BYE. A request type name such as "REGISTER" is referred to as a SIP *method*. Of these SIP methods only REGISTER is actually used within the RoBIM prototype.

A response message contains the result of the request that was provided to a UAS. All response messages contain a response code which explains the response, as well as a human-readable response message.

**Layered Protocol**

RFC 3261 defines SIP as a layered protocol (Rosenberg et al., 2002). These layers are logical and may be implemented across several components. The four layers are:

- Syntax and encoding layer which is concerned with the correct form of requests and responses as well as the UTF-8 character encoding and decoding of SIP messages.
- Transport layer which is responsible for sending SIP messages across the network as well as receiving messages and passing them to the transaction layer.
- Transaction layer which keeps track of requests and their corresponding responses.
- Transaction user layer which constructs requests and responses and sends them to the transaction layer to be sent. It also interfaces with the user, showing feedback concerning transactions.

All four of these layers were implemented by the RoBIM prototype in an extensible object-oriented way. RoBIM is therefore designed to be altered and improved upon according to future requirements. It should be noted that the RoBIM SIMPLE server is a stateful proxy because it contains a transaction layer. Stateless proxies do not process incoming messages by creating transactions but simply responds or forwards them in a passive way.

### 7.2.2 SIMPLE

The SIMPLE protocol was introduced in chapter 4 as a collection of specifications from the IMPP, SIP, and SIMPLE working groups. In this section a description of the implementation of the SIMPLE protocol in the prototype is given.

The SIMPLE specifications can be divided into specifications that extend SIP for IM, and specifications that extend SIP for presence. Figure 7.8 shows the IM specifications while figures 7.9 and 7.10 show the presence specifications. These diagrams were shown in chapter 4 but have been modified to show which specifications were implemented in part of full.

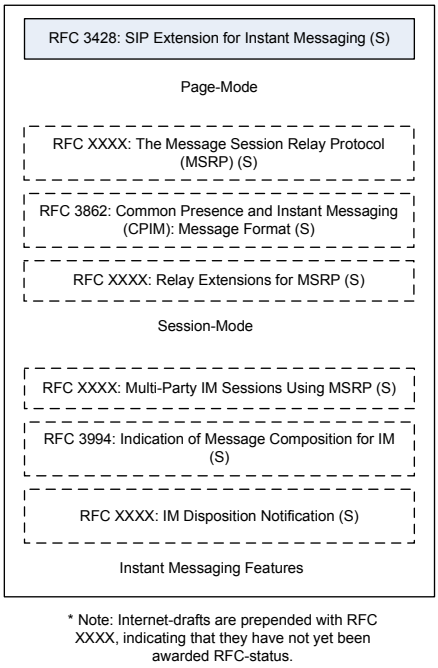


Figure 7.8: IETF SIMPLE specifications related to IM

### IM

In RFC 3428 the MESSAGE method is specified as an extension to SIP (Campbell, Rosenberg, Schulzrinne, Huitema, & Gurle, 2002). This allows out-of-dialog SIP requests to be used to carry instant messages. The MESSAGE request supports two kinds of message-body data (both are MIME types): text/plain and cpim/application. Only the former is implemented in the RoBIM prototype. An example of a MESSAGE request is shown in figure 7.11.

### Presence

The specifications relating to the core protocol machinery are shown in figure 7.9. The SIP subscribe/notify mechanism defined in RFC 3265 specifies the SUBSCRIBE and NOTIFY methods for SIP (Roach, 2002). The subscribe/notify event framework is used in RFC 3856 to enable watchers to subscribe to presentity presence information and be notified when subscribed presence information is updated (Rosenberg, 2004a). RFC 3903 allows presentities to update their presence information via another method extension to SIP called PUBLISH (Niemi, 2004). All of these mentioned were imple-

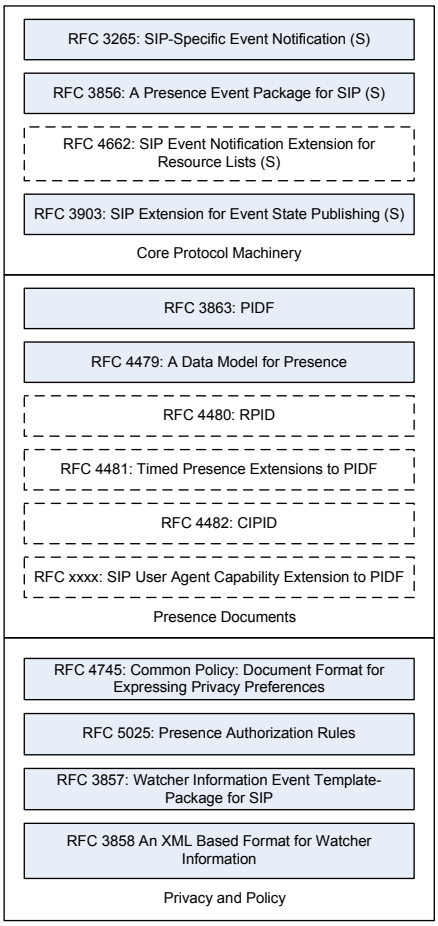


Figure 7.9: IETF SIMPLE presence specifications related to core protocol machinery, presence documents, privacy, and policy

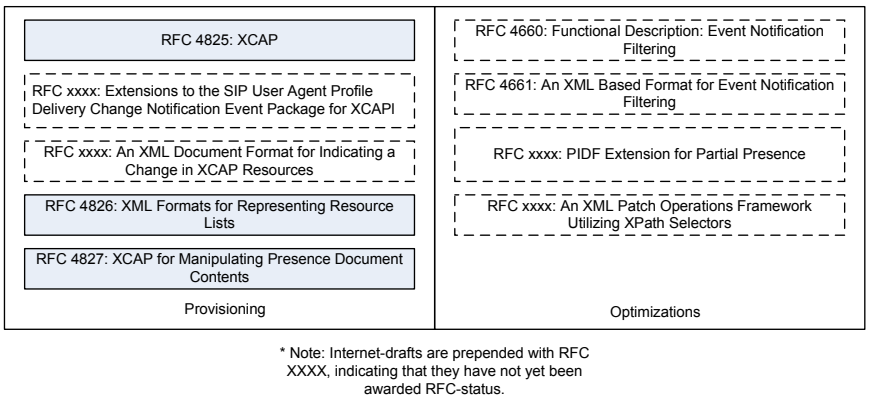


Figure 7.10: IETF SIMPLE presence specifications related to provisioning and optimization

```
MESSAGE sip:jill@hill.com SIP/2.0
Via: SIP/2.0/TCP 192.168.187.1;branch=z9hG4bK5591583
Max-Forwards: 70
To: sip:jill@hill.com
From: sip:jack@hill.com;tag=7829625
Call-ID:4560433
CSeq: 1 MESSAGE
Content-Type: text/plain
Content-Length: 10

Hello Jill
```

Figure 7.11: SIP MESSAGE request

mented in the RoBIM prototype.

"Presence documents"-related specifications include the PIDF (RFC3863) and RPID (RFC4480) data formats. These two document formats serve as a guide to the type of detailed information that presence services will be conveying to watchers in the near future. However, only the PIDF format was used in RoBIM and only in a very simplistic manner.

The detail and comprehensiveness of PIDF and RPID presence documents exemplify the need for proper presence rule policies and privacy management. Such privacy and policy related specifications are shown in figure 7.9. Of most interest is the presence authorization rules specification in RFC 5025 (Rosenberg, 2007). It contributes fundamentally to the purposes of this prototype as it is used to apply presence filtering in a role-based fashion. The presence authorization rules specification is itself based on a more generic privacy policy specification defined in RFC 4745.

Provisioning and optimization related specifications are shown in figure 7.10. Both groups were purposefully neglected with the exception of the XCAP protocol (RFC 4825 and RFC 4826) which is used to manipulate XML resource lists on a server via the HTTP protocol.

The implementation philosophy for all implemented specifications was to *implement as little as possible without sacrificing accuracy.* Together with the implementation philosophy, RoBIM also pursued a design philosophy of *design with an emphasis on extensibility and efficiency* so the prototype can easily be adapted for future research. The following discussion will describe how RoBIM implements the proposed $PH_2$ model.

```
PUT /resource-lists/users/sip:jack@hill.com/index HTTP/1.1
Content-Type:application/auth-policy+xml Host: xcap.hill.com

<?xml version="1.0" encoding="UTF-8"?> <resource-lists
xmlns="urn:ietf:params:xml:ns:resource-lists">
 <list name="Friend">
  <entry uri="sip:jill@hill.com">
   <display-name>jill</display-name>
  </entry>
 </list>
</resource-lists>
```

Figure 7.12: HTTP XCAP resource-list document update

## 7.3 Implementation of $PH_2$

The RoBIM prototype demonstrates the features of the proposed presence handling model for IM ($PH_2$). Its additional functionality as a valid Instant Messaging system allows various scenarios to be acted out and used for stimulating thought on presence information, presence policy, and instant messaging. In this section such a scenario will be illustrated.

In order to incorporate role-based presence handling ($PH_1$) into RoBIM, several enhancements had to be made and liberties taken. However, the core notion of mapping watchers to roles was implemented as $PH_1$ specifies. Similarly, the abstract concept of availability profiles was implemented in a simple yet fitting manner.

### 7.3.1 Watcher-to-role Mappings

A prerequisite for applying presence authorization rules to roles is the ability of managing watchers on a per-role basis. Traditionally IM clients do not show a list of watchers. However, almost all current IM clients enforce a symmetric trust relationship where *if presentity A is a watcher of presentity B then presentity B is a watcher of presentity A*. The RoBIM prototype implements the same strategy but separates watcher management from contact management for clarity.

RFC 4826 provides an extension to the SIP protocol that describes two XML formats for representing resource lists (Roach & Campbell, 2006). Within the RoBIM prototype one of these XML document formats, the

*resource-lists* document format, is used to store watcher-to-role mappings on the presence server.

The XML Configuration Access Protocol (XCAP), defined in RFC 4825, is used to manipulate these watcher-to-role mapping documents. XCAP uses the HTTP protocol to manipulate XML documents in part or full. Figure 7.12 shows an example XCAP message of a watcher-to-role mapping document being updated to the RoBIM SIMPLE server. In order to further understand the usefulness of watcher-to-role mappings and presence rules, as well as availability profiles, an example scenario will now be presented.

### 7.3.2 An Example Scenario

*John is an employee for a paper manufacturing company and communicates regularly with colleagues, friends, and family via Instant Messaging.*

*When John is at work he uses IM to communicate work-related information to and from Colin and Colleen. Furthermore, John occasionally exchanges non-work messages with his wife Fiona and his friend Frank among others. However, every now and then John's friends and family interrupt him while he is busy, causing a break in concentration. John's current resolve is to disconnect his IM client to prevent unwanted interruptions. In doing so, John deprives himself of communicating with co-workers and has to resort to other means of communication.*

*John connects to his IM client at home as well. He perceives IM as a very cost-effective way of communicating with friends and coordinating social events. Many of John's colleagues, including Colin and Colleen, also use IM at home. With the exception of Colleen, whom John considers to be a friend, he does not really want to communicate with his colleagues or share his online presence information with them when he is at home.*

The scenario above epitomizes the current problems with IM and presence-handling which this dissertation has addressed with the proposed $PH_2$ model. Those problems can be listed as John's needs from his IM system:

- To handle different contacts differently. In the above example John wants to treat Colin and Colleen differently; Colleen should have access to John's presence information after work while Colin should not.

- To handle a contact differently depending on the presentities current

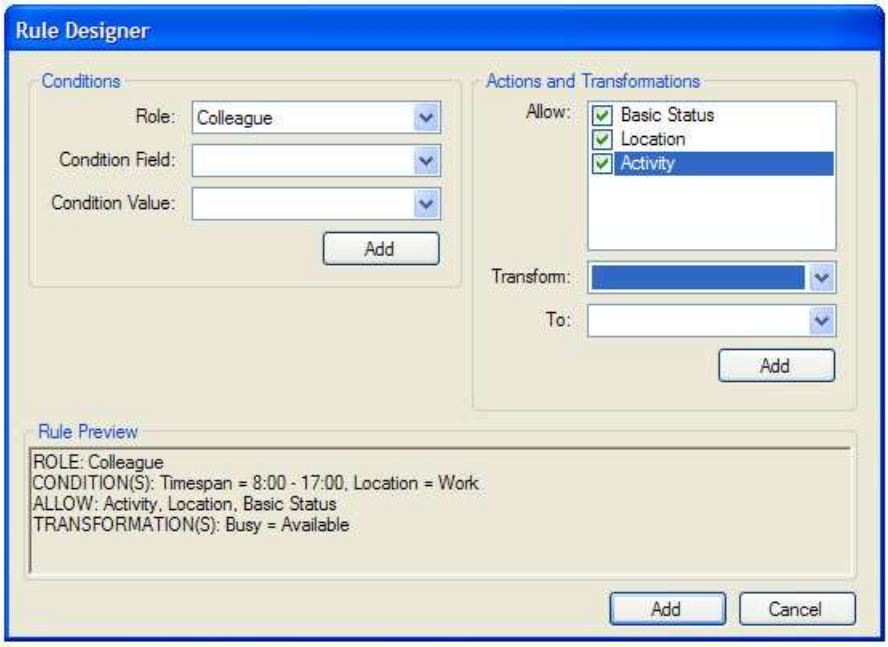


Figure 7.13: "Colleague" presence rule

state. John doesn't want Frank to bother him at work but wants to encourage communication when he is at home.

### 7.3.3 Applying RoBIM Rules to the Example Scenario

In figure 7.13 the Rule Designer is shown with a rule defined for the role of "Colleague". The watcher manager allows John to map Colin and Colleen to the role of "Colleague". According to the rule preview the rule defines extra conditions based on John's current presence information. These conditions are:

- The rule only applies between a time-of-day of 08:00 and 17:00. These hours pre-defined in the RoBIM prototype and should be interpreted as "working hours".
- The rule only applies when John is actually at work (i.e. his presence reflects a "Location" of "Work").

### Combining Rule Conditions

It could be said that John does not work on Saturdays and Sundays. Therefore the first condition should not apply on those two days. This is achieved by adding the second condition which further narrows the applicable conditions. As per the definition of RFC 5025, *all* conditions have to be met for a presence rule to apply.

It should be noted that the first condition is actually an external context variable while the second condition was based on presence information controlled by John. Therefore, John can empower and limit the use of numerous rules by changing a single presence element.

### Modifying Presence Information through Presence Rules

The actual modification of presence information is not currently supported by the presence rule document specification. However, it was included because of its powerful ability to allow a presentity to project a different "Basic Status" to different watchers. This transformation of presence information is part of the server-related processing of availability profiles as defined in the proposed $PH_2$ model.

As mentioned in chapter 2, basic presence states can be confusing to watchers. John might set his presence status to "Busy" because he is working, but that could be misinterpreted by colleagues. Therefore, by allowing John to modify his presence state automatically via presence rules he can more accurately express his availability as "Available" to colleagues while presenting "Busy" to all others. Consequently, each role is provided with a finer grained and more accurate "Basic Status" which describes the presentity's availability and willingness toward that specific role at that specific time.

### Explicitly Allowing Presence Information

In figure 7.13 it can also be seen that the presence rule will *allow Activity, Location, and Basic Status to be included in the presence document presented to all colleagues*. The action of explicitly permitting certain parts of a PIDF document is prescribed in RFC 5025 and implemented as such in the RoBIM prototype.

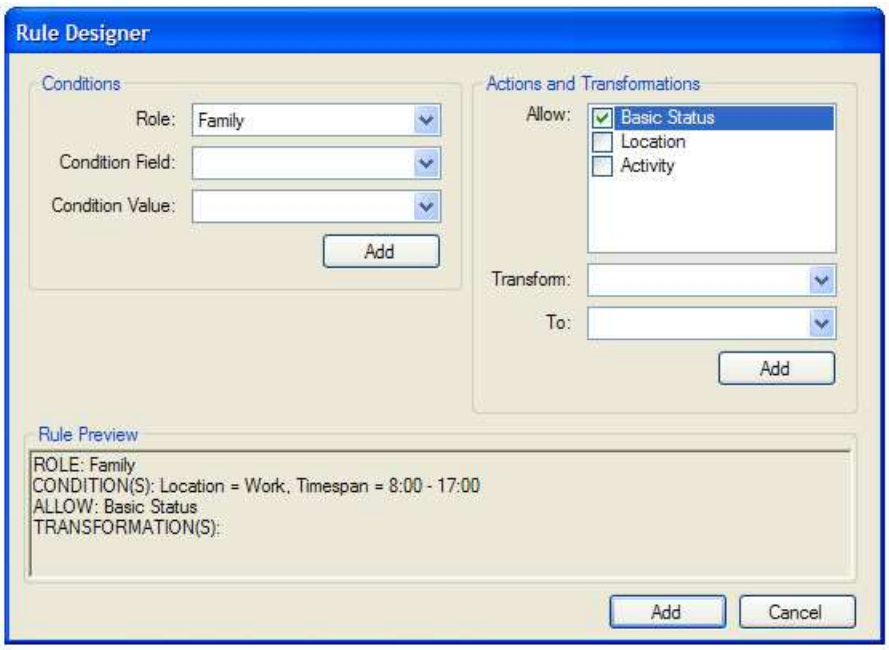


Figure 7.14: "Family" presence rule

As shown in the Allow box within figure 7.13, at the most, only "Basic Status", "Location" and "Activity" can be *allowed* within a PIDF document as implemented in the RoBIM prototype currently. Within our example scenario, John does not want his wife, Fiona, to know all the particulars of his current presence information.

Fiona fits in with the "Family" role and a presence rule was designed in figure 7.14 with her in mind. It shows that only the "Basic Status" will be included with the PIDF document during John's working hours. The conditions are identical to the ones defined in the presence rule that applies to the role of "Colleague". However, the rules may overlap and interfere with each other. Therefore, all applicable roles have to be combined to a single *presence filter* through which the full PIDF document must pass.

### Combining Presence Rules

In figure 7.15 and 7.16 the RoBIM Rule Designer is used to define two separate rules that apply to the role of "Friend". In the story of John it was said that Frank, as well as other friends, should be prevented from interrupting John when he is busy working. However, it was mentioned that there are times during the day that John welcomes IM communication with friends

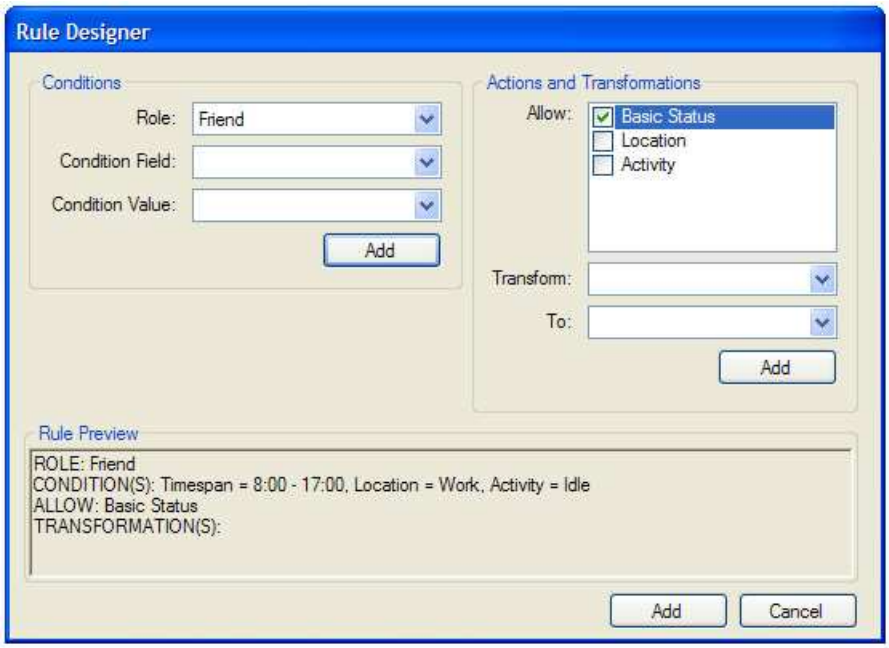


Figure 7.15: "Friend" presence rule applicable at work

and family. To achieve that, the presence rule shown in figure 7.15 specifies that friends receive John's "Basic State" presence information when he is at work but not working (i.e. he is idle).

John also wants to encourage IM communication with friends while he is at home. Therefore, another rule is depicted in figure 7.16. Its conditions specify the rule only applies when John is using IM from home. In such context, a "Friend" should be provided with full presence by providing all three the *allow* options.

Although both rules apply to the the role of "Friend" they can never overlap. This is because John cannot have a "Location" of *home* and *work* at the same time. The two rules can be said to be non-clashing. However, by slightly modifying the rule shown in figure 7.15 slightly a situation can occur where two rules need to be combined.

If John did not define that the first "Friend" rule applied to John being at work (i.e. "Location" set to *work*) a situation can exist where John's context and presence combine in such a way that both rules are put into affect.

Traditional access control such as subscribed to in RFC 5025 will award the highest rights or least restrictive access to the resource (Rosenberg, 2007). Therefore, if "Friend" rule number one only *allows* "Basic State" information

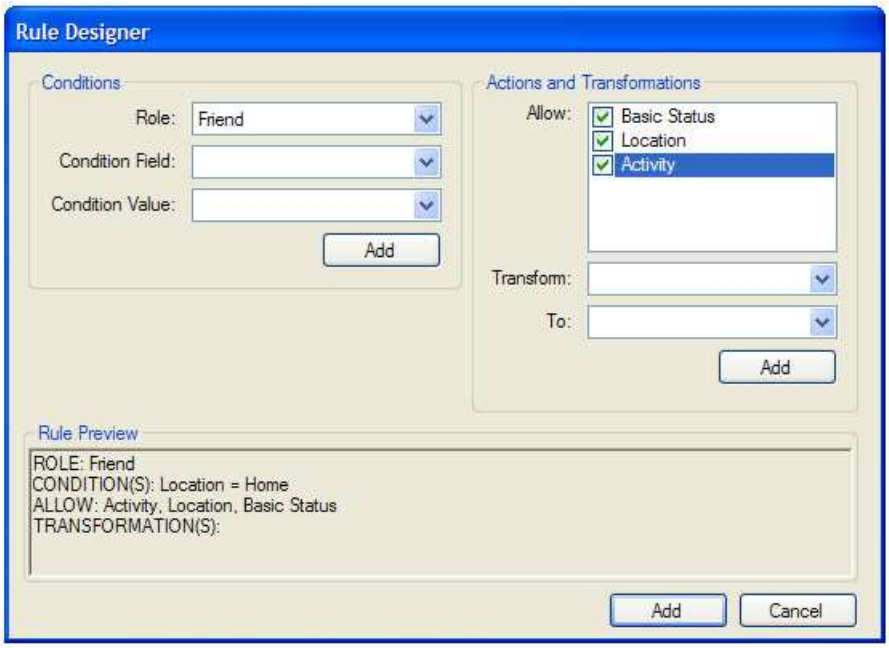


Figure 7.16: "Friend" presence rule applicable at home

and "Friend" rule number two *allows* all three options then all three options will be provided to the watcher. The transformation of presence information will require a special conflict resolution.

### The Special Case of Resolving Presence Transformation Collisions

The problem with transforming special data is that it is not possible to simply grant or deny access to it. It needs to be compared in some way. For example, if both "Friend" rules were to define different "Basic State" transformations, one changes "Available" to "Busy" and one changes "Available" to "Do Not Disturb". Only one can be allowed, but which one?

The simplest way to compare and choose which transformation to use is to associate a priority or preference in numeric value. Resolution of combining rules by making use of numeric values is the recommended approach in RFC

Table 7.2: "Basic State" values and corresponding priority.

| Basic State | Priority (%) |
|---|---|
| Available | 90 |
| Busy | 80 |
| Away | 70 |
| Do Not Disturb | 60 |

4745, the common specification for expressing privacy. By associating a preference value to each "Basic State" option a simple arithmetic comparison of which is greater can be used to choose among transformations. In table 7.2 the relative preference values for "Basic State" transformation is given. One way to identify conflicts in rules is to use some form of autonomous rule engine that continually monitors the rule-making process. The implementation of availability profiles as defined in model $PH_2$, is considered next.

## 7.4 Availability Profiles

The RoBIM prototype implements availability profiles in a very simple manner. The definition of $PH_2$ states that there are two sides to availability profiles: (a) the transformation of availability-related presence attribute values and (b) the handling of incoming messages based on the awarded availability within which the message was sent. Now (a) has already been shown in the presence rule manager, where every presence state may be transformed into a different one. However, the true strength of providing different availability to different roles (and to watchers via implication), is illustrated in (b).

The RoBIM prototype implements a single availability profile for handling incoming messages and it can be described as follows: The identity of the watcher, the roles associated with the watcher, the presentity's current "basic state", and the current presence policy is used to establish the "basic state" that was presented to the watcher. If the "basic state" is equal to "Busy", the message is not shown saliently, but only displayed in the task bar.

With regards to the aforegoing example in section 7.3.2, John would have benefitted from the simple availability profile implementation in RoBIM. More specifically, the IM client would have helped John demonstrate lowered availability to his friends while at work, simply by not grabbing John's attention when his friends' messages are inbound. John may still see that there is a new message but the demand on his attention is significantly reduced.

## 7.5 Discussion

Throughout this dissertation in general, and this section in particular, the strong need for handling watchers more effectively was shown. One of the core needs lacking in current IM systems and standards is the different handling of different watchers. $PH_2$ brings this improved handling to a model for presence in IM. It was said in chapter 2 that basic presence states are often misinterpreted or not considered when initiating conversation. Consequently, unwanted communication follows which could include interruption and lowered productivity.

This dissertation has taken an extra step beyond the need for handling watchers differently by applying a role-based presence handling model. The reason being that to manage each watcher individually is very impractical and not sustainable. Even more so if the rich set of presence attributes, specified by the IETF, is taken into account. However, role-based concepts provide a significant reduction in administration with regards to handling watchers differently.

The significant increase in various information fields carried in current presence documents like the PIDF and RPID specifications adds to the need for managing watchers differently. For example, certain contact numbers can be considered private information and should be removed from presence documents provided to untrusted watchers. Therefore, the role-based presence model was integrated with presence policy to provide fine-grained control over *who gets what information*.

Lastly, the ability to transform "basic state" information is a powerful way for presentities to reduce interruptions via watchers. For instance, an IM client may combine the priority of the presented "basic state" with the salience of an incoming message. If a watcher was presented with a "basic state" of *Available* the incoming message should be very salient, while a presented *Busy* state should be inconspicuously received.

This prototype has successfully demonstrated the per-role presence handling of watchers as well as the utility of availability profiles. Additionally, RoBIM is developed according to the IETF's SIMPLE protocol which provides an extra dimension of feasibility. With the proposed model ($PH_2$) defined, formalized, and demonstrated, this dissertation can conclude in the next chapter.

# Chapter 8

# Conclusion

In this dissertation several issues concerning the handling of presence information in IM were investigated. As a result of a thorough literature review and iterative prototype development, an enhanced model for handling presence in IM was proposed in chapter 6.

Although presence is not limited to IM only, IM was chosen to investigate presence because of its history of implementing presence information. Predictions have it that more people will be using IM overall with each person having more contacts. Consequently, concerns over privacy, presence information handling, and productivity will only increase. All three of these concerns have been addressed in this dissertation.

The following section will summarize how this dissertation addressed the problem statement in terms of achieving the objectives defined in chapter 1.

## 8.1 Revisiting the Objectives

The primary objective of this dissertation was to develop an enhanced model for presence handling in IM. A complete chapter, chapter 6, was dedicated to present the model. Each stage in the model addresses a specific concern. In order to evaluate this model the three secondary objectives defined in chapter 1 must be considered. To recap, these objectives were to:

- allow fine-grained control over what presence information is given to which watcher and when;
- manage presence information in an efficient and scalable manner; and

- provide watchers with individualized presence awareness and allow the presentity to demonstrate awareness accordingly.

Each of these secondary objectives will now be discussed in turn.

### Allow fine-grained control over what presence information is given to which watcher and when

This objective was addressed by the first stage of the proposed model ($PH_0$). $PH_0$ is a simplification of the IETF presence handling model which allows presence to be filtered on a per-watcher basis.

Per-watcher filtering allows the presentity fine-grained control over presence information distribution. Each presence authorization rule applies to a single watcher. The result is that every watcher receives an individualized set of presence information.

In the development of the prototype, the presence authorization rules specification was extended to combine conditions such as time-of-day and location. Thus finer-grained control is provided in the sharing of presence information.

### Manage presence information in an efficient and scalable manner

The increasing popularity of IM will undoubtedly lead to an increase in the number of contacts that an IM user will have to manage. This is especially true in the workplace where one's contact list may include numerous colleagues.

It was stated in chapter 2 that current implementations of IM implement only a very basic set of presence information. Thus the implementation of $PH_0$ will become more difficult if implemented in such a system. However, the IETF specifies a rich set of presence information in its PIDF and RPID data formats as discussed in chapter 4. It is foreseen that if such a rich set of presence information has to be individualized for each watcher, the management thereof will not be feasible.

In order to discover an enhanced means of administrating presence handling, chapter 5 investigated access control models. Chapter 5 concluded

that RBAC provides a very simple means of administrating access control by applying the concept of roles. Hence, RBAC was applied to stage two of the proposed model, $PH_1$, to create a more scalable per-role presence handling. The advantage of handling watchers on a per-role basis is that watchers can be grouped into roles and the presentity only has to filter presence for each role.

### Provide watchers with individualized presence awareness and allow the presentity to demonstrate awareness accordingly

The third stage of the proposed model, $PH_2$, defines availability profiles which can be used to transform availability-related presence information for a specific watcher. In chapter 1 it was said that situations could easily exist where a presentity may want to provide a different presence state to different watchers. In the prototype the transformation of availability-related presence information was implemented by enhancing the presence authorization rules specification.

Availability profiles, as defined by the proposed model, can also be used by the presentity to change the way incoming messages are handled. The prototype implemented that part of availability profiles by showing an incoming message saliently if high availability of a presentity was conveyed and showing a message inconspicuously if low availability was conveyed.

It can therefore be concluded that all of the objectives were met. However, the question can be asked whether this was achieved in a scientifically sound manner.

## 8.2 Conformance to Design Science

In chapter 1 it was stated that this research would conform to the design science research paradigm. Hevner et al. (2004) state that the purpose of design science is to uniquely address an important problem with innovation or to improve upon the solution of a previously solved problem.

The current handling of presence information in IM was shown to be a

significant and relevant problem. Consequently, a solution was developed in the form of an enhanced presence handling model for IM. In this section the seven guidelines of design science will be revisited to confirm that this research did in fact conform to sound design science.

## Design as an Artifact

The principal artifact of this research is a model that enhances presence handling in IM. In particular the model provides role-based presence handling in order to simplify administration of presence information. The model also defines availability profiles which provide the ability to individualize presence information according to the role of the recipient of that information. Availability profiles also provide a means for a presentity to handle incoming messages according to the conveyed availability. The definition of the artifact is contained in chapter 6.

## Problem Relevance

The inability to individualize presence information is one of the problems identified in chapter 1. This problem was emphasized in chapter 3 in stating that the quality of communication is directly affected by how well the medium conveys the intentions of the user. Although the current IETF presence processing model provides a means to filter presence information, there is, to the knowledge of the author, no current presence solution that transforms presence information based on the recipient of that information.

The recent IETF specifications relating to presence was presented in chapter 4 and it was revealed that a greater need for control over the provision of such sensitive information is needed. If the current presence processing model of the IETF is widely adopted, it will not scale well with regards to the administration of presence authorization rules.

The increasing popularity of IM in the workplace also contributes to the problem relevance. Within the workplace environment, it was stated in chapter 2 that IM can lead to lowered productivity despite its usefulness. However, the use of presence information reduces interruption, which can increase productivity.

Other studies within the research group have also highlighted problems

with implementing presence rules (Rutherford, 2008; Ophoff & Botha, 2007) which provides further evidence of problem relevance.

### Design Evaluation

Although no empirical methods were used to evaluate the model, other techniques were employed to show its appropriateness. Firstly, the implementation of the model in the prototype served as proof-of-concept. Secondly, scenario-based reasoning in chapter 7 illustrates how the model solves the identified problems.

### Research Contributions

Good design science requires a clear contribution in the domain of discourse. This research concerns the domains of presence and presence management in IM. The research contribution is the primary design artifact itself; in this case, a model that enhances how presence is handled in IM. This model also contributes to the research conducted by Rutherford (2008) regarding presence in multi-channel communication. Additionally, the prototype that implements the model provides an excellent platform for future related studies.

### Research Rigor

In achieving problem understanding, the author has adhered to sound qualitative research techniques. The research was firmly grounded in theoretical background by conducting a thorough literature review. As part of the theoretical background, current IM and presence standards have been investigated. Furthermore, the prototype has followed an iterative design process in order to provide practical insight throughout the research process. Formal and graphical techniques have been used to reduce ambiguity and enhance clarity.

### Design as a Search Process

This guideline implies that in order to produce an effective artifact, detailed knowledge in the area of research is required. In essence this guideline ex-

cludes ad hoc development. This research was not done in an ad hoc manner. Chapter 2 has provided knowledge on the history and current state of IM as well as how it implements presence. The social psychological roots of IM were considered in chapter 3 to provide knowledge on the effect of presence on people. Furthermore chapter 4 looked at the standards related to the implementation of presence. Access control was considered in chapter 5 to investigate how it may apply to the controlling of presence information. Knowledge about the technical aspects of the work were refined through an iterative approach to prototype development. The aforegoing serves to demonstrate a thorough search for knowledge prior to the formulation of the proposed model presented in chapter 6.

### Communication of Research

Adherence to this guideline has been achieved by submission of this dissertation. The author also collaborated on a more generic presence-handling model which is currently under review at a suitable publisher (Botha, Ophoff, Rutherford, & Victor, 2008).

The aforegoing discussion confirms that sound design science methodology was followed. As this research built on the work of others, it can serve as base for further work.

## 8.3 Future Work

The prototype was developed to be extensible and could prove very useful for future related research. Further studies concerning problems in either IM or presence could benefit from its use. The prototype could help in demonstrating both problems and possible solutions.

The third stage of the proposed model, $PH_2$, introduced the concept of availability profiles. The definition of availability profiles was provided at a very abstract level, further study could refine the concept and thus provide extensions and enhancements to $PH_2$. An investigation of the similarities and differences between availability profiles and RBAC constraints may also provide a further path of research.

The management of presence information is another interesting avenue of presence-related research. The RBAC principles applied in model $PH_1$, and implemented in the prototype, can benefit from an empirical study in order to provide practical proof of its feasibility. However, there may be a challenge in convincing users of the need for role-based presence filtering since most current implementations do not provide a rich set of presence worth controlling in a fine-grained manner.

Despite the benefits of role-based administration of presence, the cost-of-effort to manage watchers remains high. Especially considering that many IM users neglect to update their presence states according to their context. Therefore automated learning techniques, such as artificial neural networks, could be applied to aid the presentity in managing presence rules. Once again the proposed model could be extended to include such techniques with the prototype providing a useful test-bed. Furthermore, as the ability of computers to "sense availability" increases, even more research questions about presence policies could be pursued. In such a research study, the prototype may prove to be very useful.

Multi-channel communication is becoming progressively more important as Internet-enabled communication becomes more pervasive and cost-effective. One of the useful features of the prototype is its standards-based implementation. Since it is based on the SIMPLE protocol which extends the SIP protocol, extending the prototype to employ other communication mediums will not be an extensive process. SIP already supports multiple communication channels. Combined with the object-oriented design, the prototype can easily be extended to provide support for research regarding issues of presence handling in multi-channel communication.

Lastly, from an IM perspective, the prototype could be used to investigate other non-presence related concepts such as security and user interface design.

## 8.4 Final Word

It is the conclusion of the author that this dissertation has provided a useful and much needed new presence handling model to the domain of IM. In doing so, major current limitations of presence use in IM have been addressed and solved.

# Appendix A

# Accompanying Material

The following additional material is supplied on a CD attached to the back cover of this dissertation.

- Source code for the RoBIM prototype, discussed in chapter 7, is available on the CD in a folder named "Source Code".
- Compiled code is provided in the "Executables" folder.
- The folder named "Instructions" provides a short video tutorial which briefly shows some of the core features of RoBIM in action.
- Technical details concerning the prototype are provided in a folder with the corresponding name.

# References

Alvestrand, H. (2002, November). *Instant messaging and presence on the internet* (Member Briefing No. 9). Internet Society.

Argyle, M., & Dean, J. (1965). Eye-contact, Distance and Affiliation. *Sociometry*, *28*(9), 289 - 304.

Avrahami, D., & Hudson, S. E. (2004). QnA: Augmenting an Instant Messaging Client to Balance User Responsiveness and Performance. In *CSCW '04: Proceedings of the 2004 20th Anniversary Conference on Computer Supported Cooperative Work* (Vol. 6, pp. 21–30). New York, NY, USA: ACM Press.

Avrahami, D., & Hudson, S. E. (2006). Responsiveness in Instant Messaging: Predictive Models Supporting Inter-personal Communication. In *CHI '06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 731–740). New York, NY, USA: ACM.

Begole, J. B., Matsakis, N. E., & Tang, J. C. (2004). Lilsys: Sensing Unavailability. In *CSCW '04: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (pp. 511–514). New York, NY, USA: ACM.

Begole, J. B., Tang, J. C., Smith, R. B., & Yankelovich, N. (2002). Work Rhythms: Analyzing Visualizations of Awareness Histories of Distributed Groups. In *CSCW '02: Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work* (pp. 334–343). New York, NY, USA: ACM Press.

Bertino, E., Ferrari, E., Buccafurri, F., & Rullo, P. (1999). A Logical Framework for Reasoning on Data Access Control Policies. In *CSFW '99: Proceedings of the 12th IEEE workshop on Computer Security Foundations* (p. 175). Washington, DC, USA: IEEE Computer Society.

Biocca, F. (1997). The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments. *Journal of Computer-Mediated Communication*, *3*(2).

Biocca, F., Burgoon, J., Harms, C., & Stoner, M. (2001). Criteria and Scope Conditions for a Theory and Measure of Social Presence. *Presence 2001*.

Biocca, F., Kim, T., & Levy, M. (1995). The Vision of Virtual Reality.

*Communication in the age of virtual reality*, 3 - 14.

Biocca, F., & Nowak, K. (2001). Plugging your Body into the Telecommunication System: Mediated Embodiment, Media Interfaces, and Social Virtual Environments. *C.Lin & D. Atkin (Eds.), Communication Technology and Society*, 407–447.

Botha, R., Ophoff, J., Rutherford, A., & Victor, R. (2008). *A role-based access control model for presence information.*

Burgoon, J., Buller, D. B., Hale, J. L., & deTurck, M. (1984). Relational Messages Associated with Nonverbal Behaviors. *Human Communication Research*, *10*(3), 351 - 378.

Burgoon, J., & Hale, J. (1987). Validation and Measurement of the Fundamental Themes of Relational Communication. *Communication Monographs*, *54*, 19 - 41.

Cameron, A. F., & Webster, J. (2005). Unintended Consequences of Emerging Communication Technologies: Instant Messaging in the Workplace. *Computers in Human Behavior*, *21*, 85-103.

Campbell, B., Rosenberg, J., Schulzrinne, H., Huitema, C., & Gurle, D. (2002, December). *Session Initiation Protocol (SIP) Extension for Instant Messaging* (Tech. Rep. No. 3428). Internet Engineering Task Force. RFC 3428 (Proposed Standard).

Connell, J. B., Mendelsohn, G. A., Robins, R. W., & Canny, J. (2001). Effects of Communication Medium on Interpersonal Perceptions. In *GROUP '01: Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work* (pp. 117–124). New York, NY, USA: ACM.

Crampton, J. (2004). Applying Hierarchical and Role-based Access Control to XML Documents. In *SWS '04: Proceedings of the 2004 Workshop on Secure Web Service* (pp. 37–46). New York, NY, USA: ACM.

Czerwinski, M., Cutrell, E., & Horvitz, E. (2000). Instant Messaging and Interruption: Influence of Task Type on Performance. In *Proceedings of OZCHI 2000, Sydney, Australia, Dec. 4-8.*

Czitrom, D. J. (1982). Media and the American Mind.

Daft, R., & Lengel, R. (1986). Organizational Information Requirements, Media Richness, and Structural Design. *Management Science*, *32*, 554-571.

Damasio, A. (1994). *Descartes' Error : Emotion, Reason, and the Human Brain.* Penguin (Non-Classics).

Day, M., Rosenberg, J., & Sugano, H. (2000, Feb). *A Model for Presence and Instant Messaging* (RFC No. 2778). Internet Engineering Task Force.

Derks, D., Bos, A. E. R., & Grumbkow, J. von. (2008). Emoticons in Computer-Mediated Communication: Social Motives and Social Context. *CyberPsychology & Behavior*, *11*(1), 99 - 101.

Dersingh, A., Liscano, R., & Jost, A. (2005). Managing Access Control for Presence-Based Services. In *CNSR '05: Proceedings of the 3rd Annual Communication Networks and Services Research Conference* (pp. 105–111). Washington, DC, USA: IEEE Computer Society.

Dourish, P., & Bly, S. (1992). Portholes: Supporting Awareness in a Distributed Work Group. In *CHI '92: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 541–547). New York, NY, USA: ACM.

Ferraiolo, D. F., Cugini, J., & Kuhn, D. R. (1995). Role Based Access Control: Features and Motivations. In *Annual Computer Security Applications Conference.* IEEE Computer Society Press.

Ferraiolo, D. F., & Kuhn, D. R. (1992). Role Based Access Control. In *15th National Computer Security Conference* (pp. 554–563).

Ferraiolo, D. F., Sandhu, R. S., Gavrila, S., Kuhn, D. R., & Chandramouli, R. (2001). Proposed NIST Standard for Role-Based Access Control. *ACM Trans. Inf. Syst. Secur.*, *4*(3), 224–274.

Fischer, C. S. (1988). "Touch Someone": The Telephone Industry Discover Sociablity. *Technology and Culture*, *29*, 32–47.

Garramone, G. M., Harris, A., & Anderson, R. (1986). Uses of Political Computer Bulletin Boards. *Journal of Broadcasting and Electronic Media*, *30*(3), 325 - 339.

Goffman, E. (1959). *The Presentation of Self in Everyday Life.* Garden City, NY: Anchor.

Greef, P. de, & IJsselsteijn, W. (2000). Social Presence in the Photoshare Tele-application. *Presence 2000.*

Greene, D., & O'Mahony, D. (2004). Instant Messaging & Presence Management in Mobile Ad-Hoc Networks. In *PerCom Workshops* (p. 55-59).

Grinter, R., & Palen, L. (2002). Instant Messaging in Teen Life. In *Pro-*

*ceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'02)* (p. 21 - 30). New Orleans, Louisiana, USA: USA: ACM Press.

Gunawardena, C. N. (1991). Collaborative Learning and Group Dynamics in Computer-mediated Communication Networks. *Research Monograph No. 9 of the Second American Symposium on Research in Distance Education*, 14 - 24.

Gunawardena, C. N. (1995). Social Presence Theory and Implications for Interaction and Collaborative Learning in Computer Conferences. *International Journal of Educational Telecommunications*, *1*(2/3), 147 - 166.

Gunawardena, C. N., & Zittle, F. (1997). Social Presence as a Predictor of Satisfaction within a Computer-mediated Conferencing Environment. *American-Journal-of-Distance-Education*, *11*(3), 8 - 26.

Hancock, J. T., Landrigan, C., & Silver, C. (2007). Expressing Emotion in Text-based Communication. In *CHI '07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 929–932). New York, NY, USA: ACM.

Handel, M., & Herbsleb, J. D. (2002). What is Chat Doing in the Workplace? In *CSCW '02: Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work* (pp. 1–10). New York, NY, USA: ACM.

Harrison, M. A., Ruzzo, W. L., & Ullman, J. D. (1976). Protection in Operating Systems. *Commun. ACM*, *19*(8), 461–471.

Heeter, C. (1992). Being There: The Subjective Experience of Presence. *Presence: Teleoper. Virtual Environ.*, *1*(2), 262–271.

Hevner, A., March, S., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75-105.

Hildebrand, J. (2003). Nine IM Accounts and Counting. *Queue*, *1*(8), 44–50.

Ho, C., Basdogan, C., Slater, M., Durlach, N., & Srinivasan, M. (1998). An Experiment on the Influence of Haptic Communication on the Sense of Being There. In *In Proceedings of the British Telecom Workshop on Presence in Shared Virtual Environments* (pp. 525–532).

Hung, S.-Y., Huang, A. H., Yen, D. C., & Chang, C.-M. (2007). Comparing the Task Effectiveness of Instant Messaging and Electronic Mail for Geographically Dispersed Teams in Taiwan. In *Computer Standards &*

*Interfaces* (Vol. 29, pp. 626–634). Elsevier.

Hwang, H. S., & Lombard, M. (2006). Understanding Instant Messaging: Gratifications and Social Presence. In *Presence 2006: The 8th annual international workshop on presence.* Internet Society for Presence Research.

Isaacs, E., Walendowski, A., & Ranganthan, D. (2002). Hubbub: A Sound-enhanced Mobile Instant Messenger that Supports Awareness and Opportunistic Interactions. In *CHI '02: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 179–186). New York, NY, USA: ACM.

Isaacs, E., Walendowski, A., Whittaker, S., Schiano, D. J., & Kamm, C. (2002). *The Character, Functions, and Styles of Instant Messaging in the Workplace.* New York, NY, USA: ACM.

Kato, Y., Kato, S., & Akahori, K. (2007). Effects of Emotional Cues Transmitted in E-mail Communication on the Emotions Experienced by Senders and Receivers. *Comput. Hum. Behav.*, *23*(4), 1894–1905.

Kontzer, T. (2002). AOL quietly wins IM patent. *InformationWeek*.

Licari, J. (2005, Jun). Best Practices for Instant Messaging in Business. *Netork Security*, 4 - 7.

Lombard, M., & Ditton, T. (1997). At the Heart of it All: The Concept of Presence. *Journal of Computer-Mediated Communication*, *3*(2).

McClea, M., Yen, D., & Huang, A. (2004). An Analytical Study Towards the Development of a Standardized IM Application. *Computer Standards & Interfaces*, *26*, 343 - 355.

McCroskey, J., Richmond, V., & Daly, J. (1975). The Development of a Measure of Perceived Homophily in International Communication. *Human Communication Research*, *1*(4), 323 - 332.

Mcluhan, M. (1964). *Understanding Media, The Extensions of Man.* MIT Press.

Mills, E. (2005, Oct). *Yahoo, Microsoft Join IM Hands.*

MXIT Reaches the Three Million Mark. (2005). *ICT World*.

Nardi, B. A., Whittaker, S., & Bradner, E. (2000). Interaction and Outeraction: Instant Messaging in Action. In *CSCW '00: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (pp. 79–88). New York, NY, USA: ACM.

Newberry, B. (2001). Raising Student Social Presence in Online Classes. In W. A. Lawrence-Fowler & J. Hasebrook (Eds.), *Webnet* (p. 905-910). AACE.

Ni, Q., Trombetta, A., Bertino, E., & Lobo, J. (2007). Privacy-aware Role Based Access Control. In *SACMAT '07: Proceedings of the 12th ACM Symposium on Access Control Models and Technologies* (pp. 41–50). New York, NY, USA: ACM.

Niemi, A. (2004, Oct). *Session Initiation Protocol (SIP) Extension for Event State Publication* (RFC No. 3903). Internet Engineering Task Force.

Nowak, K. (2000). *The Influence of Anthropomorphism on Mental Models of Agents and Avatars in Social Virtual Environments.* Michigan State University E Lansing.

O'Kane, P., & Hargie, O. (2007). Intentional and Unintentional Consequences of Substituting Face-to-face Interaction with E-mail: An Employee-based Perspective. *Interacting with Computers*, *19*(1), 20–31.

Ophoff, J., & Botha, R. (2007). Mobile Communications: User Perception and Practice. *South African Computer Journal*(40), 63–73.

Osborn, S., Sandhu, R. S., & Munawer, Q. (2000). Configuring Role-based Access Control to Enforce Mandatory and Discretionary Access Control Policies. *ACM Trans. Inf. Syst. Secur.*, *3*(2), 85–106.

Palen, L., & Dourish, P. (2003). Unpacking "Privacy" for a Networked World. In *CHI '03: Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 129–136). New York, NY, USA: ACM.

Palmer, M. (1995). Interpersonal Communication in Virtual Reality: Mediating Interpersonal Relationships. *F. L. Biocca, M. (Ed.), Communication in the age of virtual reality*, 277 - 302.

Perse, E. I., Burton, P., Kovner, E., Lears, M. E., & Sen, R. J. (1992). Predicting Computer-mediated Communication in a College Class. *Communication Research Reports*, *9*(2), 161 170.

Peterson, J. (2004a, Aug). *Common Profile for Instant Messaging (CPIM)* (RFC No. 3860). Internet Engineering Task Force.

Peterson, J. (2004b, Aug). *Common Profile for Presence (CPP)* (RFC No. 3859). Internet Engineering Task Force.

Rettie, R. (2003). Connectedness, Awareness and Social Presence. In *6th*

*International Presence Workshop.* Aalborg.

Rheingold, H. (1993). The Virtual Community.

Rice, R. (1993). Media Appropriateness; Using Social Presence Theory to Compare Traditional and New Organizational Media. *Human Communication Research, 19*(4).

Roach, A. (2002, Jun). *Session Initiation Protocol (SIP)-Specific Event Notification* (RFC No. 3265). Internet Engineering Task Force.

Roach, A., & Campbell, B. (2006, Aug). *A Session Initiation Protocol (SIP) Event Notification Extension for Resource Lists* (RFC No. 4662). Internet Engineering Task Force.

Rosenberg, J. (2004a, Aug). *A Presence Event Package for the Session Initiation Protocol (SIP)* (RFC No. 3856). Internet Engineering Task Force.

Rosenberg, J. (2004b, Aug). *A Watcher Information Event Template-Package for the Session Initiation Protocol (SIP)* (RFC No. 3857). Internet Engineering Task Force.

Rosenberg, J. (2005, Feb). *A Processing Model for Presence* (Internet Draft). Internet Engineering Task Force.

Rosenberg, J. (2007, Dec). *Presence Authorization Rules* (RFC No. 5025). Internet Engineering Task Force.

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., et al. (2002, June). *SIP: Session Initiation Protocol* (Tech. Rep. No. 3261). Internet Engineering Task Force. RFC 3261 (Standards Track). (Obsoletes RFC 2543)

Rutherford, A. (2008). *Automating the Management of Presence Policies.* School of ICT, NMMU. (Doctoral Research Proposal)

Salin, P. (2004, September). *Mobile Instant Messaging Systems - A Comparitive Study and Implementation.* Helsinki, Finland. (Unpublished Masters Thesis)

Sánchez, J. A., Hernández, N. P., Penagos, J. C., & Ostróvskaya, Y. (2006). Conveying Mood and Emotion in Instant Messaging by Using a Two-Dimensional Model for Affective States. In *IHC '06: Proceedings of VII Brazilian Symposium on Human Factors in Computing Systems* (pp. 66–72). New York, NY, USA: ACM.

Sandhu, R. S. (1993, Nov). Lattice-based Access Control Models. *Computer*,

*26*(11), 9-19.

Sandhu, R. S., Coyne, E. J., Feinstein, H. L., & Youman, C. E. (1996). Role-based Access Control Models. *Computer*, *29*(2), 38–47.

Savicki, V., & Kelley, M. (2000). Computer Mediated Communication: Gender and Group Composition. *Cyberpsychology and behavior*, *3*(5), 817 - 826.

Schulzrinne, H., Gurbani, V., Kyzivat, P., & Rosenberg, J. (2006, Jul). *RPID: Rich Presence Extensions to the Presence Information Data Format (PIDF)* (RFC No. 4480). Internet Engineering Task Force.

Schulzrinne, H., Tschofenig, H., Morris, J., Cuellar, J., Polk, J., & Rosenberg, J. (2007, Feb). *Common Policy: A Document Format for Expressing Privacy Preferences* (RFC No. 4745). Internet Engineering Task Force.

Segerstad, Y., & Ljungstrand, P. (2002). Instant Messaging with WebWho. *International Journal of Human-Computer Studies*, *56*, 147-171.

Short, J., Williams, E., & Christie, B. (1976). *The Social Psychology of Telecommunications*. New York: Wiley.

Smale, S., & Greenberg, S. (2005). Broadcasting Information via Display Names in Instant Messaging. In *GROUP '05: Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work* (pp. 89–98). New York, NY, USA: ACM.

Smith, D. (2007, Sep). *Battle Under Way Among IM Presence Protocols* (Tech. Rep.). Gartner, Inc.

Smith, D., & Lundy, J. (2007, Apr). *MarketScope for Instant Messaging, 2007* (Tech. Rep. No. G00147732). Gartner, Inc.

Soe, L., & Markus, M. (1993). Technological or Social Utility? Unraveling Explanations of Email, Vmail, and Fax Use. *The Information Society*, *9*, 213-236.

Stone, J., & Merrion, S. (2004). Instant Messaging or Instant Headache? *Queue*, *2*(2), 72–80.

Strom, D. (2006, July). Review: AOL's AIM Pro still in the Amateur League. *Computerworld.com*.

Sugano, H., Fujimoto, S., Klyne, G., Bateman, A., Carr, W., & Peterson, J. (2004, Aug). *Presence Information Data Format (PIDF)* (RFC No. 3863). Internet Engineering Task Force.

Tang, J. C., & Begole, J. B. (2003). Beyond Instant Messaging. *Queue*, *1*(8),

28–37.

Tang, J. C., Yankelovich, N., Begole, J., Kleek, M. V., Li, F., & Bhalodia, J. (2001). ConNexus to Awarenex: Extending Awareness to Mobile Users. In *CHI '01: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 221–228). New York, NY, USA: ACM.

To, P.-L., Liao, C., Chiang, J. C., Shih, M.-L., & Chang, C.-Y. (2008). An Empirical Investigation of the Factors Affecting the Adoption of Instant Messaging in Organizations. *Comput. Stand. Interfaces*, *30*(3), 148–156.

Tran, M. H., Yang, Y., & Raikundalia, G. K. (2004). Consumption of Multiple Concurrent Identities: The Need from the Instant Messaging Virtual Community. In (Vol. 20).

Tran, M. H., Yang, Y., & Raikundalia, G. K. (2005). Supporting Awareness in Instant Messaging: An Empirical Study and Mechanism Design. In *OZCHI '05: Proceedings of the 19th Conference of the Computer-Human Interaction Special Interest Group (CHISIG) of Australia on Computer-Human Interaction* (pp. 1–10). Narrabundah, Australia, Australia: Computer-Human Interaction Special Interest Group (CHISIG) of Australia.

Trevino, L., Lengel, R., & Daft, R. (1987). Media Symbolism, Media Richness, and Media Choice in Organizations: A Symbolic Interactionist Perspective. *Communication Research*, *14*, 553-574.

Tu, C. H. (2000). Critical Examination of Factors Affecting Interaction on CMC. *Journal of Network and Computer Applications*, *23*(1), 39 - 58.

Tu, C. H. (2002a). The Impacts of Text-Based CMC on Online Social Presence. *Journal of Interactive Online Learning*, *1*(2), 1 - 24.

Tu, C. H. (2002b). The Relationship Between Social Presence and Online Privacy. *The Internet and Higher Education*, *5*(4), 293 - 318.

Tu, C. H., & McIsaac, M. S. (2002). An Examination of Social Presence to Increase Interaction in Online Classes. *American Journal of Distance Education*, *16*(3), 131 - 150.

Viégas, F. B., & Donath, J. S. (1999). Chat Circles. In *CHI '99: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 9–16). New York, NY, USA: ACM.

Vos, H. de, Hofte, G. H. ter, & Poot, H. de. (2004). IM [@Work] Adoption of Instant Messaging in a Knowledge Worker Organisation. In *Hicss.*

Walter, J. (1992). Interpersonal Effects in Computer-Mediated Interaction; A Relational Perspective. *Communication Research*, *19*(1), 52–90.

Walther, J., & Burgoon, J. (1992). Relational Communication in Computer-Mediated Interaction. *Human Communication Research*, *19*(1), 50 - 88.

Weiner, M., & Mehrabian, A. (1968). Language Within Language: Immediacy, a Channel in Verbal Communication. *Psychoanalytic Quarterly*, *39*, 489 - 493.

Zhang, X., Li, Y., & Nalla, D. (2005). An Attribute-based Access Matrix Model. In *SAC '05: Proceedings of the 2005 ACM Symposium on Applied Computing* (pp. 359–363). New York, NY, USA: ACM.