A Model for Privacy-Aware Presence Management in Mobile Communications

by

Jacobus A. Ophoff

A Model for Privacy-Aware Presence Management in Mobile Communications

by

Jacobus A. Ophoff

Thesis

submitted in fulfillment of the requirements for the degree

Philosophiae Doctor

in

Information Technology

in the

Faculty of Engineering, the Built Environment and Information Technology

of the

Nelson Mandela Metropolitan University

Promoter: Prof. Reinhardt A. Botha

January 2011

Declaration

I, Jacobus A. Ophoff, hereby declare that:

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

Jacobus A. Ophoff
January 7, 2011

Abstract

As humans we find communicating natural and compelling. Over the centuries we have created many innovations which enable and improve communication between people; during the past decade mobile phone networks have brought about a technological revolution in this area. Never before have people been so connected to one another. Today we have the ability to communicate with almost anyone, anytime, anywhere.

Our increased connectivity and reachability also leads to new issues and challenges that we need to deal with. When we phone someone we expect an instant connection, and when this does not occur it can be frustrating. On the other hand it is equally disruptive to receive a call when one is busy with an important task or in a situation where communication is inappropriate. Social protocol dictates that we try to minimize such situations for the benefit of others nearby and for ourselves. This management of communications is a constant and difficult task. Using presence – which signals a person's availability and willingness to communicate – is a solution to this problem. Such information can benefit communication partners by increasing the likelihood of a successful connection and decreasing disruptions.

This research addresses the problem of staying connected while keeping control over mobile communications. It adopts a design-science research paradigm, with the primary research artifact being a model for privacy-aware presence management in mobile communications. As part of the model development knowledge contributions are made in several ways.

Existing knowledge about the problem area is extended through a quantitative analysis of mobile communications management. This analysis uses a novel survey, collecting useful empirical data for future research. This includes how people currently manage their communications and what features they expect from a potential "call management" system.

The examination and use of presence standards, as a foundation for the model, provides a comparison of the main presence technologies available today. A focus on privacy features identifies several shortcomings in standards which, if addressed, can help to improve and make these standards more complete.

The model stresses the privacy of potentially sensitive presence information. A unique perspective based on social relationship theories is adopted. The use of relationship groups not only makes logical sense but also assists in the management of presence information and extends existing standards.

Finally, the evaluation of the model demonstrates the feasibility of a practical implementation as well the ability to extend the model in next generation mobile networks. Thus the model presents a solid foundation for the development of future services. In these ways the proposed model contributes positively towards balancing efficient mobile communications with the need for privacy-awareness.

Acknowledgements

It is a pleasure to thank those who have made this thesis possible:

- I owe my deepest gratitude to my supervisor, Prof. Reinhardt Botha, whose patience and kindness, as well as his academic experience, have been invaluable to me. Without his encouragement and support this thesis would not have been possible.
- I am grateful to the staff members at Nelson Mandela Metropolitan University for their kind assistance.
- I would like to thank my entire extended family for their love and constant support throughout my studies.
- Last, and most importantly, I would like to thank my wife Madri for her love and patience. She has, and always will be, my joy and pillar of strength.

And also that every man should eat and drink, and enjoy the good of all his labour, it is the gift of God.
ECCLESIASTES 3:13 KJV

Contents

List of Figures

List of Tables

List of Definitions

Part I

Background

Chapter 1

Introduction

The past decades have seen uninterrupted global growth of information and communication technologies (ICTs). Mobile subscriptions have set the highest records. International Telecommunication Union (2009) statistics show that the number of mobile subscribers have grown from about 318 million in 1998 to over 4 billion in 2008. Currently this translates into a penetration rate of 61 per cent.

In comparison the number of fixed telephone lines are decreasing while the number of Internet users are growing at a much smaller rate. Despite differences between regions and between developed and developing economies (International Telecommunication Union, 2009), as well as the issue of multiple SIM card ownership (Sutherland, 2009), these statistics show a clear shift towards mobile telephony as the dominant communications technology.

The pervasive nature of mobile communications is easily understood when looking at its advantages. Amongst others, it provides the ability to coordinate actions irrespective of time and place (Ling, 2004, pp. 57–80). Mobile communications make us more connected and by increasing our accessibility it expands our social network (Palen, 2002). Furthermore it provides a sense of security, which is often valued by elderly users (Ling, 2004, pp. 35–54).

So great has the impact of these technologies been that one could argue that it has significantly changed how our society operates. Unfortunately, the anytime anywhere paradigm of mobile communications can also lead to negative consequences. This is frequently noticed in everyday life.

1.1 Mobile Disruption

Without control mobile communications risk becoming disruptive and disorganizing (Rennecker and Godwin, 2005). These consequences can be far-reaching. Ling (2004) states that many people find mobile phones disturbing, and there are numerous situations where the use of mobile phones are seen as inappropriate.

According to Ling (2004, pp. 125–142) there are three general domains in which mobile communications can cause disruption: public settings with extensive norms governing behaviour (such as restaurants), interpersonal interactions and on an individual, internal level. It is clear that the disruption caused by an incoming communication affects the recipient, people in the immediate vicinity, and also changes the social status and behaviour patterns.

Rennecker and Godwin (2005) show that communication technologies can both facilitate and interrupt work activities. People experience shorter delays in obtaining information needed for a task at the cost of interrupting the work of others. This contradiction stems from our increased availability and lack of sophisticated communications management tools. Another aspect of the problem is that the caller usually has no motivation to delay communication until such a time that it is convenient for the receiver (Dabbish and Kraut, 2004). The problem of disruptions is extremely relevant when we frame it against the current state of information overload. Having to deal with a multitude of facts and tasks as efficiently as possible has meant that our attention has become scarce – put another way we are trading in the economics of attention (Davenport and Beck, 2000). Operating in this environment requires us to manage our attention, and correspondingly our communications, as efficiently as possible if we want to lead a productive life.

The disruptive effect of mobile communications on bystanders is also evident. A prominent researcher in this area, Richard Ling, states that "both qualitative and quantitative data suggest that the mobile telephone is seen as an invasive influence in public spaces" (Ling, 2004, p. 123). In public spaces, such as restaurants, theaters, elevators, trains, social functions, and meetings, using a phone can have various negative effects: feelings of awkwardness and embarrassment, the creation of social partitions, and forced eavesdropping being some examples (Ling, 2004, pp. 123–143).

The act of answering a call results in a transition from the current situation to an interaction with the caller (Light, 2008). Such a sudden shift in attention can make bystanders uncomfortable and causes tension by destabilizing the boundaries between people (Palen and Dourish, 2003). In addition, an unexpected call often causes embarrassment for the receiver when it occurs in a public setting (Feenberg, 2003). It is easy to agree with these effects because of personal experience and word-of-mouth recollections from other people. While it can be argued that the receiver always has the choice of not answering a call, there may still be a resultant mental distraction. For example, the receiver may wonder why someone was calling and whether it was important.

Recently public debate has surrounded the proposed allowance of mobile communications on airplanes (Jones, 2008). The level of noise versus convenience clearly divides opinion. Thus far it has been impossible to reach consensus over this issue. The improper use of mobile communications can also lead to more severe outcomes. Research has shown the dangers of using a mobile phone while driving (Redelmeier and Tibshirani, 1997; Lamble et al., 1999). In such situations mobile phone use decreases our reaction time and diverts our attention from more important matters (Ling, 2004, pp. 49–53). Laws banning mobile phone use in vehicles try to curb this problem, but often prove hard to enforce. The results can be fatal, for ourselves and for other road users.

Thus mobile communications induce complexity through the environments where we use the technology. We are required to use phones responsibly and also consider those around us. This responsibility for adhering to laws and social etiquette falls on the shoulders of every user. In the face of these challenges people often reconsider their perception of acceptable use and adopt a more tolerant attitude (Palen et al., 2000; Love and Perry, 2004). Research suggests that a range of dynamic factors influence our communications: the communications medium, relationship between caller and receiver, status differences, affinity towards a contact, expectations of reciprocity and culture all play a role (Rennecker and Godwin, 2005). Perhaps this explains why an effective solution to this problem is yet to be found.

1.2 Context Information

While the convergence of communication channels with the Internet is delivering richer communication experiences this research focuses specifically on mobile communications in the traditional sense – telephony and text messaging over a mobile network. In this environment the lack of context available between caller and receiver is a cause of many of the previous issues. According to Dourish (2004) context is "a slippery notion" which "slips away when one attempts to define it." He defines context as an emergent property which is dynamic and arises from activity. A simpler definition is provided by Abowd and Mynatt (2000) who propose to base a definition on the "five W's":

1. Who: the identity of a person and other people in their environment.
2. What: the current activity of a person.
3. Where: the current location of a person.
4. When: the influence of time, such as when an activity takes place or the duration of time spent at a location.
5. Why: the reason for a person's actions.

It is not unreasonable to assume that such information can dramatically reduce disruptions and socially awkward situations. Indeed it has been shown that transmitting some context information before starting a session can greatly enhance the productivity of communications (Ljungstrand, 2001). This can be particularly valuable in telecommunications, where it is estimated that only a small percentage of calls end in productive conversations (Rosenberg, 2000).

Unfortunately, mobile communications does not lend itself easily to the incorporation of context. Currently the Caller ID (Petersen, 2002, pp. 161–162) feature is the only context indication and is only available to the receiver. However, it is often unavailable because it can be switched off by the caller.

1.3 Structured Context: Presence

Context can be used to present an accurate and relatively complete picture of a person's situation. But to use such information effectively it needs to be standardized so that its meaning is understood by users. An example of such a standard is presence.

Presence is a concept which has different meanings for different people and is hard to pin down. This research defines presence as an indication of the ability and willingness of a person to communicate (Rosenberg, 2006). While these terms may seem similar there is a distinction between them. Ability provides a technical indication: the user is away from the communications device, inactive, or present to communicate. Willingness refers to the user's attitude or availability of attention: for example, do not disturb me because I am very busy, I am unavailable because we are having a meeting, or I can talk. Only by combining these attributes do we get a true sense of a person's presence for communication.

Presence is a service that is well-known from the instant messaging (IM) world. It forms an integral part of the IM experience by indicating a user's availability for communication. As such it makes communication and collaboration more efficient while reducing disruptions (Vaughan-Nichols, 2003). Due to its success presence has also spread to applications other than IM. Presence can now be found in word processors, spreadsheets, social networks and enterprize communication applications. According to Gartner research, presence is a proven and rapidly evolving technology which will enable novel business opportunities across industries, resulting in major shifts in industry dynamics (Johnson et al., 2007). Applications such as unified communications and collaboration, presence-based contact routing and enterprize IM are just some of the potential possibilities enabled by presence.

However, apart from all the obvious benefits presence information can deliver to enrich communications, several concerns still remain. For a presence service to succeed user discipline is necessary. Users need to update their presence state on a regular basis to make it useful to other parties. Automated systems can assist in this process, but in doing so often reduce the quality of the information (e.g. changing the presence state after a period of inactivity).

A more serious concern is a breach of user privacy through the misuse of presence information. User (and communications) monitoring is becoming an important social issue and presence information can also be abused in this respect. People are more likely to have a resistance to presence if this information will be used for accountability purposes (Becker, 2004). Once presence information has been made available to a user, protection against monitoring becomes problematic and additional steps are necessary to prevent this. Most users would prefer to keep their current status hidden unless specifically required for communications decisions. As the scope of presence information grows, the privacy requirements also become more important (Rosenberg, 2000).

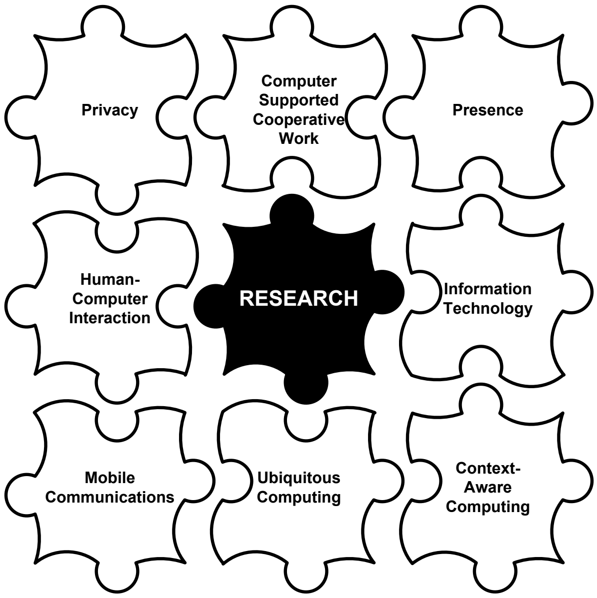

Figure 1.1: Domain of Discourse

Presence provides information regarding the context of a user thus enabling people to make smarter communication choices. While IM has been a driver for presence (Poikselkä et al., 2004), it is predicted that presence will be a universal feature in future communication networks such as the IP Multimedia Subsystem (Camarillo and García-Martín, 2004). Until then this research examines the use of presence in mobile communications as a means to include context in the interactions between caller and receiver.

1.4 Domain of Discourse

This research incorporates aspects from various fields of study. Most of the fields are closely related, while in certain fields a component within the research area is relevant to this study. A visual description of the domain of discourse is given in Figure 1.1.

The research is rooted in the mobile communications domain. This domain is extensive, encompassing topics such as network architectures, protocol design, and end-user services. The research concentrates on communication-related topics such as call setup and data services. While research on improving the efficiency of communications has been conducted in many areas, for instance email and video-conferencing, the mobile domain presents unique challenges; standard network practices need to be followed and limitations need to be taken into consideration. In addition the research outputs should be adaptable to next generation mobile networks, such as the IP Multimedia Subsystem. In this research the term 'mobile communication' refers specifically to mobile telephony, which excludes other communication channels such as Voice over IP (VoIP) and IM. This focus is due to the unique problems faced by users and the lack of context information in the communication process. In addition, this is still the predominant way of communicating while mobile.

The goal of improving the efficiency of communications is closely linked with human-computer interaction (HCI) research. HCI is a well-established field with numerous subdomains. In general, HCI "is the study and the practice of usability" (Carroll, 2002, p. xxvii). In this sense the research addresses the interaction between caller and receiver, making the experience more efficient and pleasant for each user. This improved interaction can be achieved by extending the current capabilities of the phone.

The research also covers issues of coordination which are similar to research within the domain of computer supported cooperative work (CSCW). CSCW studies issues surrounding the collaboration of groups and coordination of activities using computer systems (Carstensen and Schmidt, 1999). Improving the coordination of phone calls, between a caller and receiver who can be considered as mobile collaborators, is a fundamental aim of this research. Such coordination is supported by sharing context information between individuals, similar to the awareness dimension within CSCW.

Considering that modern mobile phones are in essence miniature, portable computers which can be used anytime and anywhere, creates a strong association with the field of ubiquitous computing (ubicomp). Ubicomp, originally coined by Weiser, is "a vision of people and environments augmented with computational resources that provide information and services when and where desired" (Abowd and Mynatt, 2000). Indeed mobile communications may be one of the best examples of a ubicomp technology. The research aims to assist the user of mobile communications by augmenting the existing capabilities of the device. Using context information, obtained from sensors and personal information managers, it is possible to better predict communications availability.

The automated use of context information within the research also draws knowledge from the context-aware computing field within ubicomp. Context-aware computing studies systems that are aware of their context and adapt to it; such systems can take action automatically without involving the user unnecessarily (Loke, 2007, pp. 7–8). As a context-aware system the research will enhance mobile communications by sensing the user's context and using such information as part of an intelligent communications manager. Calling parties can be informed of unavailability and action can be taken to divert unwanted calls without user involvement.

Sensing and representing context is not a simple process. Within the field of information technology the research makes extensive use of presence to accomplish this. Presence allows a standardized presentation of context with well-defined protocols, components and interactions. While the use of presence increases the data processing and transfer load, it provides standardization which is an important consideration in a domain as diverse as mobile communications. Although it has its limitations, presence is a unique and compelling feature (Tang and Begole, 2003); the research aims to extend presence to mobile communications.

Lastly, the privacy of context information must be protected. Context, presented as presence, can be detailed and therefore sensitive information. While many interpretations of privacy exist, the research follows the view that privacy must be linked to the situation and the social relationship between the caller and receiver. Essentially the research provides a mechanism to control the information given to interested parties under specified conditions. Unfortunately, secondary control over imparted context is not covered by presence standards; such issues might be addressable by privacy law. In addition the research aims to make the management of privacy transparent to the user. Making privacy understandable and usable is challenging but especially important in pervasive computing environments (Karat et al., 2006).

This research combines these diverse fields into a unique whole to address the research problem.

1.5 Problem Statement

The use of mobile communications can lead to negative consequences for the receiver of a call and for bystanders in the vicinity of a caller or receiver. Problematic consequences consist of interruptions leading to disruption and embarrassment among others. Efforts by the caller or receiver to negate these consequences often lead to other undesirable effects. For example, switching a phone off decreases the availability of a person. In addition, the capability to block or divert calls is a limited solution as it does not give the receiver any opportunity to differentiate between callers or call priority.

The lack of context information is a major cause of the above problems. There is currently no standard way for a caller to be aware of the receiver's situation or vice versa. This problem is worsened by the use of multiple communication channels because there is no integrated context standard between them. Presence standards present a potential solution to the sharing of context information between caller and receiver but has privacy implications due to the nature of the information being shared.

In mobile communications a fundamental conflict exists between the desire for availability and the wish to maintain a high level of control over communication and personal privacy. Parties need a way to balance availability, interruptions leading to disruption as well as privacy requirements.

This balance needs to be addressed on a technological and social level. Technology needs to provide an efficient channel for presence sharing without compromising personal privacy. In addition ease of automation, simple management and social behaviour need to be considered to create a practical solution.

1.6 Research Objectives

The primary objective of this research is the development of a prescriptive model for controlling disruptions in mobile communications using established presence standards. To achieve this objective a number of secondary objectives must be addressed:

- investigating current theories and solutions regarding mobile communications management,
- determining the perceptions of disruption as well as current practice and possible suggestions for addressing them,
- examining the features and capabilities of presence standards for transmitting context information as well as their extensibility into the mobile domain, and

- establishing privacy requirements and the mechanisms available in presence standards for privacy protection.

The secondary objectives form a whole which, when integrated, should contribute towards the successful motivation, development and evaluation of the model.

1.7 Research Design

According to Marczyk et al. (2005, p. 1) research tries "to reduce the complexity of problems, discover the relationship between seemingly unrelated events, and ultimately improve the way we live." This last desire to better or refine is the essence of design science which is the research approach adopted by this study.

There is general agreement that design science originates from a chapter in Herbert Simon's *The Sciences of the Artificial* on the science of design. Unlike the natural sciences, which tries to understand reality, design "is concerned with how things ought to be, with devising artifacts to attain goals" (Simon, 1996, p. 114). A natural phenomenon in engineering schools, design science has achieved academic respectability through the incorporation of scientific methods. However, despite several decades of practice there is not yet agreement on all aspects of design science research and the community is still "engaged in a discourse of discovery" (Baskerville, 2008).

March and Smith (1995) state that design science is technology-oriented and "attempts to create things that serve human purposes." Hevner et al. (2004) clarify this statement by saying that design science seeks a solution to a real-world problem of interest to practice. If an optimal solution can not be found it should at least be "satisficing" (Simon, 1996, p. 119).

The problem addressed in this research is typical of a design science project in that it is a real-world problem which has unstable requirements and consists of complex interactions between elements of the problem and solution (Hevner et al., 2004). Next the philosophical grounding of the research is presented.

1.7.1 Philosophical Grounding

The theoretical perspectives reflect the researcher's basic beliefs about the world and plays an important part in determining the course of the research project. At a philosophical level two primary research paradigms exist, referred to as positivist and phenomenological. These paradigms are also referred to by other terms such as quantitative and qualitative, while some prefer to use the term interpretivism rather than phenomenological. According to Collis and Hussey (2003, p. 56) it is more accurate to call these paradigms positivist and phenomenological "because it is possible for a positivistic paradigm to produce qualitative data and vice versa."

The positivist paradigm is based on scientific observation and empirical inquiry, dealing with facts rather than values (Gray, 2004, p. 18). It believes that the world exists independent of the researcher and regardless of our awareness of it (Collis and Hussey, 2003, p. 52). When making observations the researcher should be objective and methods and conclusions must be valid and reliable (Cresswell, 2003, p. 8). Post-positivism has emerged as an important refinement of positivist thinking, responding to critics of the paradigm by realizing that we can not claim facts and absolute truths when studying the behaviour and actions of humans (Cresswell, 2003, p. 7). In contrast, the phenomenological paradigm "is concerned with understanding human behaviour from the participant's own frame of reference" (Collis and Hussey, 2003, p.53). It believes that the world is socially constructed and that truth is based on our experience of reality (Gray, 2004, pp. 21–22). The researcher is a part of the reality being observed and has an effect on it (Collis and Hussey, 2003, p. 53).

It is difficult to separate these research paradigms completely because, as theoretical perspectives change, one paradigm can comprise qualities of the other. Thus it is appropriate to consider the theoretical perspectives as part of a continuum with the two primary paradigms at each end (Collis and Hussey, 2003, p. 51). While not subscribing to a pure positivist paradigm this research does adopt a post-positivist perspective in that it aims to solve the problem above understanding it. However, design science provides the most accurate philosophical description of the research.

It is argued that design science can also be considered a research paradigm (Baskerville, 2008). A concise understanding is provided by Vaishnavi and Kuechler Jr. (2008, pp. 16–19) who compare design science philosophies with the previously mentioned paradigms. Design science changes the state of the world through the purpose-driven development of artifacts, and thus researchers are comfortable with alternative realities. Knowledge is gained through the construction and evaluation of artifacts. The researcher values

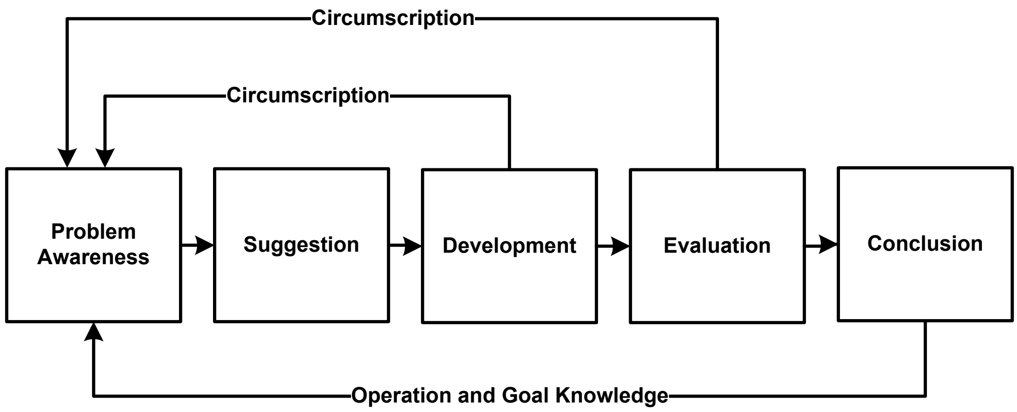

Figure 1.2: General Design Cycle (Vaishnavi and Kuechler Jr., 2008, p. 12)

creativity and control of the environment in addition to truth and understanding. It is also proposed that the philosophical grounding of a design science researcher changes as the design science project progresses. This is due to the creation of a new reality through constructive intervention and subsequent observation and interpretation of the results. Next the methodology of the research is discussed.

1.7.2 Methodology

According to March and Smith (1995) design science consists of two fundamental actions namely build and evaluate. Building constructs an artifact to address a problem and the evaluation measures how well it performs. These two activities usually follow a set process, referred to as the "general design cycle" by Vaishnavi and Kuechler Jr. (2008, p. 12). This process is illustrated in Figure 1.2.

All design begins with problem awareness. Awareness can be drawn from multiple sources, such as the existing literature in the area. Drawing from this knowledge a suggestion for a problem solution is formed. Next a process of development follows where the suggestion is implemented as an artifact.

In the evaluation phase the artifact is studied and deviations from expectations are tentatively explained. Finally the conclusion leads to the consolidation of the results in a thesis or dissertation. At each stage of the general design cycle publication of ideas and results may be produced.

March and Smith (1995) propose four artifacts as the output for design science research:

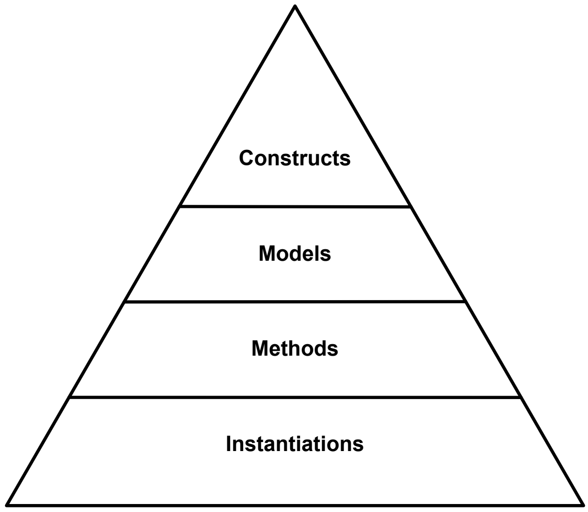

Figure 1.3: Design Science Artifacts

Constructs The vocabulary of the domain in which problems and solutions are defined and knowledge is shared. Constructs evolve and are constantly refined during the research process.

Models A representation of the problem and solution domain, consisting of relationships among constructs.

Methods The process of arriving at a solution (how-to knowledge) based on the underlying constructs and model of the solution space.

Instantiations The implementation of an artifact in a working system, demonstrating feasibility and allowing evaluation. An instantiation may precede the complete understanding of underlying constructs, models, and methods.

The four artifacts are illustrated in Figure 1.3, which also shows how each artifact builds on the previous. In addition, Vaishnavi and Kuechler Jr. (2008, p. 14) propose a fifth artifact as output, referred to as "better theories", which they define as the revelation of relationships between the elements of an artifact and experimental knowledge of artifact construction.

The implementation, partially or fully successful, is evaluated according to the functional specifications of the suggestion. This process of development, evaluation and further suggestions is iteratively performed until a fully successful artifact is designed. This iteration process is also referred to as circumscription.

After the implementation of a successful artifact the process reaches a conclusion. New knowledge is generated by the circumscription process and final artifact. Vaishnavi and Kuechler Jr. (2008, p. 12) emphasize that "the circumscription process is especially important in understanding design science research because it generates understanding that could only be gained from the specific act of construction."

In each phase of the general design cycle outputs are produced (Vaishnavi and Kuechler Jr., 2008, pp. 19–22). The artifacts produced by this research include constructs, a model and an instantiation. However, while not the focus of the research, implicit methods and better theories may be present as well. The constructs present the elements of the problem and solution domain, such as privacy and presence. The model represents the relationship between the constructs in solving the research problem.

Finally an evaluation should be done according to design science principles. In order to evaluate the research design it is necessary to compare it with accepted practice. Hevner et al. (2004) establish seven requirements for design science research:

1. Design as an artifact: the research output should be a purposeful artifact which addresses an important problem.
2. Problem relevance: the problem should be relevant in the research community.
3. Design evaluation: the functionality, completeness and usability of the research output should be demonstrated.
4. Research contributions: effective research must provide clear contributions in the research area.
5. Research rigour: rigourous methods should be applied in both the construction and evaluation of the research output.
6. Design as a search process: an iterative search process should be used.
7. Communication of research: research should be presented to a wide audience.

The continuous dissemination of research resulted in various publications, which are listed in Appendix A. These consisted of an article in an accredited journal as well as several papers presented at international and national subject specific conferences and published in the conference proceedings. To conclude this chapter a layout for the rest of the research is presented.

1.8 Chapter Layout

The research consists of eleven chapters organized in four parts, as illustrated by Figure 1.4.

The first four chapters form the background and examine the problem area, current theory, and proposed solution standards. Chapter 1 is an introduction, which includes a description of the problem area, research objectives, as well as the proposed research design. Next, Chapter 2 investigates current theories relating to mobile communications management. A taxonomy of features is created by analysing prior projects in this domain and aggregating common functionality. The identified features form an integral part of the model. Chapter 3 investigates the perception of disruption and current practice in managing such problem situations through an attitudinal survey. Theories in current literature are combined with the results of the survey to provide an in-depth look at the issues surrounding the problem. The last chapter in the background, Chapter 4, examines the advantages of distributing context information between caller and receiver, and the applicability of presence standards therein. Additionally the issue of privacy, which is highly relevant when dealing with sensitive information, is investigated and the presence standards' capabilities for protecting presence data assessed.

The next part proposes and defines the model. Chapter 5 defines the conceptual foundation of a privacy-aware presence model for managing mobile communications. Constructs are derived from the theories examined in the background and the relationships between entities are defined. This sets the stage for a detailed look at the distinguishing features of the model in the following chapters. Chapter 6 defines the model's use of context and presence to indicate availability for communication. Next, Chapter 7 defines the call management features and how the model uses presence information in call management. Finally, Chapter 8 examines the important issue of privacy and presents a novel contribution in using social relationship theory to control access to sensitive data.

The third part evaluates the model through an instantiation and discussion of its extensibility. Chapter 9 evaluates the model by presenting an instantiation, an implementation of the constructs and model artifacts.

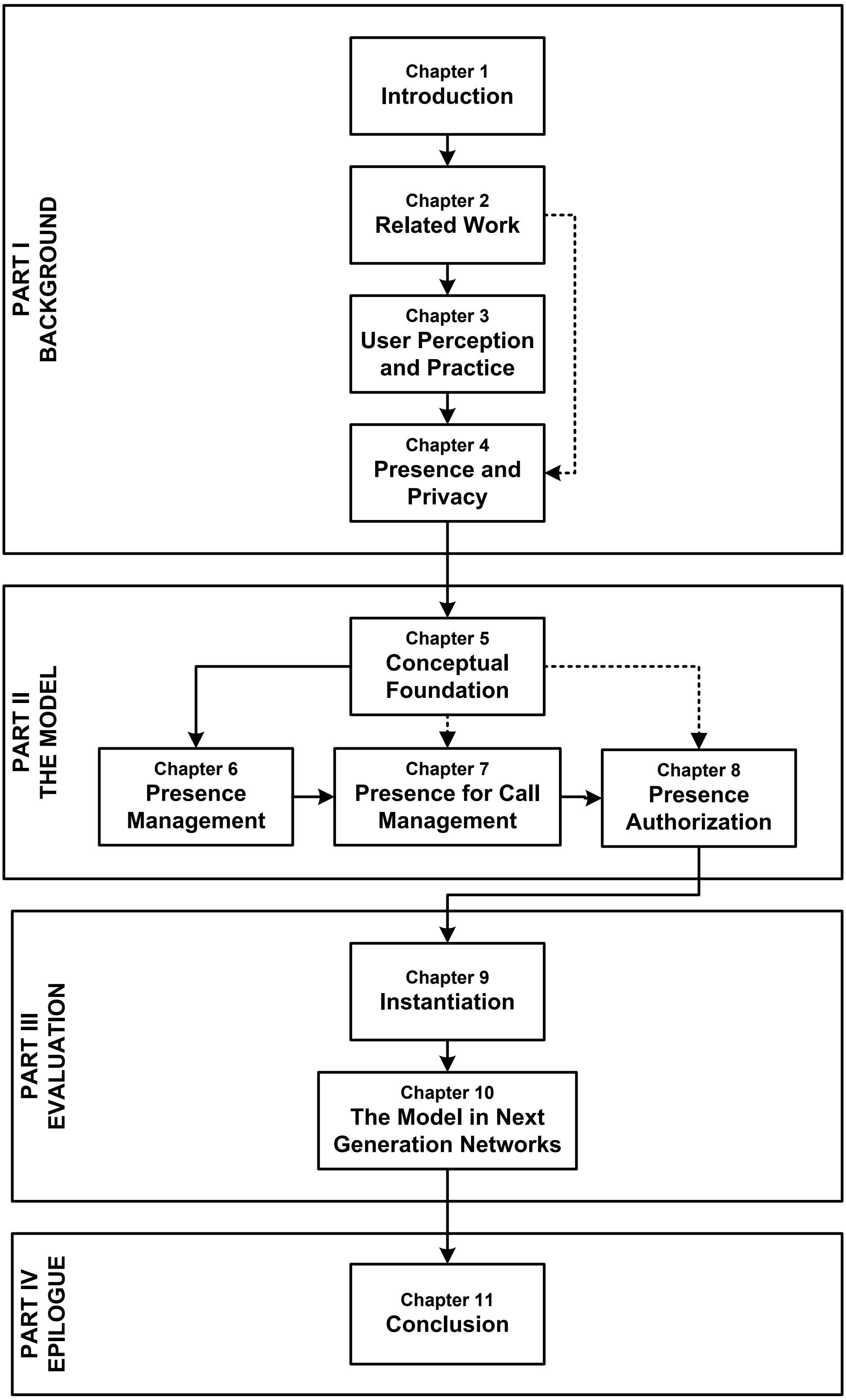

Figure 1.4: Chapter Layout

Chapter 10 concludes this part by showing the extensibility of the model in future communication networks.

The epilogue, consisting of Chapter 11, summarizes the research and evaluates it to determine whether the objectives have been achieved. It also includes a discussion for further research.

1.9 Looking Ahead

In summary, the author wishes to borrow the sentiments of Booth et al. (2008, p. 186) and reiterate it for this research:

> Reader, after my best efforts, here's what I believe—not the whole or final truth, but a truth important to me and I hope to you. I have tested and supported that truth as fully as time and my abilities allow, so that you might find my argument strong enough to consider, perhaps to accept, maybe even to change what you believe.

Chapter 2

Related Work

Writing about future transaction systems, Chaum (1985) expressed the opinion that "advances in information technology have always been accompanied by major changes in society". The rise of mobile communications in the late twentieth century, and the accompanying changes in communications capabilities, is another example which confirms this statement. Enabled by mobile communications, the anytime anywhere communications paradigm has changed the way we manage our activities, do business and maintain relationships. However, as the previous chapter has shown, these changes have not come without some undesirable consequences.

Issues, such as disruption and managing communications, waste valuable resources including our attention and time. Current technologies provide little assistance, forcing helpless users to endure these inevitable problems. The consequences vary from personal embarrassment and annoyance to serious results such as fatalities. Several common examples of these consequences can be given: forgetting to silence your phone and it ringing during an important event, a multitude of annoyed cinema-goers when a phone rings during a movie or the number of deaths caused by drivers talking on a phone. To minimize negative communication effects laws are passed and acceptable etiquette is defined, for instance the banning of phone use while driving or silencing a phone during dinner parties. In addition, research is studying ways in which technology can assist users.

This chapter reviews research efforts addressing the undesirable consequences of mobile communications using technology. It takes a concept-centric approach and synthesizes the literature by discussing each identified concept. Such an approach organizes the review and helps the reader to make sense of the accumulated knowledge on a topic (Webster and Watson, 2002). First the approaches researchers have taken to address problem are reviewed. Next the types of information used in these approaches are discussed. This is followed by examining the functional solution architectures. Finally, the approaches to common privacy issues are reported.

2.1 Communication Management Approaches

Researchers have approached the problem of managing availability for mobile communications from different perspectives. Three main techniques for coordinating communication can be identified (De Guzman et al., 2007):

- allowing the receiver to decide on incoming calls,
- negotiating the call, or
- displaying an availability cue to the caller.

These techniques are reviewed in more detail in the next sections.

2.1.1 Receiver-oriented

Receiver-oriented approaches are practically easy to implement as they require no interaction with the caller. Usually such approaches include the ability to change the call notification, for example to silent or vibrate. The action is based on the callers identity, which needs to be available for such approaches to be useful. A drawback of these approaches is that it forces receivers to make a decision on communication requests in advance and thus the possibility exists that important calls are missed or unimportant calls are accepted (De Guzman et al., 2007).

The *Taming of the Ring* project allowed users to pre-record voice messages which get played back to callers depending on the receiver's context and the callers identity. Callers could also indicate the urgency of the call which allowed the receiver to make a more informed choice. Findings indicated that callers liked getting voice feedback (instead of a normal call rejection) and they welcomed being able to indicate the urgency of the call (Pering, 2002).

Quiet Calls allowed mixed-mode synchronous communication whereby a caller could talk while the receiver responds through pre-recorded messages

(Nelson et al., 2001). This allowed the receiver to move to a location where speech would not be disruptive, before starting to talk. While not addressing the decision to accept or deny a call, the system allows a receiver to continue a conversation without disturbing others in the immediate range of the conversation.

The *Personal Reachability Management* system allowed receivers to specify the situations in which they are willing to receive a call (Reichenbach et al., 1997). Rules are specified for a situation and are evaluated in order to determine when a call is accepted. Support for complex, situation independent rules were abandoned because of the complexity of evaluating them (Rannenberg, 2000b). Indeed it was found that most receiver's preferred having only a few rules such as fully reachable, reachable in case of emergency and not reachable (Ammenwerth et al., 2000). This is used in combination with caller information to determine whether the receiver is notified of a call by the phone ringing. Thus, before the receiver is personally involved the call is negotiated by the personal reachability manager (Reichenbach et al., 1997). Of course even if the rules allow a call through the receiver can still manually deny it.

2.1.2 Negotiated

Negotiated approaches allow a caller and receiver to negotiate an agreeable time to communicate. Previous research has focused on negotiating when to speak and what will be spoken about (Reichenbach et al., 1997; Wiberg and Whittaker, 2005).

Wiberg and Whittaker (2005) modelled a lightweight negotiation of an interaction request. Callers can indicate that they would like to talk now or at a later time. The receiver can accept the proposed time or make a counter proposal for a different time. Negotiation continues until an agreement is reached. However, even though the negotiation is lightweight it may still be disruptive as it requires the attention of both the caller and receiver.

The system developed by Reichenbach et al. (1997) allowed negotiation based on the content of the call. In setting up a call callers can specify additional information about the requested communication. This call context is transmitted to the receiver during the signalling phase and is then negotiated by the respective personal reachability managers (Reichenbach et al., 1997), which are software agents acting on behalf of the user. Depending on the receiver's preferences there are three possible outcomes of a call proposal (Rannenberg, 2000a,b):

- The call is accepted and the caller (and receiver) is notified that the call is proceeding,
- The call does not meet the receiver's specifications and is rejected, with the caller having the option to leave a message or request a callback, or
- Additional information is requested by the receiver which the caller needs to provide for the call to proceed.

Negotiation of a call is implemented using a simple three-step model. First a caller specifies a proposal containing the requested measures and offers. This proposal is transmitted to the receiver. Second the receiving personal reachability manager compares the proposal with the receiver's preferences and produces a counter proposal which is sent back to the caller. Third the caller's personal reachability manager compares both proposals and connects the call if they match. If a match is not found the caller is asked whether the receiver's counter proposal should be accepted (Rannenberg, 2000a,b). To avoid repeated negotiation inquiries or failures a three-level security scope is defined whereby callers can overstate their requests and receivers understate their offers. This allows the personal reachability manager to automatically conclude negotiation when there is a small discrepancy in proposals (Rannenberg, 2000a,b).

2.1.3 Caller-oriented

From the caller's point of view there is currently no information regarding the receiver's availability before making a call. Caller-oriented approaches aim to provide cues about the receiver's availability to the caller before making a call. This allows the caller to make an informed decision on whether to proceed with the call, balancing his or her communication needs against the receiver's context (De Guzman et al., 2007).

The *Live Addressbook* project explored the usefulness of providing information such as availability and location about a caller's contacts. Their main findings indicated that users will attempt to keep their information updated but automatic detection is a useful strategy. They also found that availability information is often unreliable and that an interactive negotiation may be preferable (Milewski and Smith, 2000).

The *AwareNex* project experimented with the use of connectivity, device usage and communication activities as indicators of availability. Running on a Palm or Blackberry device, their system assists a caller with the most likely locale to reach another user being highlighted by a simple algorithm (Tang et al., 2001).

Another system, tested in a hospital environment, the *AwarePhone*, tried to achieve context-mediated social awareness by presenting context cues to users. They argued that having multiple context cues (e.g., calendar information, status and location) is advantageous, but also indicate some privacy concerns about revealing location information (Bardram and Hansen, 2004).

Raento et al. (2005) designed and developed a context-aware platform called *ContextPhone*. One of the main goals of the platform is to provide context as a resource for use by applications; it represents the user's sensed context in a human-readable form for social interaction. Built on top of this platform the *ContextContacts* application aims to enhance the awareness of others' situations by allowing users to exchange context information automatically (Oulasvirta et al., 2005). The application does not reason about the receiver's availability but instead offers context cues to callers (Raento et al., 2005). Explicit actions create context information and the receiver can see how callers use that information (Raento and Oulasvirta, 2008). Thus the application functions much like a presence system with dynamic status updates according to user context.

The *Calls.calm* project allowed callers to see more information about the receiver through an interactive web page, displayed on the phone. The page presented callers with a personalized greeting, context information, continuity information and available communication channels. Their experimental results suggest that a smooth transition between synchronous and asynchronous communications is beneficial (Pedersen, 2001).

The *Lilsys* project used sensors to sense elements of the receiver's environment such as sound, light and motion. This information is used to infer moments of unavailability, which is displayed to potential callers. Their results indicated that callers often worked around availability indicators, still making calls but starting the conversation with an explanation for the interruption (Begole et al., 2004).

The *Context-Call* project gave callers a WAP-based indication of the receiver's context. After viewing the information the caller has the choice to proceed with the call, cancel it, or leave a message (Schmidt et al., 2000).

In addition to these mobile projects, much work on supporting collaboration by providing awareness has been done in work-oriented desktop environments. A good example is the *Coordinate* project (Horvitz et al., 2002), which uses predictive learning models to forecast presence and availability. By examining information such as a user's calendar, location, and activity across multiple devices, the expected cost of interruption for a user is calculated. Other projects supported by this research, *Notification Platform* (Horvitz and Apacible, 2003) and *Bestcom* (Horvitz et al., 2003), have focussed on the further development of interruption models, user privacy concerns, and automating the mediation of communications.

Finally, the *UPCASE* project focused on using context to mobile applications and networking services (Santos et al., 2010). Their results showed that social networking services provide a convenient way to share information between users.

A visual representation of the three approaches and the projects following each is presented in Figure 2.1. As can be seen the majority of related work focused on a caller-oriented solution to the problem. These solutions used a variety of context information to indicate the receiver's availability to callers. The next section will review these context sources.

2.2 Context Sources

It is possible to infer availability from a wide variety of data. This section will review what data researchers have found to be most successful.

Location is an important part of user context as is evident from the number of projects which have used such information (Milewski and Smith, 2000; Tang et al., 2001; Pedersen, 2001; Bardram and Hansen, 2004; Begole et al., 2004; Raento et al., 2005; Khalil and Connelly, 2006; De Guzman et al., 2007; Bentley and Metcalf, 2008). Mobile phones make it relatively easy to retrieve location information as they are constantly connected to a cellular network. In addition an increasing number of mobile phones include an integrated GPS receiver and other sensors which can connect to short-range networks via Bluetooth or WLAN technology. This allows extremely accurate location information to be retrieved.

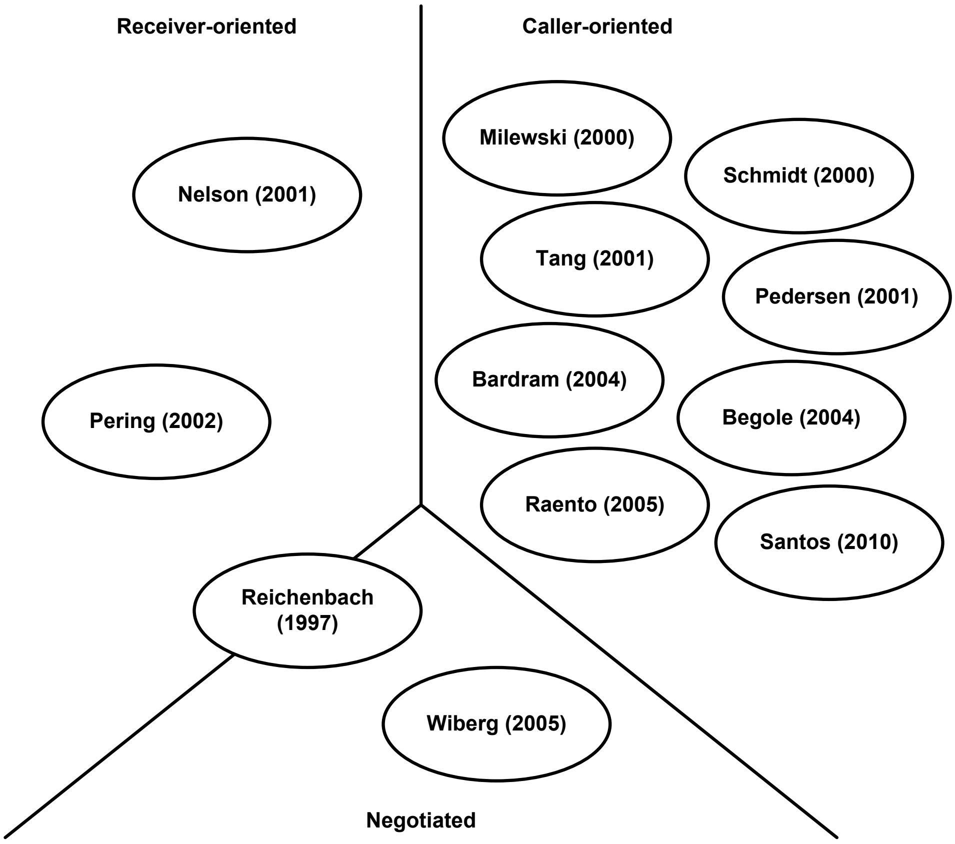

Figure 2.1: Communication Management Approaches

Network location, based on the cellular tower which the phone is connected to, is a popular source of information because it is available on most mobile phones. Projects which have used this resource include Raento et al. (2005). However, because network location is based on a unique cell tower identifier it is without meaning to a person. Although costly, a possible solution is to automatically fetch a description from a service that provides city and district names for each identifier (Oulasvirta et al., 2005). Alternatively the user could be prompted to enter a description when the phone moves to a new location, which is the commonly chosen approach (Oulasvirta et al., 2005; Khalil and Connelly, 2006). Oulasvirta (2008) also took the approach of omitting infrequently visited locations while including the last known location to support understanding of where the person is coming from.

An alternative source of location information is through the use of GPS receivers (Marmasse et al., 2004). This allows a highly accurate determination of a person's location, but suffers from the same lack of meaning as cell tower identifiers and thus also needs some form of naming service.

Finally it is also possible to infer location from sensors such as infrared, Bluetooth or WLAN (Bardram and Hansen, 2004). Small components can be worn on a person and will receive a signal when in proximity to a beacon. As the beacon's location is known the persons location can be inferred. Even a mobile phone can be used to sense context from the surrounding environment. Kononen et al. (2010) have found that such devices, with limited computational ability, can still calculate reasonably accurate context data.

In addition to the location, Oulasvirta et al. (2005) also captured the *duration of stay* as a contextual cue. It was found that as time progressed the accuracy of information could be lowered without negative consequences, e.g. >1h or >1d could be used as simple categories (Oulasvirta, 2008).

Previous studies indicate that users perceive a strong correlation between disclosing location and activity information (Bentley and Metcalf, 2008; Khalil and Connelly, 2006). Different kinds of *activity* information were frequently seen as important. *Physical availability* was found to be an important context cue (Pedersen, 2001; Wiberg and Whittaker, 2005; Raento et al., 2005; Khalil and Connelly, 2006; De Guzman et al., 2007; Santos et al., 2010). This encompasses activities such as sleeping or driving and is often not easy to determine. However, sensors such as small motion detectors or GPS receivers can be used effectively in this regard. The distance between the receiver and the phone was found to be relevant by De Guzman et al. (2007). Callers may not attempt to initiate communication if they know that the receiver is unlikely to reach the phone in time.

Social availability refers to the immediate social environment a person is in. This can include the setting as well as people speaking in the vicinity of the receiver. A popular method to gather such information is the calendar appointments of a person (Tang et al., 2001; Bardram and Hansen, 2004; Raento et al., 2005; Khalil and Connelly, 2006). Another method is to use short-range sensors, such as sound and motion, and inference techniques (Raento et al., 2005). Users can also judge their own social availability and indicate as much on incoming calls (Nelson et al., 2001; Pedersen, 2001; Wiberg and Whittaker, 2005; De Guzman et al., 2007).

Emotional availability refers to whether a receiver is in the mood to talk (Milewski and Smith, 2000; De Guzman et al., 2007). Automated mechanisms to postpone calls to a later time have been proposed by Wiberg and Whittaker (2005).

Whether the receiver is currently occupied is indicated by *task status*.

Many systems have been developed to infer availability from calendar information, device usage and real-time sensors (Milewski and Smith, 2000; Khalil and Connelly, 2006; De Guzman et al., 2007; Bentley and Metcalf, 2008).

De Guzman et al. (2007) suggested that incoming callers are interested in the current *time* at the receiver's location. This could indicate whether the call would be disruptive because, for example, the receiver is still sleeping.

Various *environmental cues* have been used by systems as context information. As sensors change state inferences can be made about a person's availability. Relevant sensors could be monitoring motion, sound or phone activity (Begole et al., 2004; Raento et al., 2005; Bentley and Metcalf, 2008; Santos et al., 2010). The importance of environmental cues regarding location and activity was also noted by Bentley and Metcalf (2008). In a real-world study based on recording mobile phone conversations, they found indicating availability to be the second most common use of such information (after disclosure as a means of creating social awareness). Their study also emphasized that users can make accurate predictions about availability based on previous interactions and communications. Thus it may not be necessary to implement complex algorithms to deduce availability, but only to inform people of simple context such as environmental cues.

One of the most flexible ways to indicate context is for the receiver to specify a *free text* message which callers can view (Schmidt et al., 2000; Begole et al., 2004; Raento et al., 2005). This allows greater freedom of expression but places the burden of updating context on the user as automatic sensing is not performed.

The *manner of communication* which the receiver or caller requests can also be a valuable context cue. Apart from indicating the communications method this could also allow a caller to deduce the availability of the receiver. Alternative to voice calls include email, instant messaging and text messaging (Reichenbach et al., 1997; Tang et al., 2001).

The *security requirements* of the caller and receiver was also deemed important by Reichenbach et al. (1997). This included maintaining the anonymity of the caller if so desired, or confirming identity through digital certificates. In addition the encryption of the communications channel allows communication partners to transfer context information without fear of interception.

Finally, the context of the caller was also considered by Reichenbach et al. (1997) and Pering (2002). This includes information about the *subject* of the call and the *urgency* of the call, as judged by the caller. Urgency can be expressed by a statement based on the caller's own judgement, indication of a function, presenting a voucher, providing a reference and offering a financial surety (Reichenbach et al., 1997; Rannenberg, 2000a,b). Simulation studies indicated that users valued knowing the subject of the call before answering (Ammenwerth et al., 2000).

To conclude this section Table 2.1 presents a summary of the context information included by previous research efforts. Where possible related context items have been grouped in close proximity to each other. The projects have also been listed in chronological order so that the perceived importance of various context sources can be judged over time. It can be seen that location has been used very often, while context such as physical availability and task status has recently gained popularity. Methods to indicate the context of the caller have been surprisingly few.

2.3 Functional Architectures

There are various ways in which context can be sensed, stored and communicated. This section will examine the techniques employed most commonly by previous projects.

2.3.1 Communications Strategy

The simplest way to communicate context information between receiver and caller is over the Internet. Many projects employed an architecture similar to instant messaging, where information is pushed to callers when it changes (Milewski and Smith, 2000; Tang et al., 2001; Bardram and Hansen, 2004; Begole et al., 2004; Raento et al., 2005). Santos et al. (2010) also published information to social networking sites, which are currently a popular way to convey status information. In these cases the data capabilities of the phone and network will have a significant effect on the speed of communication and usability of the system.

However, mobile phones also allow other communication methods, such as text messages. These channels can handle lightweight communications very affectively and can travel with minimal routing between communication partners (Wiberg and Whittaker, 2005; Raento et al., 2005). However, delivery is best effort and thus cannot be guaranteed.

Table 2.1: Context Sources

	Location	Duration of stay	Physical availability	Social availability	Emotional availability	Task status	Time	Environmental cues	Free text	Manner of communication	Security requirements	Subject (caller)	Urgency (caller)
Reichenbach et al. (1997)										X	X	X	X
Milewski and Smith (2000)	X				X	X			X				
Schmidt et al. (2000)									X				
Nelson et al. (2001)				X									
Pedersen (2001)	X		X	X									
Tang et al. (2001)	X			X						X			
Pering (2002)													X
Bardram and Hansen (2004)	X			X									
Begole et al. (2004)	X							X					
Raento et al. (2005)	X	X	X	X				X	X				
Wiberg and Whittaker (2005)			X	X	X								
Khalil and Connelly (2006)	X		X	X		X							
De Guzman et al. (2007)	X		X	X	X	X	X						
Bentley and Metcalf (2008)	X					X		X					
Santos et al. (2010)			X					X					

Projects have also used the voice channel to communicate context to the caller. Nelson et al. (2001) and Pering (2002) have employed this method to play back pre-recorded messages based on the receiver's context and caller's identity. Assuming such playback capabilities, this presents a solution which will work between any two mobile phones.

Schmidt et al. (2000) created a prototype implementation using the wireless application protocol (WAP). It allowed users of a WAP-enabled mobile phone to specify and store their context by sending a WAP-request to a WAP server. A special WAP application, which replaced the normal interface for making phone calls, retrieved and displayed the context to callers based on the number they were calling.

Finally, Reichenbach et al. (1997) employed data communication during the signalling phase of a call. Because they did not present context to the caller prior to the call this allowed the call to be negotiated across the mobile network. However, this technique suffers the drawback of being slow and thus leads to a poor user experience (Ammenwerth et al., 2000).

2.3.2 Data Processing and Storage

Projects which used communication systems based on instant messaging architectures stored context information on a server in the network (Milewski and Smith, 2000; Tang et al., 2001; Bardram and Hansen, 2004; Raento et al., 2005). This was also the case for Schmidt et al. (2000) and Pedersen (2001) who stored receiver information in a database on a network server and exposed the information to callers through HTML or WML pages.

It was found to be important that data is processed automatically, thus giving users a sense that the information is updated and timely (Raento and Oulasvirta, 2008). However, for users with a large number of contacts a high frequency of context updates would occur. An alternative solution is to use a pull-only mechanism, choose which contacts to follow or use an algorithm to indicate a change in behaviour (Oulasvirta, 2008). A common architecture consists of three parts: first there is a presence publisher which gathers sensor data and transmits it. Second a presence listener receives data and integrates it into the application user interface. Third a customization component allows logging of all actions (Raento et al., 2005).

Raento and Oulasvirta (2008) implemented an approach based on the principle of self-disclosure and reciprocity – if context or a subset of information is not disclosed it will also not be received . However, in this case users have no way of knowing when their information is being looked at and may feel unnecessarily tracked (Oulasvirta et al., 2005). To address this the application provides users with control over who receives what context. Contacts can be grouped and information can be sent to all users or for a group at a time. Only single identities are supported where users are bound to their real name and mobile phone. However, a user can be a part of any group allowing control over what information is transmitted to them (Raento and Oulasvirta, 2008).

In contrast with the above approaches, Reichenbach et al. (1997) placed a high value on the privacy and security of context information. Steps that were taken into consideration include securing the environment by not implementing the system as a network service, detecting repeated information requests, protecting against unintentional information exposure and the ability to audit the system. It was recommended that data should be placed where the owner can control it, such as on the phone itself (Reichenbach et al., 1997). Additionally messages were secured by encrypting them, location information was managed by the system and identity information was authenticated using digital certificates (Reichenbach et al., 1997). Their system consisted of two parts which handled communication negotiation and forwarding of requests respectively (technological limitations prevented a unified implementation). The parts of the system exposed several services such as user interface components, a phone book, evaluation of the current communication context, reachability rule storage, an interpreter for reachability rules, security functions and connection methods (Reichenbach et al., 1997).

2.3.3 Usability and User Interface

Previous research has mainly focused on extending or replacing the normal interface for making calls (Reichenbach et al., 1997; Milewski and Smith, 2000; Schmidt et al., 2000; Pedersen, 2001; Tang et al., 2001; Bardram and Hansen, 2004; Raento et al., 2005). Such extensions extend the phone book by adding available context information next to each person in the list. Information can be presented as text, such as by (Tang et al., 2001), using icons or symbols such as by (Milewski and Smith, 2000; Begole et al., 2004) or a combination of the two approaches such as by (Raento et al., 2005).

The impact of the user interface was studied closely by Raento et al.

(2005). Their ContextContacts application was tightly integrated into the mobile phone, extending the existing call application with context information. In addition it was found necessary to include context in the recent calls list as it is used frequently to initiate a call (Oulasvirta et al., 2005). The positioning of context close to content and associated communication functionality was deemed an important factor in the integration with the phone (Oulasvirta, 2008). Context was predominantly displayed using icons to save space, support visual search and grab attention (Raento and Oulasvirta, 2008). Information was also slowly greyed out as it became stale, thus giving a visual indication of its accuracy (Oulasvirta et al., 2005). Context information was also quickly obtainable with the application supporting access in less than 3 seconds; it was found that on average users looked at this information for 1 to 4 seconds before placing a call (Oulasvirta, 2008).

Another approach is to expose an interface in addition to the default phone application (Nelson et al., 2001; Pering, 2002; Wiberg and Whittaker, 2005). Such an approach allows greater freedom in the amount of information which can be presented to the user. It also allows the user access to different communications mechanisms, such as playing back pre-recorded messages (Nelson et al., 2001; Pering, 2002). However such approaches requires more attention from users as they have to navigate away from default applications to manage communications.

2.4 Privacy Issues

As context information is of a highly personal nature it is natural to assume that users will be concerned about who has access to such information. It may be said that social networking and media is changing this attitude and that users are more willing to share personal information. However, further research is needed to confirm this – users may just be limited in their knowledge of the risks or lack good alternative applications which respect privacy. This section reviews reports from previous studies regarding privacy concerns. Three salient points are given by Schmidt et al. (2000) which summarize many of the privacy concerns people perceive:

1. People want to be in control of what about them is visible to others.

2. People want to know what others know about them.

3. People like to share information selectively.

In general, research indicates that privacy is less important than is currently thought and that users are willing to share personal information in exchange for useful services (Khalil and Connelly, 2006; Raento and Oulasvirta, 2008). However, this does not mean that the privacy of information is not valued (Danezis et al., 2005). Rather, the social relationship as well as the type of information influences privacy.

Research has shown that privacy concerns depend significantly on the relationship between caller and receiver (Consolvo et al., 2005; Khalil and Connelly, 2006; Raento and Oulasvirta, 2008). Users are more likely to share availability information with social relations such as significant other, family and friends (Khalil and Connelly, 2006). Similar findings have been shown for the sharing of location information (Consolvo et al., 2005). Other factors influencing the willingness to share information with a particular person include the user's current activity and mood, the level of detail as well as why the information is needed (Consolvo et al., 2005).

It has also been shown that different kinds of context are perceived with varying levels of privacy. Information such as location and activity are perceived as more sensitive than company and conversation (Khalil and Connelly, 2006). However, it has been found that users often share such information in as much detail as possible, or not at all (Consolvo et al., 2005). This is in contrast to other projects which have found allowing granular information to be important (De Guzman et al., 2007; Raento and Oulasvirta, 2008).

Controlling the granularity of information displayed to different groups and having the ability to fake some or all presence information is an important aspect in giving users full control over their information (Raento and Oulasvirta, 2008). However, granularity also provides callers with different views on the receiver and allows them to infer context with varying levels of accuracy (De Guzman et al., 2007). Thus there is a fine balance between receiver's maintaining control over information and such information remaining useful for callers.

While anonymity can only be achieved if supported by the underlying network, pseudonymity is an important and desirable feature which systems should support (Reichenbach et al., 1997; Raento and Oulasvirta, 2008). This affords callers a degree of control over their information as well.

2.5 Conclusions

From the summary of related work presented in this chapter it can be seen that many people are dealing with the management of mobile communications. It is clear that there are numerous approaches, each with positive and negative aspects.

It would seem logical to put control in the hands of the receiver so that they can decide when and how to be contacted, rather than allowing the caller to interpret any context indicators. With this in mind it is interesting that the majority of previous research focused on the opposite perspective (caller-oriented approaches in Figure 2.1).

One of the greatest difficulties for researchers in this area is the great diversity in mobile communications. Not only do mobile phones vary a great deal but networks and operators also function differently within and between countries. This makes it difficult to achieve a universal solution to the problem. Standardization is difficult to achieve, but Chapter 4 examines a common messaging and data representation solution which can be used in this regard.

From the amount of work going on in this area one can conclude that the problem is real and topical. However, to confirm the opinion of actual users the next chapter presents the results from a survey on the perception and practice of mobile communications.

Chapter 3

User Perception and Practice

While the previous chapter presented several projects focused on managing mobile communications almost no data exists to understand the problems and possible solutions for communications management in the mobile domain. While numerous surveys have been conducted in the field of mobile communications, few have focussed exclusively on communication management. While the problem seems obvious most research has not looked at user perception in this area, which is what this chapter intends to do.

Some of the research in this area includes a Finnish survey which was conducted by Aarnio et al. (2002) to analyse the adoption and use of mobile services. Nickerson and Isaac (2006) developed a research model and survey to examine the acceptable use of mobile phones in social settings. In a multinational survey among students they examined how user characteristics affect the preference for mobile phone use in different social settings (Nickerson et al., 2008). A study of the factors influencing the adoption and use of mobile phones was conducted by Kaba et al. (2008). They focused on the microeconomic factors in developing countries by collecting data from users in Guinea.

As descriptive survey was chosen as research method for its ability to provide attitudinal information which can be generalized from a sample to the population (Cresswell, 2003, pp. 153–162). Surveys provide high measurement reliability and construct validity, but can lack depth and sometimes be too context specific (Mouton, 2008, pp. 152–153). The next section defines the survey research questions.

Table 3.1: Variable to Secondary Research Question Map

Variable Name	**Research Question**
Practice	How does the respondent make use of Caller ID and switching the phone off to manage calls?
Privacy and trust	Is the respondent concerned about privacy?
Willingness to pay	What is the respondent's current spending pattern?
Perceptions	What is the respondent's perception of various call-management features?

3.1 Research Questions

The survey focused on two primary research questions – how people manage their mobile communications use, and what features they perceive as useful in a communications management solution. By analysing the responses it can be judged whether previous research has focused on relevant mechanisms to solve the problem. It also establishes the need for further work in this area, for which the collected data may serve as possible input. However, the current chapter does not judge whether a communications management solution will enhance the experience of mobile communications. Nor does it conclude whether such a solution is possible or what the underlying architecture should be.

To evaluate user practice and build a picture of mobile communications usage the survey also defined several secondary research questions. These research questions were aligned to four variables: practice, privacy and trust, willingness to pay and perceptions. Within these variables specific survey items were identified based on informal user discussions and personal experience. Table 3.1 maps the variables mentioned above to the secondary research questions.

The survey probes user practice by asking users to provide a description of their present communications management behaviour. Questions include how often users switch off their phone to avoid being disturbed, as well as their use of the Caller ID feature in both a business and social environment. Currently these are the only methods available to users for implementing their own communications management.

Next the survey examines the perception of privacy threats and trust in the mobile communications environment. This concerns the misuse of information, specifically Caller ID, by receivers without the knowledge of the caller. The level of trust users place in their network operator is also questioned.

The users' willingness to pay for mobile applications and services is estimated by their current mobile phone upgrade practices and their consumption of mobile applications and services. Because price plans and contracts differ across the globe the items are predominantly based on the South African market. However, it is believed that the questions are general enough to provide useful information in other markets as well.

Lastly the survey asks the user's opinion on various communications management features, both for incoming and outgoing scenarios. These features, based on previous research (Reichenbach et al., 1997), include caller identification, the subject of the call and the call priority. In addition, users were also asked to identify features they perceived as useful. Their willingness to wait, before a call is established is also questioned. However, such data needs to be confirmed through further observation.

By examining these research questions, the survey aims to establish current communications management practice as well as the expectations of a potential communications management solution – possibly in the form of a mobile communications application or service. The next section elaborates on the research design.

3.2 Research Design

The survey was conducted during 2006 using a web-based questionnaire, which was hosted by the Nelson Mandela Metropolitan University (`http://www.nmmu.ac.za/mobile`). The design of the questionnaire was done in conjunction with a statistical consultant and a pilot study by selected respondents was used to produce the final version. Appendix B lists the survey items together with abbreviated answer options.

The population of interest consisted of people who regularly use mobile communications and often find themselves in busy situations. Therefore the sample intended drawing respondents from professional environments with a fairly high degree of technical knowledge. To this end the survey was submitted to two well-known information technology mailing lists – the International Association for Information Systems (`http://www.isworld.`

Table 3.2: User Characteristic Items

Variable Name	**Item on Survey**
Demographics	Questions 1–4, and 8: gender, age, job title and function, and number of mobile phones owned
Availability	Questions 5–7: personal secretary, meetings, and unavailability for communications
Technical Use	Questions 9–11: network services, voicemail, and phone features

org) and the Computer Society South Africa (http://www.cssa.org.za). An article explaining the purpose of the survey and inviting participation was also posted on a popular South African technology news site (Thompson, 2006).

During the data collection period, 89 valid survey responses were received. All responses were treated anonymously and confidentially. Ideally a higher response rate would have been preferred, but restrictions were imposed by time and budget constraints. The survey was conducted over a two month period to allow for sufficient data analysis thereafter. The research budget also limited the promotion of the survey to the above-mentioned channels. However, it is believed that the results will be confirmed by future surveys using a larger sampling frame. The following sections present an overview of the survey design as well as a description of the data analysis.

3.2.1 Survey Design

The survey questions were designed to answer the previously defined research questions (see Section 3.1) while also allowing the deduction of characteristics which could be used to profile users. As such, three additional variables were introduced: demographics, availability and technical use. Table 3.2 and 3.3 maps the variables to specific questions on the instrument.

The variables in Table 3.3 – practice, privacy and trust, willingness to pay and perceptions – were discussed in Section 3.1. The demographical data intended to provide a more well-defined description of each respondent. While the survey intended to assess the status of all respondents, demographical data may reveal facts which could be of interest. Additionally, Question

Table 3.3: Research Question Items

Variable Name	**Item on Survey**
Practice	Questions 12–14: switching off and Caller ID
Privacy and trust	Questions 15–18: Caller ID, misuse of information, and trust in network operator
Willingness to pay	Questions 19, 20 and 24: hardware and software or services
Perceptions	Questions 21–23: notifications and initial waiting time

8 eliminates potentially undesirable respondents who do not own a mobile phone.

Next each respondents availability for communications was established. Of interest was whether the respondent had a personal secretary performing the role of communications manager. As such respondents were not the focus of the survey it tried to determine whether their responses, in terms of communications management, differed from others. Preferable respondents were those who manage their own communications and regularly find themselves in situations where they are unable to answer the phone.

Finally, since respondents with technical expertise were of interest, the survey tried to gauge the use of technical phone features and network services. Despite the fact that the use of such features are dependent on many factors, such as usability, these are believed to be the best available measure of technical expertise. The survey also questioned the users' willingness to use a technical feature if it provided a useful service.

3.2.2 Data Analysis

For the previously defined research questions the survey data is reported primarily as response frequencies and percentages for each question. Percentage values have been rounded which result in some totals not adding up to 100%. This is also the case for the previously defined user characteristic variables.

Additionally an analysis of the user characteristics reports any visible clusters. Thereafter any discovered clusters are examined to determine their responses to the research questions. The *effect size* of these responses are determined which assesses how big the practical importance of the relationships are (Cohen, 1988).

Finally, Chi-square analysis using the demographic data is reported, using the research questions as well as any visible clusters to determine statistically significant relationships. Following, the survey data is presented according to the variables and structure defined above.

3.3 Results

This section presents the frequency distributions for the valid responses (see Section 3.3.1). Thereafter visible clusters are reported (see Section 3.3.2) as well as the strength of the cluster relationships to the research questions (see Section 3.3.3).

3.3.1 Frequency Distributions

The variables defined in Table 3.2 and 3.3 are presented in the same order. Percentages in figures have been rounded to the nearest decimal for convenience.

Demographics

Reviewing the responses from a demographical perspective it was found that a wide range of users completed the survey. The male (63%) versus female (37%) ratio was realistic for the target group. While various age groups were represented in the sample, over 96% of the respondents were 25 years or older. Out of these 26% were 45 years or older. From an employment perspective the majority (42%) of respondents were educators, with researchers/analysts (18%) and consultants (15%) the next highest groups. All the respondents used at least one mobile phone on a regular basis, while 15% of the respondents used two or more mobile phones. This confirms that all the responses were of value to the study.

A Chi-square test using the demographic data revealed minimal significance between the research questions and the gender, age and mobile phone ownership of respondents. Relationships at the 5% and 10% level of significance are discussed in later sections, where appropriate.

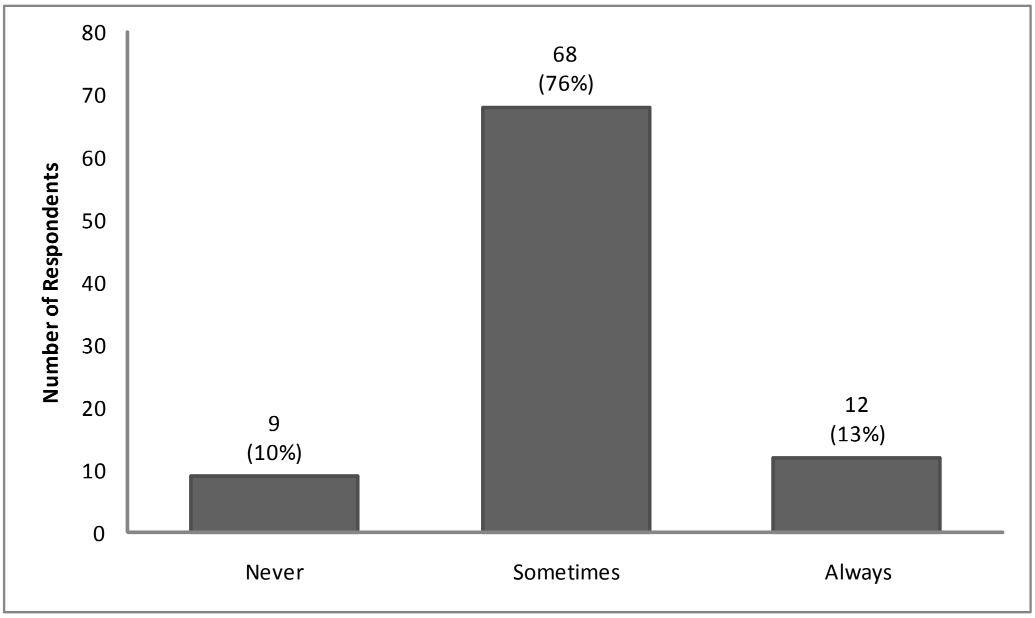

Figure 3.1: Unavailable for an Important Call

Availability

To understand the respondents' availability for communications the survey asked several questions related to the amount of time they find themselves in situations where mobile conversations were awkward. The overwhelming majority of respondents (90%) did not have a personal secretary to manage their communications.

Most of the respondents regularly attended meetings which was the predefined situation during which the use of mobile communications is generally minimized. The average number of meetings attended was five per week, while some respondents attended that many every day.

When asked how frequently they were 'unavailable' while expecting an important call the results were as indicated in Figure 3.1. As can be seen the majority of respondents do sometimes miss an important call because of the situation they find themselves in.

Technical Use

The survey evaluated respondents' technical use based on the usage of network services as well as their mobile phone. Figure 3.2 depicts the responses regarding the use of three network services – Caller ID, the short message

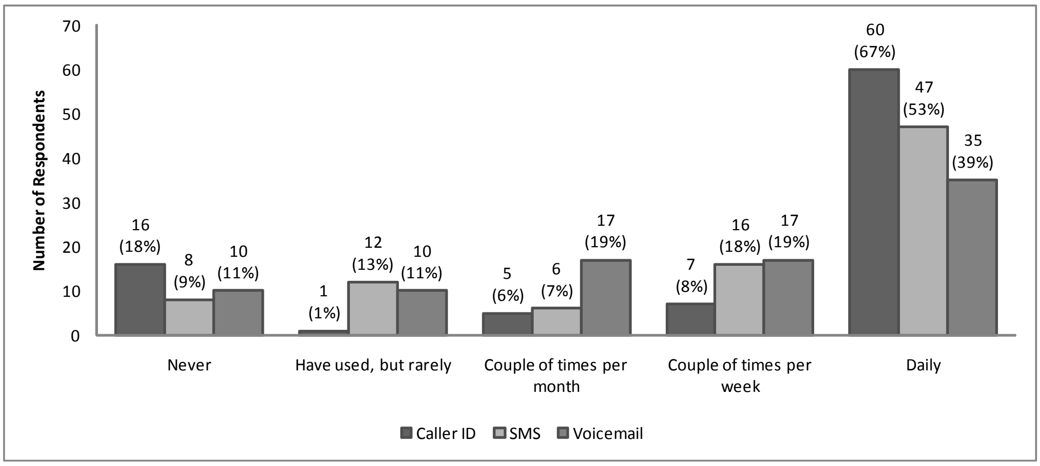

Figure 3.2: Network Service Use

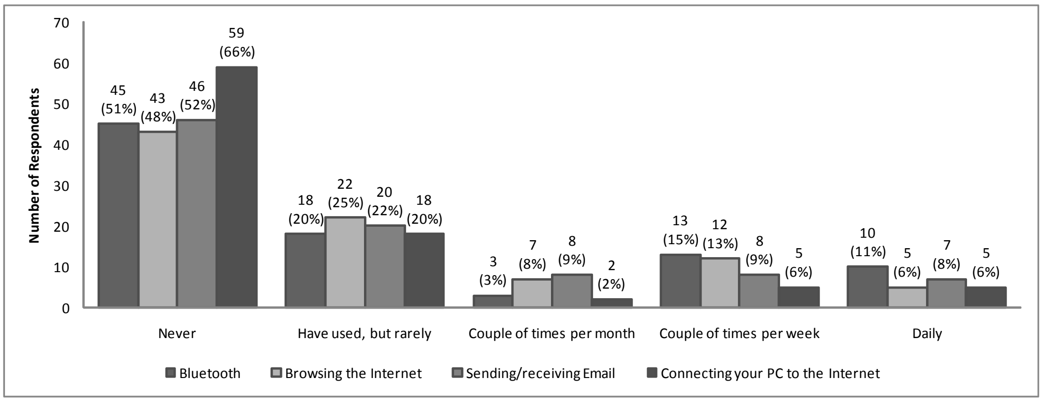

Figure 3.3: Phone Feature Use

service (SMS) and voicemail (Petersen, 2002, pp. 839, 971). More than half of the respondents used the Caller ID feature and SMS on a daily basis. The use of voicemail is slightly less frequent with the majority of respondents (60%) also indicating that they never changed their voicemail message.

Regarding the use of their mobile phone, Figure 3.3 indicates the collected responses. As can be seen the majority of respondents hardly ever used their phone for the proposed functions, mainly using it for communications purposes. Assuming the availability of these features this indicates a sample with limited technical utilization of their phone; alternatively it may point to usability issues with these features which need to be addressed in the future.

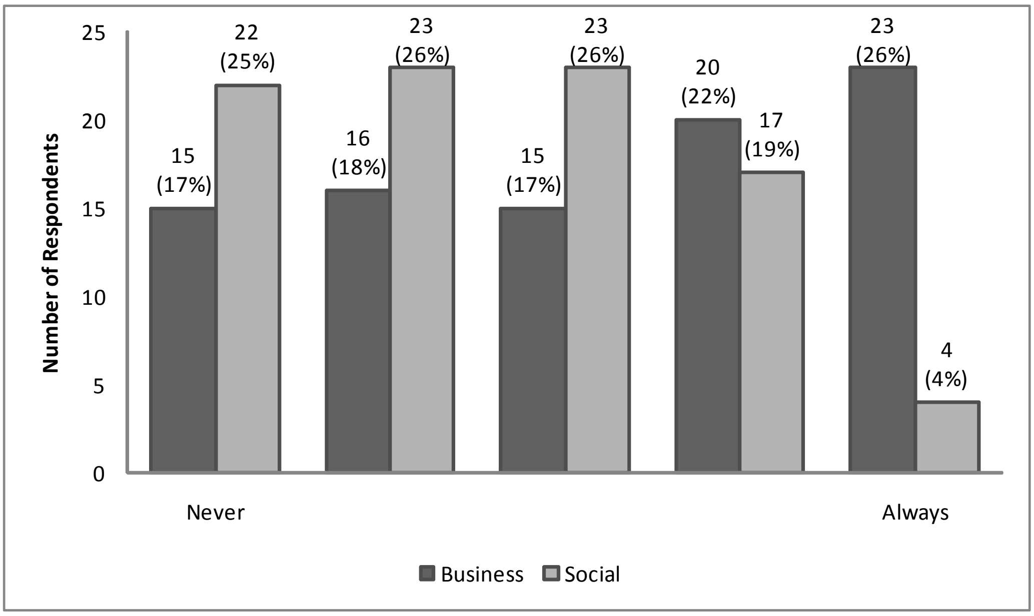

Figure 3.4: Avoiding Disruption in Different Settings

Practice

To establish how users dealt with the disruptive nature of mobile communications the survey asked them to describe their behaviour as well as their use of the Caller ID feature as an information source. Figure 3.4 shows how frequently users switched off their phone(s) to avoid disruptions. As expected most users switched off their phone(s) to avoid disruptions at work. Surprisingly a number of users seemed to be willing to deal with such disruptions to remain available. In contrast, social occasions seemingly did not warrant switching the phone off and only a small percentage of people did so.

When asked about their use of the Caller ID feature the majority of users (80%) indicated that they use the information to filter received calls, while only 18% of respondents are never concerned with who the caller is. Figure 3.5 indicates the results when asking users about their use of the Caller ID feature to manage their identity when making calls. While the majority of users were not concerned with hiding their identity it was interesting to note that quite a few users selectively withheld their identity depending on the recipient of the call.

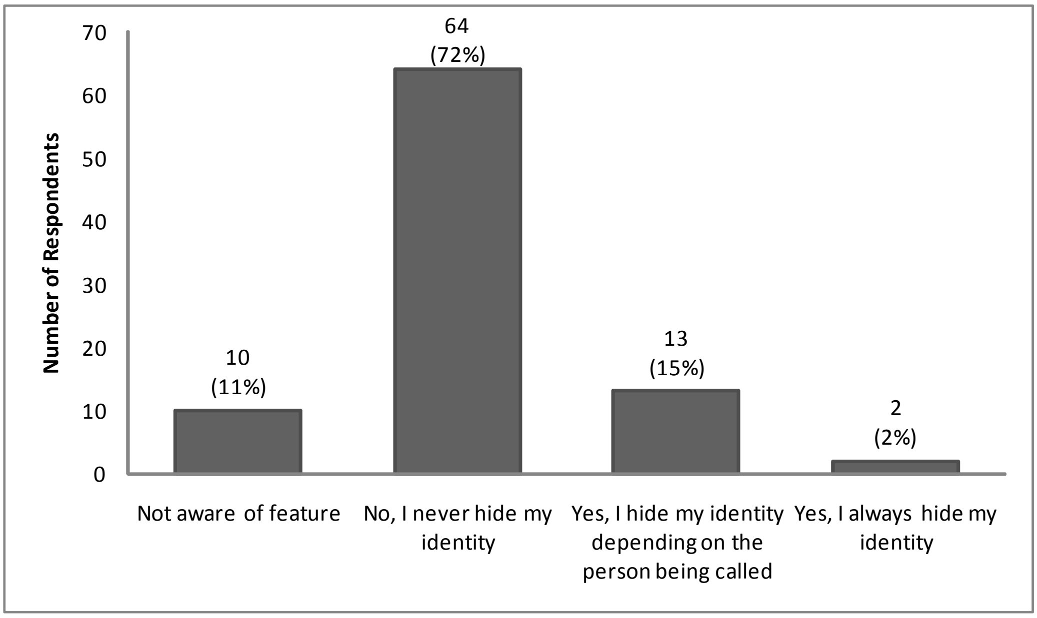

Figure 3.5: Caller ID Use (Making Calls)

Privacy and Trust

A series of questions were used to determine the respondents' concerns about the misuse of their information and their perception of privacy threats and trust in the mobile communications environment. When asked how they feel about the statement, "The use of the Caller ID feature can be seen as a threat to your privacy", 52% of respondents either strongly disagreed or disagreed with the statement while 33% were neutral. Only a small percentage (16%) perceived a threat to their privacy.

Regarding their concern about the misuse of their information respondents answered as indicated in Figure 3.6. As can be seen the majority of respondents were not overly concerned about the misuse of their information by the recipient(s) of their calls. When looking at the network operator the level of concern was only slightly higher. When asked if they ever query the accuracy of the information provided by the network operator, 30% never and 64% sometimes question the information indicating a relatively high level of trust.

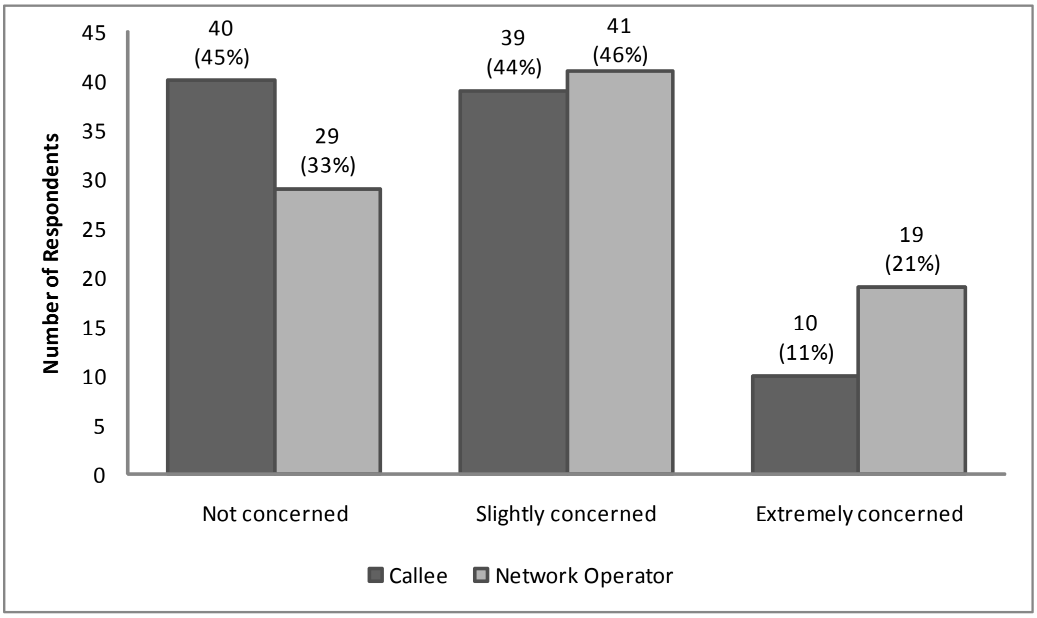

Figure 3.6: Concern about Information Misuse

Willingness to Pay

Respondents' spending patterns with regards to the upgrading of their mobile phone is reported in Figure 3.7. Most respondents upgraded their phone as soon as possible while quite a few also upgraded when needing a feature absent on their current phone.

When asked how often they downloaded applications/multimedia, about half (51%) of the respondents indicated that they never do and a further 42% that they only did so sometimes. It was also found that varying factors, such as costs linked to the time of day, did not have a big influence on the mobile spending habits of users, with 55% indicating that this never influenced their communication decisions and 35% stating that it influenced them sometimes.

Although the statistical significance was not substantial (above the 5% level) Chi-square analysis using the demographic data revealed two relevant relationships. The majority of female respondents (70%) indicated that they did not take cost into account when sending SMS messages or generating data traffic (Chi-square=5.50346; p=0.063822). A difference also exists between those respondents who own a single mobile phone and those who own multiple phones. Owners of a single phone were less likely to download content than those who owned multiple phones (Chi-square=5.17806; p=0.075098). This

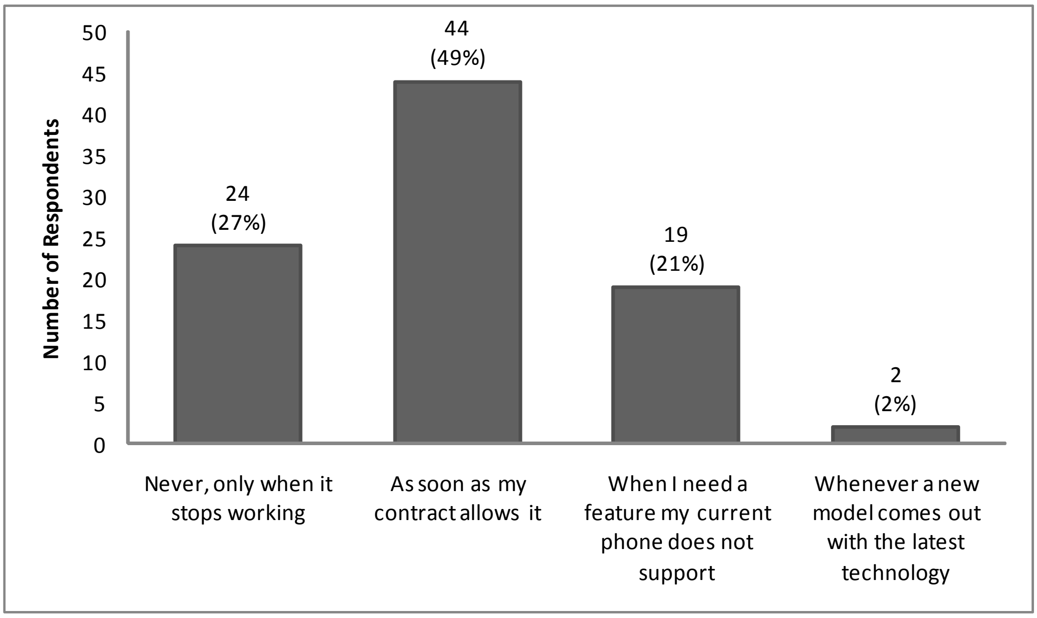

Figure 3.7: Phone Upgrading

could be interpreted as a second phone being used more actively in a job function, or perhaps a greater willingness to experiment without the loss of communications if a fatal phone error occurs.

Perceptions

The respondents' opinion regarding various communications management features were measured. Their perceived usefulness of various items of information was evaluated as well as their willingness to provide the same information. The results for perceived usefulness are given in Figure 3.8. As can be seen most respondents perceived caller identification to be the most important, followed by the subject of the call and then the priority. When asked what other information they would find useful the most common answers were the location of the caller and the estimated talk time.

The willingness of respondents to provide the above information is indicated in Figure 3.9. It can be noted that these figures are similar to the perceived usefulness of each item of information. This supports previous findings that users do perceive the benefit of presence information and are willing to provide it (Milewski and Smith, 2000).

When asked if they would be willing to wait longer for a call to be estab-

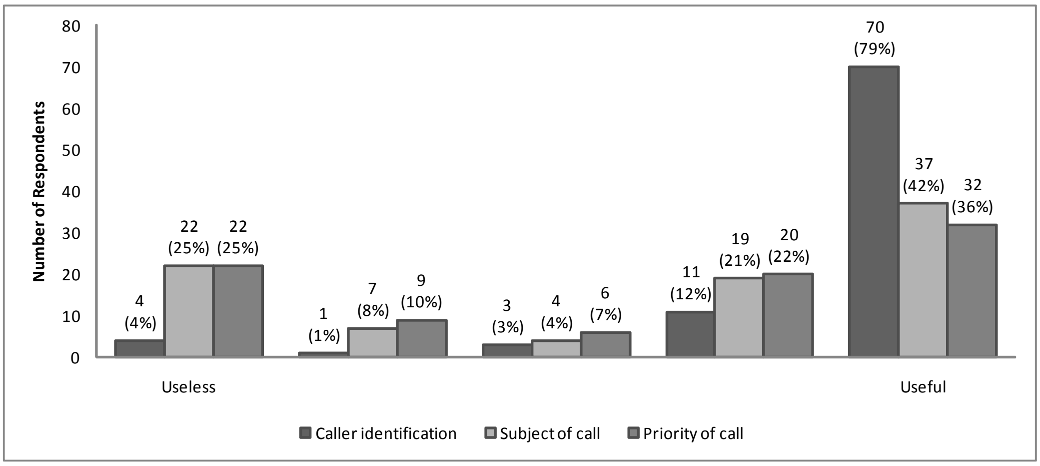

Figure 3.8: Perceived Usefulness

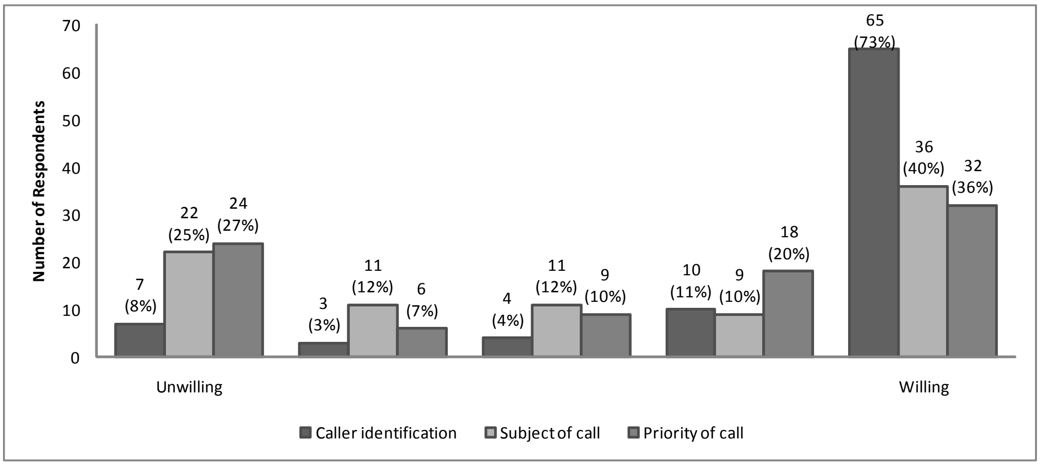

Figure 3.9: Willingness to Provide

lished if the above information were available, 53% of respondents indicated that they would wait up to ten seconds longer while most of the others (42%) were not willing to wait at all.

3.3.2 Cluster Analysis

An analysis of the characteristics data (Questions 5–11) revealed several user groups. The survey questions were standardized to mean 0 and standard deviation 1 so that they have equal influence in determining the clusters. A cluster analysis of Questions 5–7 revealed two groups of people who differ in their availability. Cluster 1 contained the 'not-so-busy' people, in other words those who have no secretary, do not attend that many meetings and are mostly 'Sometimes' unavailable. Cluster 2 contains the 'busy' people, in which half of the people have secretaries, attend mostly more than five meetings a week and are mostly 'Often' unavailable. A Chi-square test using the demographic data revealed that the majority of respondents under 45 years of age (86%) fell in Cluster 1, while 30% of those over 45 years fell in Cluster 2 (Chi-square=3.02181; p=0.082155). This is not surprising as older workers often have more commitments and increased managerial responsibilities.

A cluster analysis of Questions 9–11 revealed three groups of people who differed in terms of technology use. Cluster 1 consisted of people with high technical utilization that on average are the most frequent users of network services – Caller ID, SMS and voicemail – as well as phone features – Bluetooth, browsing the Internet, sending/receiving email and connecting their PC to the Internet. In contrast to this cluster, Cluster 3 consisted of people with the least technical utilization. On all the attributes where Cluster 1 is 'most frequent' the respondents in Cluster 3 are on average the least frequent users. Cluster 2 is very similar to Cluster 3 in that it also consisted of people with low technical utilization, except that on Caller ID and SMS use (two easy features) they were more frequent users than Cluster 3. Note that the clusters did not differ in terms of how often the voicemail message was changed (Question 10).

Chi-square analysis revealed one interesting relationship between the demographic data and the technology-use clusters at the 5% level of significance: membership to the clusters differed between males and females (Chi-square=6.88311; p=0.032019). An inspection of the descriptive data showed that while male respondents were distributed fairly evenly across the technology-use clusters 64% of female respondents fell in Cluster 2.

3.3.3 Cluster Effect Size

To establish whether the relationships between the identified clusters and their answers to the research questions were statistically significant two tests were performed. A Chi-square test only revealed statistical significance between a limited number of relationships. This can be partially attributed to the number of responses that were received. In addition it was decided to calculate an index that measured the size or strength of the relationships, also known as the effect size (w). This was used to assess the practical importance of the relationships. Cohen (1988) provides guidelines for the interpretation of w:

- Small effect: $w = 0.1$
- Medium effect: $w = 0.3$
- Large effect: $w = 0.5$

The effect size of the clusters compared with their responses to the research questions are indicated in Table 3.4.

As can be observed a medium to large effect size existed between the technology-use clusters and many of the research questions, indicating statistical significance. The following section interprets the survey results and draws conclusions relevant to communications management.

3.4 Discussion

The survey results indicate that most respondents do miss important calls because of their current situation and would benefit from some form of communications management. When looking at current practice (see Section 3.3.1) it is interesting that a large number of people do not switch their phone off at work to avoid disruption. This number rises even higher in a social setting despite the invasive influence of unexpected phone calls and the general perception of inappropriateness by the public. The value of being available for communication seems to outweigh all the negatives associated with such practice.

Figure 3.10 and 3.11 show the results when comparing the availability clusters with how often they switched off their phones to avoid disruption. Compared with the overall results the busy people – likely those older than 45 years – more often tend not to switch off their phone in a business environment, indicating a likely need for communications management. As expected the comparison between the two groups was more similar in a social setting (0.09 effect size) as the nature of the situation became more equal for all.

As expected, the Caller ID feature serves as the most important filter mechanism for users. Its use is even higher (94%) when looking at the busy people. However, it is interesting that quite a number of users selectively hide their identity based on the recipient of the call and that the majority of

Table 3.4: Effect Size

Research Question	**Availability Clusters**	**Technology-use Clusters**
Practice		
12 (Business)	0.20	0.24
12 (Social)	0.09	0.23
13	0.17	0.40
14	0.14	0.34
Privacy and trust		
15	0.22	0.30
16	0.06	0.41
17	0.12	0.25
18	0.14	0.38
Willingness to pay		
19	0.25	0.38
20	0.14	0.43
24	0.07	0.27
Perceptions		
21 (Identification)	0.15	0.40
21 (Subject)	0.28	0.37
21 (Priority)	0.18	0.33
22 (Identification)	0.30	0.36
22 (Subject)	0.17	0.45
22 (Priority)	0.15	0.35
23	0.13	0.31

these people come from the not-so-busy group. It is thought that this is an effort to implement basic communications management when the time allows it and serves as a motivation for further research into a solution.

The success of a communications management solution will depend on two main factors – the technical capabilities of users and their mobile phones, and the usability of such a solution. In examining technical use (see Section

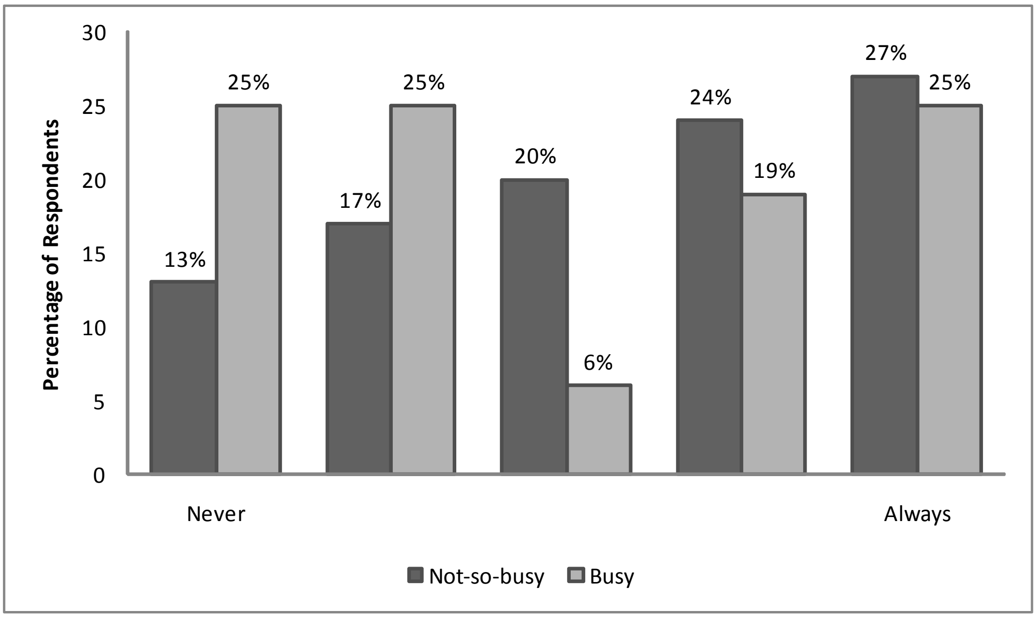

Figure 3.10: Avoiding Disruption in a Business Environment (Availability Clusters)

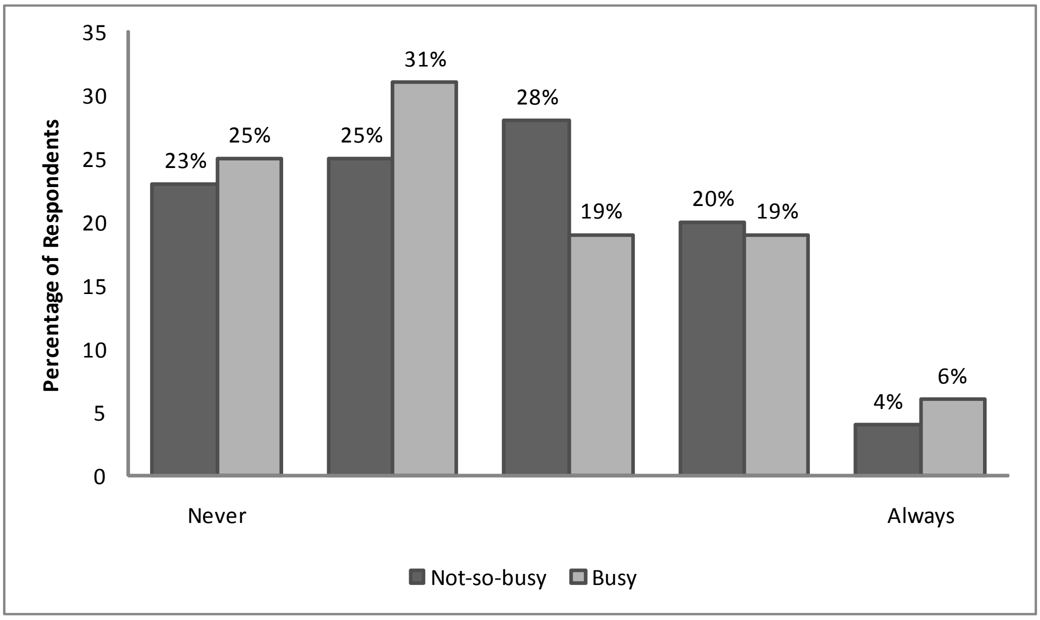

Figure 3.11: Avoiding Disruption in a Social Setting (Availability Clusters)

3.3.1) it is interesting to note that the use of voicemail is generally quite low, perhaps indicating that most people prefer not to leave a message when the recipient is busy. This can become a costly affair and can be successfully prevented with some availability indication. Data communications also do not seem to be a major part of most users' daily activities (or indicates unsupported features) but would form an essential part of a communications management solution, thus indicating a potential issue for further investigation.

In trying to establish the spending habits of users (see Section 3.3.1) it was seen that most users regularly upgrade their mobile phone. This bodes well for a solution requiring the latest technologies. Users', and in particular female subscribers', communications also do not seem to be influenced by varying costs with the ability to connect when the need arises taking precedence.

Looking at the perceived usefulness of various communications management features (see Section 3.3.1) the identity of the caller is by far the most important factor for users. It is interesting to see users' perception of the priority of a call even though it is a very subjective item. The addition of location information and estimated talk time, suggested by respondents, are also worthy additions to this list. This supports previous research projects which have shown positive feedback in using location information (Milewski and Smith, 2000; Bardram and Hansen, 2004; Oulasvirta et al., 2005).

When comparing the technology-use clusters with the perceived usefulness of various communications management features the results are as described in Figures 3.12, 3.13 and 3.14. This relationship exhibits some practical importance with an effect size ranging from 0.33 to 0.40. Compared with the overall response (see Figure 3.8) the people with high technical utilization indicated a higher value across all the features, suggesting that they are perhaps more likely to use communications management in their daily routine. Similar slightly higher figures were encountered in their willingness to provide the above information. These values are consistent with the use of features such as Caller ID by the different technology-use clusters.

There was a universal – across all clusters – willingness to wait longer for a call to be established if communications management features were available. An investment of another ten seconds seems to be most users' compromise for such features. However, this needs to be confirmed by ethnographic field studies as the experience of actual time versus a theoretical value may change user opinion.

In examining the use of mobile communications two interesting factors

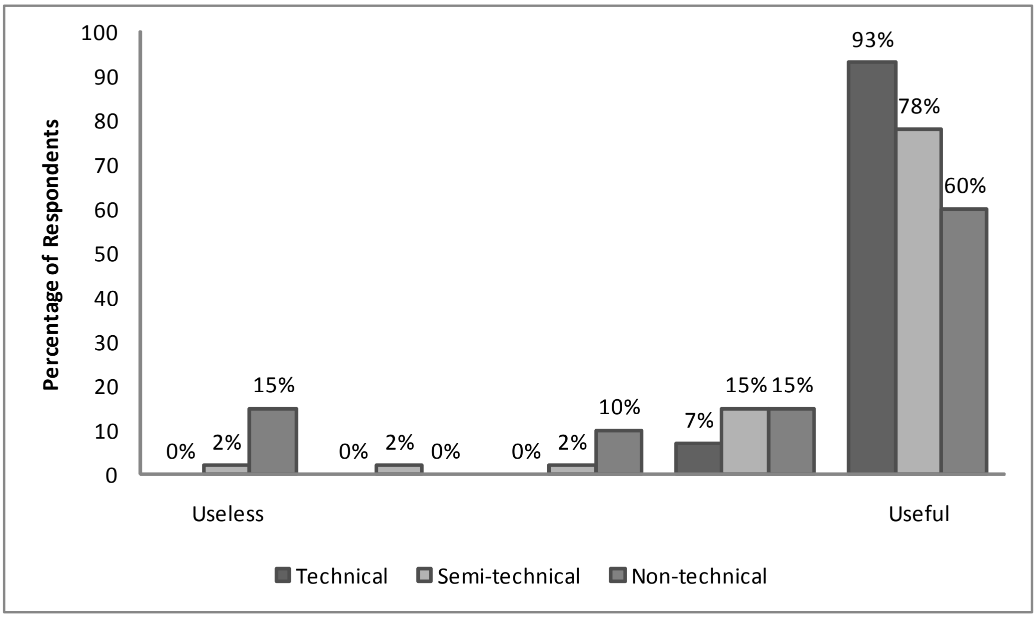

Figure 3.12: Perceived Usefulness of Caller Identification (Technology-use Clusters)

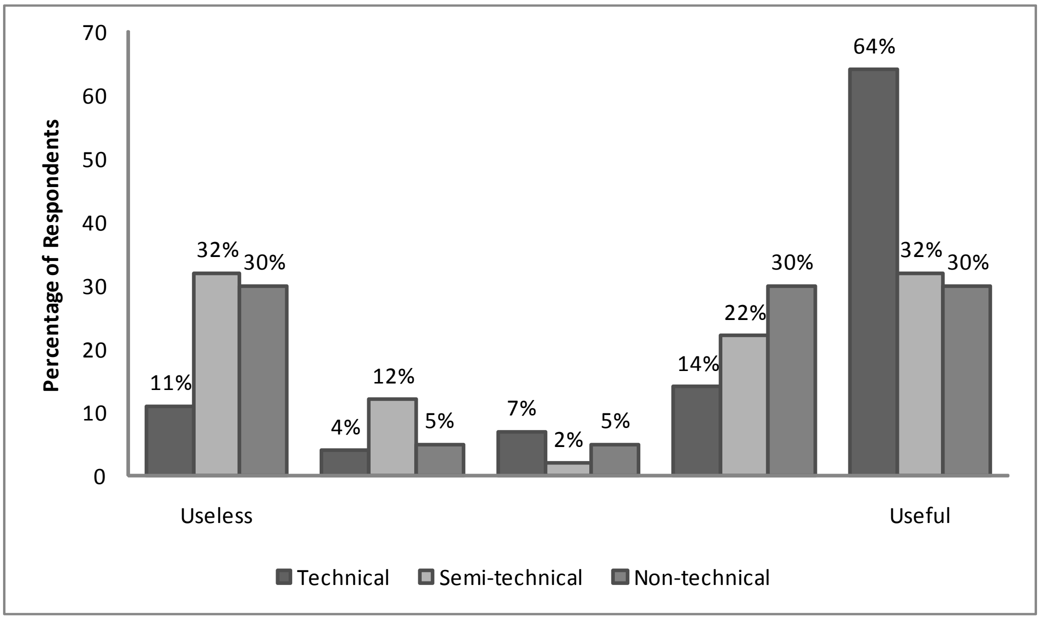

Figure 3.13: Perceived Usefulness of Call Subject (Technology-use Clusters)

can be noticed. First it is noted that the majority of users never change their voicemail message (see Section 3.3.1). Second it is seen that most users

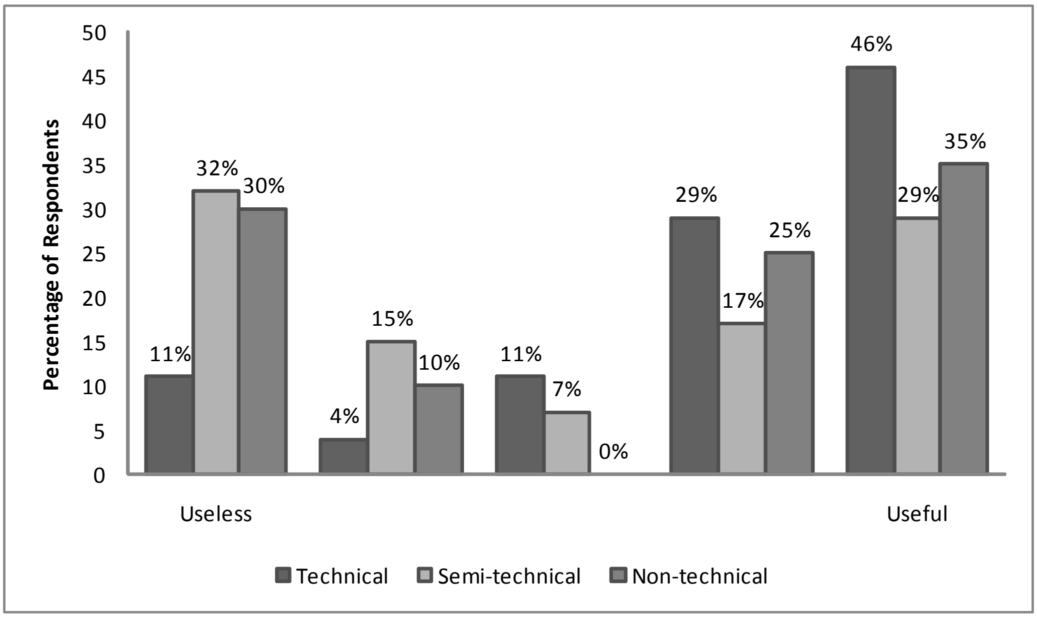

Figure 3.14: Perceived Usefulness of Call Priority (Technology-use Clusters)

do not download applications/multimedia especially when they only own a single phone. These factors emphasize the importance which good usability design will play in the success of a communications management solution. Getting users to download, configure and use a solution will be an important factor in the success of such a system which largely depends on the user base.

Even though we live in an era where identity theft is a major concern most respondents are not overly concerned about the misuse of their information by others (see Section 3.3.1). Similar findings have been made by Karatzouni et al. (2007) during focus group studies. While this may not be a big problem in the mobile world at the moment future developments may change user opinion in this regard. Users are slightly more concerned about the misuse of their information by their network operator but still not excessively so, indicating a fairly high level of trust in this regard.

3.5 Conclusions

This chapter presented the results from a survey exploring user mobile communications perception and practice. This data extends existing knowledge on mobile communications management. Previous research regarding the validity of communications management as a useful communications aid was confirmed by the fact that an overwhelming percentage of respondents were willing to invest both time and money in a solution which could provide such features.

The data collected by this survey also shed some light on other aspects in the area of mobile communications. In particular the identification of target groups, useful additional information and possible caveats are of interest. Since only about 10% of users are never unavailable while expecting an important call it seems that a need exists for additional features to assist users in enhancing the efficiency of their mobile communications. The survey data predicts that a communications management system will appeal to people with high technical utilization and a fairly busy schedule. This seems likely to include an older age group with males being more inclined to use technically oriented features. Although willingness to wait longer for a call to be established will have to be confirmed by ethnographic field studies, the current feedback suggests that not-so-busy people might also be persuaded by additional call features.

The survey has shown that users perceive caller identification, the subject and priority of the call as important features of a solution. In addition, respondents suggested caller location and estimated talk time as valuable information. The survey shows that when implementing a solution as an application the danger exists that users might not download it especially if they only own a single phone. Because of the diverse population using mobile communications usability will always be a critical factor in the success of any solution in this area. Following proper usability engineering guidelines, such as prescribed by Nielsen (1993), is vital in this regard. Future work in this area will undoubtedly benefit by making the user the central figure in the quest to resolve these issues and enhance the efficiency and experience of mobile communications.

Based on these results three issues need to be considered. First, how can the context information be obtained in a manner which is unobtrusive and does not inconvenience the user? Second, how should context be integrated into the phone and made part of the user experience to provide value to the user in managing communications? Third, what is the most effective way to distribute context to a large number of users? Fortunately a potential solution exists in the form of presence. The next chapter will examine this technology and its usefulness to convey context information. In addition potential privacy issues are also considered.

Chapter 4

Presence and Privacy

Presence is an important technology in communication applications. Millions of people use presence every day as part of instant messaging (IM), which allows real-time messages to be exchanged between people. Although presence and IM are usually perceived as a single technology they are in fact distinctly different.

While IM enables the exchange of messages, presence indicates when a person is online and available for communication. Thus presence can also be of value in other communication applications such as mobile communications. Some of the projects in Chapter 2 have used presence to communicate context between users. Thus presence is a candidate technology for implementing a mobile communication management solution.

This chapter aims to provide a better understanding of presence. The term is interpreted in various ways and the chapter starts by providing a standard definition as it is understood here. Thereafter the evolution of the technology is discussed by looking at presence from a historical perspective. Next the chapter examines the architecture and various aspects which form part of a presence system. Finally the important issue of privacy and how best to address it is debated.

4.1 A Definition of Presence

The term *presence* can be interpreted in different ways depending on the context it is used in. In the context of this research presence refers to whether a user can be contacted right now. Knowing about presence is useful because it saves communication time.

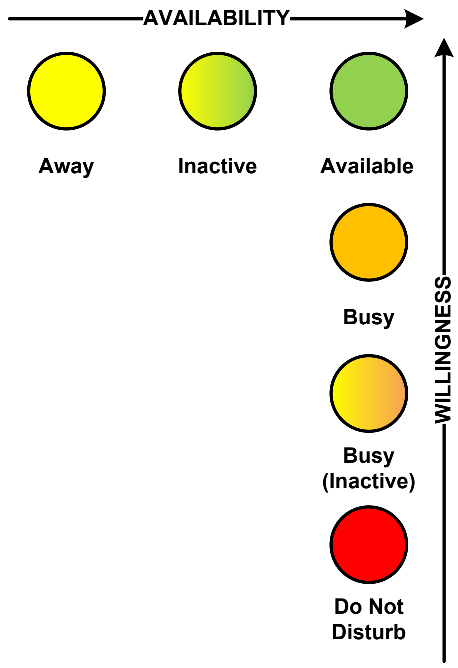

Figure 4.1: Presence Dimensions (Microsoft TechNet, 2010)

Presence can be as simple as knowing whether a user might be available by calling a particular phone number. However, many more complex situations exist where presence can convey the location of a user or the ability of a user to take a call, amongst others. Because presence can convey a large number of attributes it can be useful in many different situations.

At a fundamental level presence conveys a user's availability and willingness to communicate. The difference between availability and willingness is illustrated in Figure 4.1. When a user is both available and willing to communicate this can be indicated by a presence state of 'available'. It is reasonable to expect a user who is available to respond to communication requests. However, other combinations of availability and willingness does not guarantee that instant communication is possible. For example, a user may be available (online) but not willing to communicate which is indicated by a self-defined presence state of 'do not disturb' (Microsoft TechNet, 2010).

Presence is an accepted and successful technology because it is built on two important principles:

- The sharing of presence with another user is voluntary. Usually the decision to share presence is based on a relationship with another user and a trust decision. This separates a user who has access to your presence from everyone else on the network (Saint-Andre et al., 2009, p. 31).

- Presence access is usually bidirectional. When you allow other users access to your presence you also receive access to their presence (Saint-Andre et al., 2009, p. 32).

As different presence standards exist various definitions of presence have been formulated. According to Saint-Andre et al. (2009, p. 31) presence allows you to "know when a contact of yours is online and available for communication." At a slightly more technical level Rosenberg (2006) states that "presence conveys the ability and willingness of a user to communicate across a set of devices." According to the Open Mobile Alliance (2008) presence information can be a dynamic set of status, reachability, willingness and capability information pertaining to a user.

In summary, presence indicates whether a user can be contacted and what that user's context is. Before looking at how presence works it is useful to understand how the technology evolved. This knowledge also helps to distinguish between the implementation difference of presence standards in the rest of this research. The next section discusses the history of presence.

4.2 Presence History

Presence has its roots in communication applications which have always been popular amongst users. Early examples include:

1. Bulletin board systems which allow users to log on to a server and exchange messages.

2. These were followed by distributed Internet-based chat systems of which some are still in use today (Salkintzis and Passas, 2005, p. 320).

3. The first IM service was launched by the Mirabilis company in 1996 and is called ICQ (I seek you), whereafter many further implementations followed. These early IM services were closed systems which made it impossible for users to communicate across networks.

Dissatisfied with the inconvenience of closed IM systems, Jeremie Miller released an open-source IM implementation in 1999 called Jabber. Jabber immediately gained support from the community which led to the formation of the Jabber Software Foundation (JSF) in 2001. The JSF's primary goal was to oversee Jabber and coordinate the growing number of open-source and commercial projects using the technology.

Starting in 1998 a similar initiative was set in motion by the Internet Engineering Task Force (IETF) to create standards for IM and presence. This group, called the Instant Messaging and Presence Protocol (IMPP) working group, defined semantics for common presence services but was unable to deliver a unified protocol.

After the dissolution of the IMPP working group the IETF continued its work on presence through the Session Initiation Protocol (SIP) working group. Under its guidance a working group called the Session Initiation Protocol for Instant Messaging and Presence Leveraging Extensions (SIMPLE) was formed in 2000, which developed an interoperable standard for instant messaging and presence. The standard is compliant with the requirements developed by the IMPP working group, which defined the basic model of presence and IM. SIMPLE is an IP-based protocol which utilizes SIP. This also makes the protocol especially suitable for future converged networks such as the IP Multimedia Subsystem.

In 2002 the JSF also decided to seek formalization of the base Jabber protocols under the guidance of the IETF. As part of this process the Extensible Messaging and Presence Protocol (XMPP) working group was formed whose task it was to standardize the Jabber protocol in line with IETF requirements. This process was completed in 2004 when XMPP was accepted as an IETF technology in conformance with the requirements originally set out by the IMPP working group. Further extensions to the base XMPP specification continues under the supervision of the JSF, which in 2007 renamed itself to the XMPP Standards Foundation (XSF).

Parallel to these efforts to standardize presence on the Internet, wireless handset manufacturers pursued similar aims in the mobile communications domain. In 2001 Ericsson, Motorola, and Nokia formed an initiative known as the Wireless Village (WV). The purpose of the initiative was to create innovative mobile instant messaging and presence services. The WV recognized the need for an industry standard which would be interoperable across mobile networks and the Internet. They identified presence as the "key enabling technology" for the initiative (The Wireless Village, 2001). As a result of this initiative a specification for a standard presence protocol was produced, called the Instant Messaging and Presence Service (IMPS).

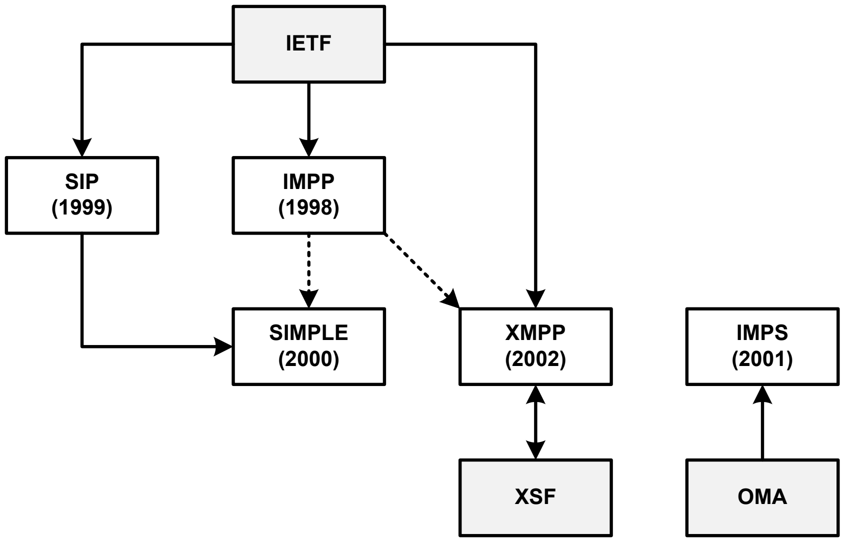

Figure 4.2: Presence Standards and Communities

In 2002 the WV was consolidated into the Open Mobile Alliance (OMA). The OMA integrated a number of existing forums to become the leading industry provider of interoperable mobile data services. It consists of companies representing almost all industry segments, such as mobile operators, device and network suppliers, information technology companies, and content and service providers. The OMA collaborates closely with other organizations to ensure compatible specifications. Development of IMPS has continued under the guidance of the OMA with the latest version of the specification being published in 2007.

Thus three main international standards for IM and presence exist: IMPS, SIMPLE, and XMPP. Figure 4.2 summarizes the organizations involved in the development of these standards.

Each standard adds its own requirements on top of a basic presence architecture and feature set. The next section discusses the common functionality which can be expected from any presence system and compares the completeness of the standards against each other.

4.3 How Presence Works

Presence is a complex technology in which multiple servers, roles and data items are defined. This section will discuss the common functionality found in today's presence standards.

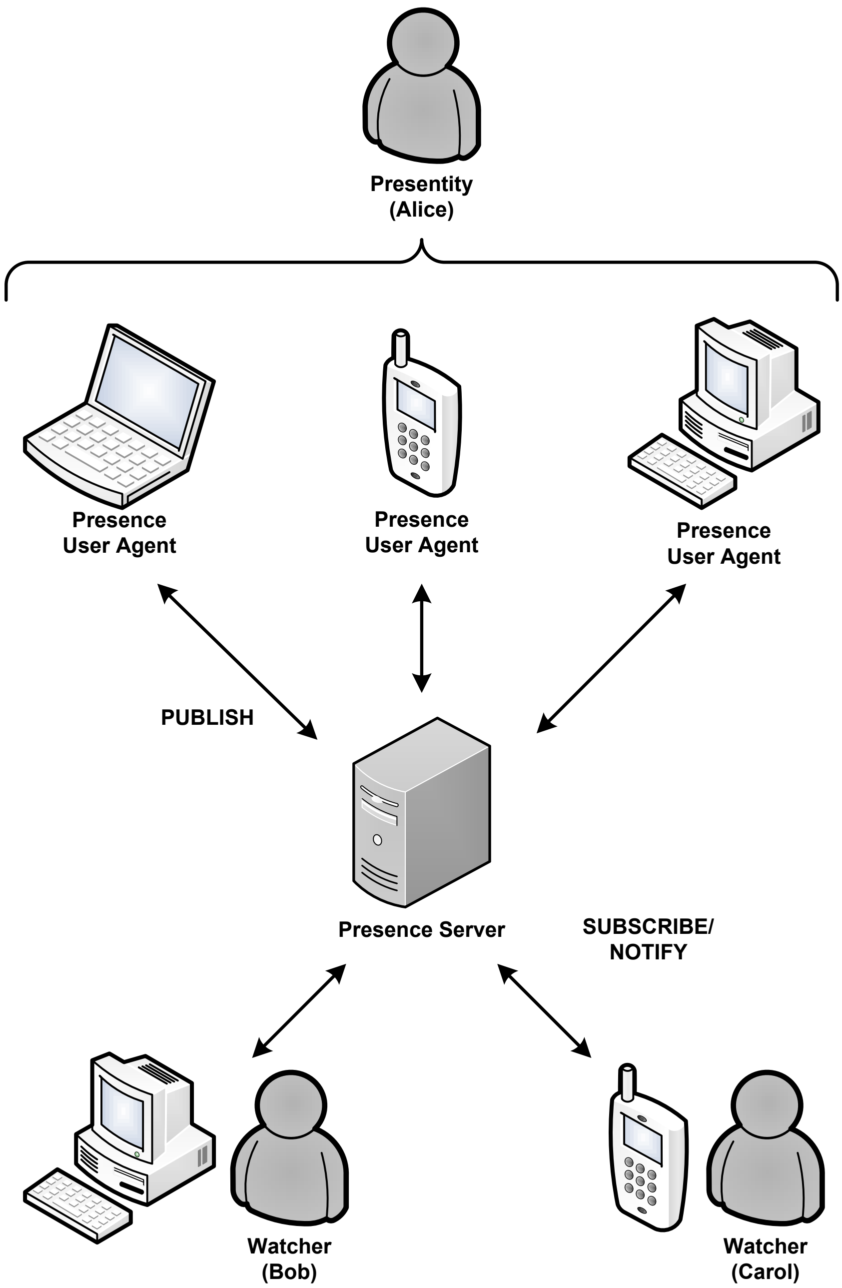

Figure 4.3: Basic Presence Functionality

4.3.1 Basic Presence Functionality

A presence system creates an environment where users can be informed of each other's availability and willingness to communicate. To achieve this several roles are defined in the context of presence. These are illustrated in Figure 4.3.

A user who provides presence information to the presence server is called a presence entity, or more succinctly, a *presentity*. Presentities are usually people but could also be non-human, such as a radio station publishing the title of the song currently playing (Open Mobile Alliance, 2007b). In the figure Alice fulfills the role of presentity.

A presentity can have a number of presence user agents (PUAs), such as a desktop computer, smartphone or notebook. Through its interaction with the user each PUA supplies presence information to the presence server. For example, the smartphone knows whether Alice is currently engaged in a call, while the laptop and desktop computer knows whether Alice is logged in, her next calendar appointment, etc.

A *watcher* is an entity interested in the presence of a presentity. Watchers are usually people but can also be non-human, e.g. a network server or corporate calendar. In the figure Bob and Carol fulfill the role of watchers. A watcher can request presence information in several ways.

- A watcher who is only interested in the current presence information of a presentity is called a *fetcher*. A special kind of fetcher is one that fetches information on a regular basis. This is called a *poller*.

- A watcher can also subscribe to a presentity's presence information. In this case the watcher requests to be notified of future presence updates of a presentity. Such a watcher is called a *subscriber* and will be kept up to date about the presentity's presence information by notifications.

As illustrated in Figure 4.3, presentity Alice has three PUAs. A client running on any of these PUAs publishes Alice's presence to the presence server. Meanwhile, watchers Bob and Carol want to subscribe to Alice's presence information.

Bob and Carol each run a client that sends the presence server a request for Alice's presence. The presence server consults Alice's presence settings in order to determine if Bob and Carol are permitted to subscribe to her presence. If they are, then the presence server sends a message containing information about Alice's current presence state. Whenever Alice's presence state changes the presence server sends a message to Bob and Carol's clients informing them of the change.

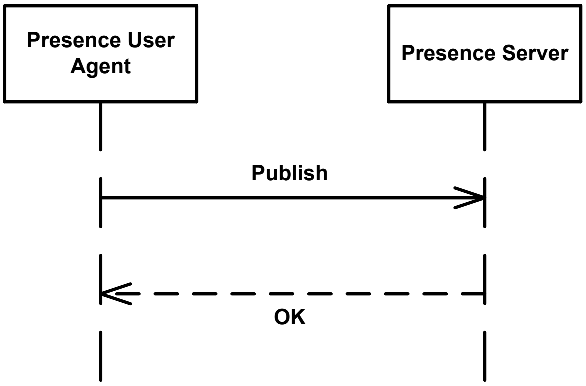

Figure 4.4: Publication of Presence Information

4.3.2 Publish/Subscribe/Notify Architecture

Presence uses three primary types of messages. These messages ensure that a presentity's current context is captured as presence information and that watchers can successfully request and receive such information. In general this is referred to as a publish/subscribe architecture.

A PUA publishes presence information to a presence server whenever a change in the presentity's context occurs. The receipt of the information is acknowledged by the presence server. This simple exchange is illustrated by Figure 4.4.

The distribution of presence information to watchers is handled by the presence server on behalf of the presentity. A watcher who wishes to subscribe to changes in a presentity's presence information sends a subscribe request to the presence server. The server acknowledges receipt of the request and determines whether the watcher should be allowed access to the presentity's presence information. Usually this is done by prompting the presentity to approve the request. If the request is accepted the presence server notifies the watcher of the presentity's current presence information. Subsequently, each time the presentity's presence information changes the presence server will notify the watcher of the new presence state. This process is shown in Figure 4.5.

4.3.3 Presence Life Cycle

The role of the presence server is much more comprehensive than the previous sections may have alluded to. It performs various valuable functions as

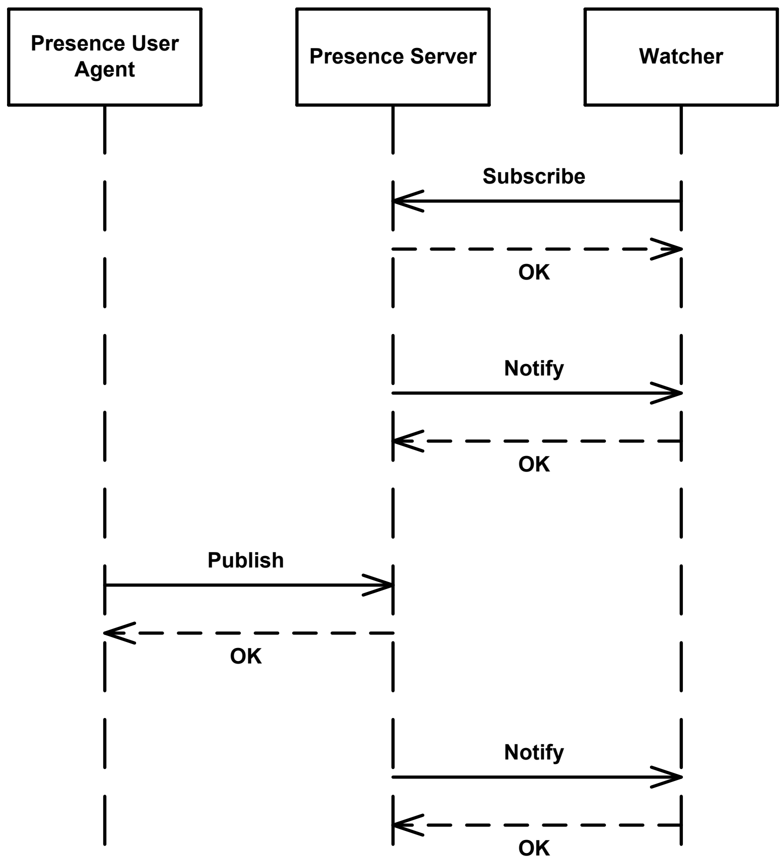

Figure 4.5: Subscription and Notification of Presence Information

information flows from a presentity to a watcher. This presence life cycle is illustrated by Figure 4.6

A presence server performs several functions which handles changes in a presentity's presence information:

1. It allows a presentity to upload a privacy policy document which specifies the information to which each watcher is entitled to. This allows a presentity to notify only specific watchers of any changes in presence information.

2. As PUAs publish new presence information the presence server merges the data into a unified document. This is done using a composition policy that specifies the rules regarding merging presence information. For instance, Alice may answer a call on her phone, leaving her laptop or desktop computer idle. A change in presence may also be brought about without the involvement of the presentity, for example losing cellular connectivity.

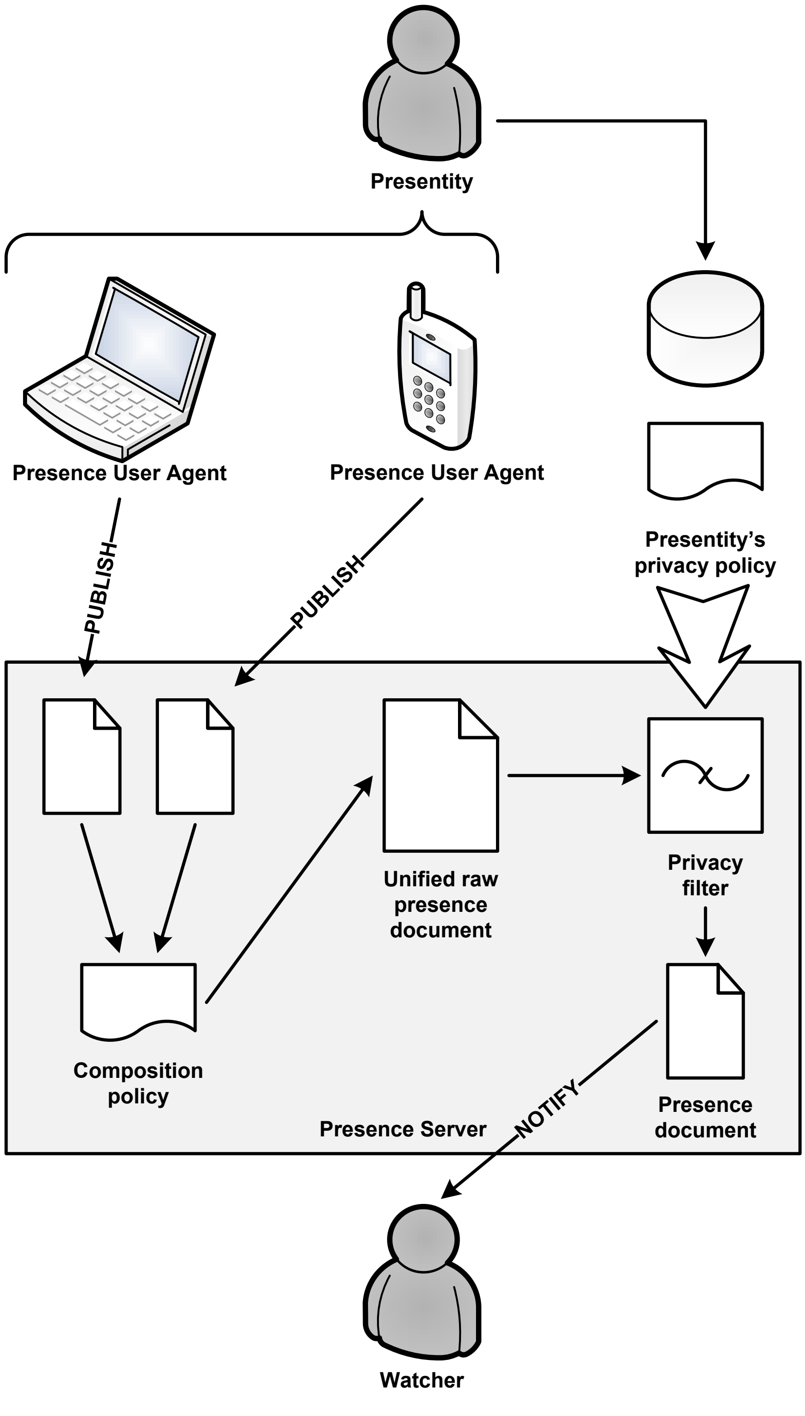

Figure 4.6: The Role of the Presence Server

3. The unified document is filtered using the privacy policy of the presentity. This filtering process removes any information that the presentity does not want to provide to a given watcher. At this stage it is possible for several parallel presence documents to be created, each pertaining to a specific watcher.

4. The presence server sends the watcher a notify request containing the presence information.

4.3.4 Other Features

Several further features and enhancements are used to make presence more valuable and efficient. This section provides an overview and compares the major presence standards across these features.

A presentity's contact list, also known as a roster or buddy list, forms an essential part of presence. The contact list is stored on the presence server which allows a presentity to connect from anywhere and still retrieve the contact list. As a presentity connects to the presence server the current state of each contact in the presentity's contact list is retrieved.

Presentity's typically want to know which watchers are subscribed to their presence information. This information is also necessary for authorizing pending watcher subscriptions. Keeping track of watcher presence subscriptions is supported across all standards through the use of a watcher list. In XMPP this information is part of the buddy list, which is referred to as a roster (Saint-Andre, 2004). In general a buddy list is used to represent the contacts to which a person has an active presence subscription.

Watcher notifications allow a presentity to be informed about subscription requests to his/her presence information. While IMPS offers the functionality for retrieving watcher information it does not offer real-time notifications about updates to this list (Salin, 2004).

While all standards support reactive presence authorization, the IMPS also offers proactive authorization (Open Mobile Alliance, 2005, 2007a). This allows a user to authorize access to presence information before anyone has requested it.

Rich presence is a common extension to the basic information available through presence. Rich presence includes more detailed information about a presentity, such as what users are doing, what music they are listening to or which videos they have watched, their activity, mood, location, etc. The

IMPS standard supports slightly fewer options in this regard, while XMPP extends the range of attributes through extension specifications.

Presence can become a very bandwidth expensive service when detailed presence is sent to a large group of watchers. All the standards allow a watcher to control the amount of presence notifications being received from a presentity. This is useful to reduce bandwidth consumption or when a watcher is temporarily not interested in this information.

Presence information can be sensitive and of a private nature. A mechanism needs to be in place which allows the presentity to selectively distribute presence information to watchers. This encompasses a list of authorized watchers and rules for transforming presence documents to filter out sensitive information. The presentity should also be able to manage these resources as needed. A privacy policy allows the presentity to define advanced rules regarding the distribution of presence information. XMPP offers basic presence notification blocking (inbound and outbound) using privacy lists while IMPS offers presence information filtering through buddy lists (Saint-Andre, 2004; Open Mobile Alliance, 2005). These mechanisms allow a presentity to remove certain pieces of information contained in a presence notification or to block notifications entirely. While SIMPLE also provides these basic features it additionally provides presence notification filtering based on the state of the presentity or a validity period (Schulzrinne et al., 2007).

It is not easy to determine the differences between the major presence standards at a quick glance. To try and make the differences more clear Table 4.1 compares the IMPS, SIMPLE, and XMPP standards across the previously mentioned features.

The management of privacy is still one of the main areas where presence standards are evolving. Presence can expose a lot of information about a user, including rich presence. This can be highly sensitive because of its personal nature and a user must be able to select which information is revealed about him/her. Additionally, a user should be able to inspect the information being published, particularly if it is generated automatically.

However, providing privacy through the use of privacy policies raises several concerns. Client capabilities need to be standardized to provide consistency in the upload, modification and support of privacy policies. This is needed so that users can be certain about what privacy controls are in place. Furthermore, the creation of specific privacy rules can be a complex and laborious process, discouraging the effort needed for proper privacy protection.

Table 4.1: Comparing Presence Standards

Feature	**IMPS**	**XMPP**	**SIMPLE**
Publish requests	x	x	x
Subscribe requests	x	x	x
Notify requests	x	x	x
Contact list	x	x	x
Watcher list	x	x	x
Watcher notifications		x	x
Presence authorization	x	x	x
Rich presence	x	x	x
Watcher filter	x	x	x
Advanced privacy policies			x

Most importantly, it is acknowledged that the results of combining several rules can be "non-obvious" to end users, thus reducing their effectiveness (Rosenberg, 2007b).

From a user's perspective it would be beneficial if a way could be found to model privacy policies according to their social perspectives and daily routine. The next section examines privacy, and whether social relationships provide more precise conditions for revealing presence information.

4.4 Informational Privacy

Privacy is an important topic in current society. In our electronic world, with growing networks of data, there is increasing concern about the privacy of personal information. This issue is a focal point for both the research community and the public at large. As we become more dependent on an online, networked world, privacy will most likely grow in importance.

Today the topic of privacy is most commonly associated with the protection of personal information. However, the notion of privacy has evolved over time and within various aspects of everyday life. Tavani (2008, pp. 135–141) discusses four distinct kinds of privacy: Physical or accessibility privacy is one of the earliest forms and protects the individual from unwarranted physical intrusion. Decisional privacy relates to freedom from interference in personal choices, plans, and decisions. Psychological or mental privacy expresses the protection of intimate thoughts and the integrity of one's personality. Lastly, informational privacy is defined as having control over or limiting access to personal information. Because we are considering presence technology and information we are therefore dealing with informational privacy issues.

In the digital age informational privacy is at the forefront of public concern. Tavani (2008, pp. 139–140) argues that the use of computers and information technology impacts informational privacy in four ways: The amount of information that can be collected and stored is virtually unlimited. The speed at which information can be exchanged is much faster than before. Information can be retained indefinitely. Finally, and most important, the kind of information that can be collected is much more detailed and reveals more about a person than ever before. Moor (1997) summarizes the situation well when he says that "the problem of computer privacy is to keep proper vigilance on where such information can and should go".

Using presence exacerbates the impact of computers and information technology when looking at the above factors. In particular, the quality of available presence raises privacy concerns. Also, a situation which was naturally private now needs normative protection from information access by others. Natural privacy refers to situations where people are protected from intrusion or observation by natural or physical circumstances; normative privacy exists when a situation is protected by ethical, legal, or conventional norms (Moor, 1997). For example, whereas presence was previously only known to co-workers in the same office, presence now distributes the same information to a much wider audience in any number of locations, thus requiring normative privacy protection measures. However, before examining the means by which such concerns are addressed the next section first provides a definition of privacy.

4.4.1 A Definition of Privacy

Privacy is a dynamic concept that is not easily pinned down. A reason for this is that it continuously evolves, influenced by our current social environment (Tavani, 2008, p. 132). Furthermore, definitions of privacy are shaped by whether it is seen as a right or an interest, as a concept which can stand alone, or something which is derived from more basic concepts (Tavani, 2008, p. 133). In an attempt to provide a concise definition this research follows the advice of Tavani and Moor (2001), who believe that a good theory of privacy must account for the justification, concept, and management of privacy.

Two common justifications of privacy is that it has instrumental and intrinsic value. Instrumental values serve as a means to achieve something else which is desirable. Intrinsic values are valued as something good in themselves. However, Moor (1997) maintains that privacy entails even more than these values. He argues that certain "core values" are shared across all human cultures and are needed for survival. While privacy is not a core value as such, it is an expression of a core value, namely security. In large, interactive societies privacy expresses the need for security, which allows a culture to survive and prosper. Linking back to an instrumental/intrinsic view, privacy is instrumental in supporting all the core values as well as intrinsically valuable as an expression of security (Moor, 1997). A benefit of such argumentation is that it also provides a useful distinction between privacy and security.

When considering informational privacy, two concepts or theories dominate current thinking: restricted access and control. Most informational privacy analyses use variations of these two theories (Tavani, 2008, p. 141). According to the restricted access theory one has privacy when others can be limited from accessing your personal information (Tavani, 2008, pp. 141–142). In opposition, the control theory advocates that we have privacy if we control information about ourselves (Tavani and Moor, 2001). However, both these theories have been criticized and fall short on various points. In an attempt to synthesize the important elements of each theory Moor (1997) proposed a framework called the "control/restricted access theory", later renamed to the Restricted Access/Limited Control (RALC) theory of privacy (Tavani, 2008, pp. 144–146). RALC emphasizes the importance of situations or zones, also found in the restricted access theory, which protect privacy by specifying different levels of access for different individuals. This allows an individual to make the final decision about what information to protect and how much privacy they desire. Thus the notion of privacy applies to a zone and not the information in itself; "...to protect ourselves we need to make sure the right people and only the right people have access to relevant information at the right time" (Moor, 1997). RALC also recognizes the importance of control in managing one's privacy. According to Tavani and Moor (2001) this is expressed in three ways: People can choose the situations that offer acceptable privacy to them. Consent can be given to others to access information, thus waiving the right to privacy. Finally, individuals should be able to access their information and correct it if necessary. Thus, in the RALC theory privacy is not dependent on absolute control over information; rather, limited controls allow people to manage their privacy effectively (Tavani, 2008, p. 145).

The view that privacy is linked to the situation in which information is transferred, and not the information itself, is supported by Nissenbaum (2004). She refers to her model as the theory of “contextual integrity”; it is a normative model that provides a philosophical description of privacy in terms of personal information transfer (Barth et al., 2006). This extends her previous work on privacy in public, which examines the effect of information technology on public information and advocates similar protection for all types of information (Nissenbaum, 1997, 1998). Contextual integrity is built on the perspective that everything in life takes place within a context. People act in a certain capacity or role within a context to achieve relevant ends (Barth et al., 2006). Such contexts can be defined at various levels of detail and are partly constituted by norms which determine aspects such as roles, expectations, behaviours, and limits (Nissenbaum, 2004). According to Barth et al. (2006) informational norms, norms applying to the communication of information, are defined by two key principles: appropriateness and transmission. Information can not be classified simply as public or private; instead the type of information is defined by the appropriateness of that information in a given context. Transmission principles control the flow of information and includes, for example, confidentiality, reciprocity, and consent. According to the theory of contextual integrity these principles need to be respected in order for an individual to have privacy.

According to the RALC theory, control is important in relation to the management of privacy. Such control needs to be adaptable to accommodate changing contexts. Palen and Dourish (2003) proposes that in actual fact one is managing boundaries “between different spheres [contexts] of action and degrees of disclosure within those spheres. Boundaries move dynamically as the context changes.” They propose three boundaries: A disclosure boundary for selective disclosure of personal information. An identity boundary in which we distinguish different views of others at different times and treat them accordingly. And lastly, a temporal boundary in which our response to a situation draws from past experience. Dourish and Anderson (2006) also discuss the trade-off between risk and reward which affects our decision to share personal information. For example, users may control the degree of location accuracy in a presence system, such as room, building, street, or city level, thus making a trade-off between privacy and specificity (Dourish and Anderson, 2006). Naturally, such assessments are highly subjective; however, they are extremely relevant in the context of presence technology.

By defining privacy in the above manner this research has emphasized the importance of context, also referred to as a zone or situation, in the assessment of privacy. Whether we feel comfortable with the privacy of presence information will depend largely on our current context as well as the relationship we have with the watcher. The next section examines the role of social relationships in more detail.

4.4.2 The Role of Social Relationships

Presence information expresses our willingness and availability for communication. However, through regular communication, gaining knowledge of a person's lifestyle, and building a relationship, we can also predict their interruptability to a certain degree. How do our relationships affect the personal information we share with one another? There is a substantial amount of theory and research concerning the topic of interpersonal communication. Within this body of work the discussion dealing with relationships are of special interest.

It is generally accepted that relationships are not innate, but are formed and develop gradually over time as exchanges between people take place (Roloff, 1981, pp. 61–62). These social relationships are greatly influenced by the time or the need we have to develop a relationship (Trenholm and Jensen, 1996, p. 352). Consequently, many unique relationships develop between people.

Hartley (1993, p. 177) observes that in everyday life we often recognize complex, individual relationships arranged into groups. He proposes, for example, three such groups: family, friends, and co-workers. Such groups are easily expressible classifications of the type of relationship we have with a specific person. For example, instead of saying that one has a confiding, respectful relationship with someone, we would abstract it and simply call each other friends.

One of the major factors contributing to the growth of a relationship is the communication of information about oneself to another person, referred to as self-disclosure (DeVito, 1992, pp. 114–118). According to Weaver II (1993, pp. 159, 494) such an act involves risk, but is necessary for a strong and enduring relationship; if the act is reciprocated the relationship grows and mutual trust is built between the individuals involved. The amount of information being shared is dependent on the state of the relationship as well as the sensitivity of the information (DeVito, 1992, p. 368).

Social relationships move through various stages characterized by changing levels of communication and self-disclosure. DeVito (1992, p. 428) explains that these stages range from initial contact and involvement to intimacy, possibly followed by deterioration and, finally, dissolution. While trust increases when a relationship is growing, it is an almost universal truth that a deterioration of a relationship leads to a marked decrease in trust (DeVito, 1992, p. 426).

Thus social relationships play an important role in our perception of privacy. It is likely that we would share more presence information with someone we are familiar with or who reciprocates the act. When combining context and social relationships this research makes two observations: In a particular context the type of presence information one shares will differ per watcher according to their relationship with us. Such a relationship may be expressed on an individual or group basis. In addition, the shared presence will also change over time and context as the relationship develops. While a person may presently feel comfortable sharing certain presence with a watcher in a particular context, this may change in the future. As Dourish and Anderson (2006) note, "privacy is not simply a way that information is managed but how social relations are managed."

4.5 Conclusions

Untimely interruptions can often be attributed to a lack of knowledge regarding the current situation of a user. Presence can provide valuable cues in this regard. Presence is analogous to the dial tone on a phone line which indicates whether a person is available for communication. However, in reality the dial tone only indicates network connectivity and no information about the person on the end of the line. Presence provides rich context information about users in addition to their network connectivity state.

Presence can make communication and collaboration more efficient and is rapidly spreading, from its traditional foundations in IM, to other applications. This chapter has provided an overview of the presence architecture and the processing of presence information. However, such information can be highly sensitive and private, and must be distributed in a secure manner. This chapter has discussed privacy in terms of social relationships and has argued for such an approach as a condition for revealing presence information. This addresses the privacy issues in revealing presence.

This chapter concludes the investigation into background information. In the next part the research continues by defining a model for the management of mobile communications using presence in a privacy aware manner.

Part I of the thesis examined information relevant to this research. A general introduction (refer to Chapter 1) formed the basis for the following chapters. It presented the problem of managing mobile communications and the possible consequences of such a disruptive influence. Following this Chapter 2 analysed existing research to determine the status quo in the area. A classification of the related work indicated the essential components of existing architectures. At this point it was decided to affirm the relevance of the research area, as existing data is predominately grounded in qualitative studies. Chapter 3 discussed the resultant survey which provides quantitative support for this research. Part I concluded by examining presence (refer to Chapter 4) as an enabler for more efficient communications. The chapter also emphasized the privacy of personal information as an important consideration for presence systems.

Part II continues the thesis by developing a prescriptive model to control disruptions in mobile communication. The model is is influenced by concepts and knowledge from prior research, as well as information obtained through the survey conducted by this research. The model is also strongly based on presence with a focus on protecting the privacy of sensitive user information.

Part II

The Model

Chapter 5

Conceptual Foundation

After defining the research objectives, reviewing related work and examining relevant technologies the thesis now moves to propose a solution that will satisfy the problem. This is the first of four chapters which define a model that addresses the balance between availability for mobile communication and disruptive interruptions. The distributed model definition allows a logical progression from a conceptual to a well-defined and detailed view.

The model is prescriptive in that it defines the core features and functionality from which an implementation can be developed. This core is drawn from previous research, relevant technology architectures as well as features specific to mobile communication. By combining proven theories and the latest technologies, with a specific focus on user privacy, a unique combination is achieved. As such it is believed to present a novel approach to the research problem.

The current chapter provides a conceptual overview of the model. To clarify the reader's thoughts, and to encourage a user-centered design approach, the next section reviews the problem domain from a high-level perspective. Thereafter the functional scope is discussed, defining the core features and functionality of the model. Next an overview of the model's primary constructs and relationships follow which form the building blocks for the following chapters. Finally, a chapter layout shows the topics that will be addressed in the next three chapters.

5.1 Understanding the Problem

A drawing can often help in the understanding of a problem and the representation of the situation in which a problem occurs. This section expresses the problem situation using a tool taken from the Soft Systems Methodology domain, known as a rich picture.

A rich picture is a tool which allows a problem to be defined and expressed visually. Checkland (1981) originally defined a rich picture as "the expression of a problem situation compiled by an investigator, often by examining elements of structure, elements of process and the situation climate." In complex problem domains a rich picture identifies all the stakeholders and their relationships effectively and concisely (Monk and Howard, 1998; Fillery et al., 1996). It establishes the issues which concern the people involved and focusses on the interactions which take place (Fillery et al., 1996).

Figure 5.1 presents a rich picture of the problem domain addressed by this research. The figure tries to illustrate several points. A caller often wonders about the availability of a receiver – guessing when to call makes communication much less efficient. In addition, the context of the caller is usually not available to the receiver, but may be important to determine the success of a call. A receiver can be in different situations when a communication is received and this affects whether it is seen as an interruption. The receiver's situation may be affected by the current task or people in the vicinity. If a call is not welcome it can be disruptive to the receiver as well as bystanders in the immediate vicinity. It would be useful if the receiver's mobile phone can adapt according to the current context. This context may also affect how calls are handled.

A receiver is likely to attend to calls if there is a relationships with the caller – social relationship theories suggest that a hierarchy exists, such as family, then friends and co-workers and finally other callers. Thus social relationship often dictates which communications are attended to. Finally, a receiver has a certain presence state depending on his or her context and the relationship with the caller. This also affects the perception of a communication.

Having discussed the problem domain and interactions which take place the next section summarizes these concepts by defining the core features and functionality of the model from a logical perspective.

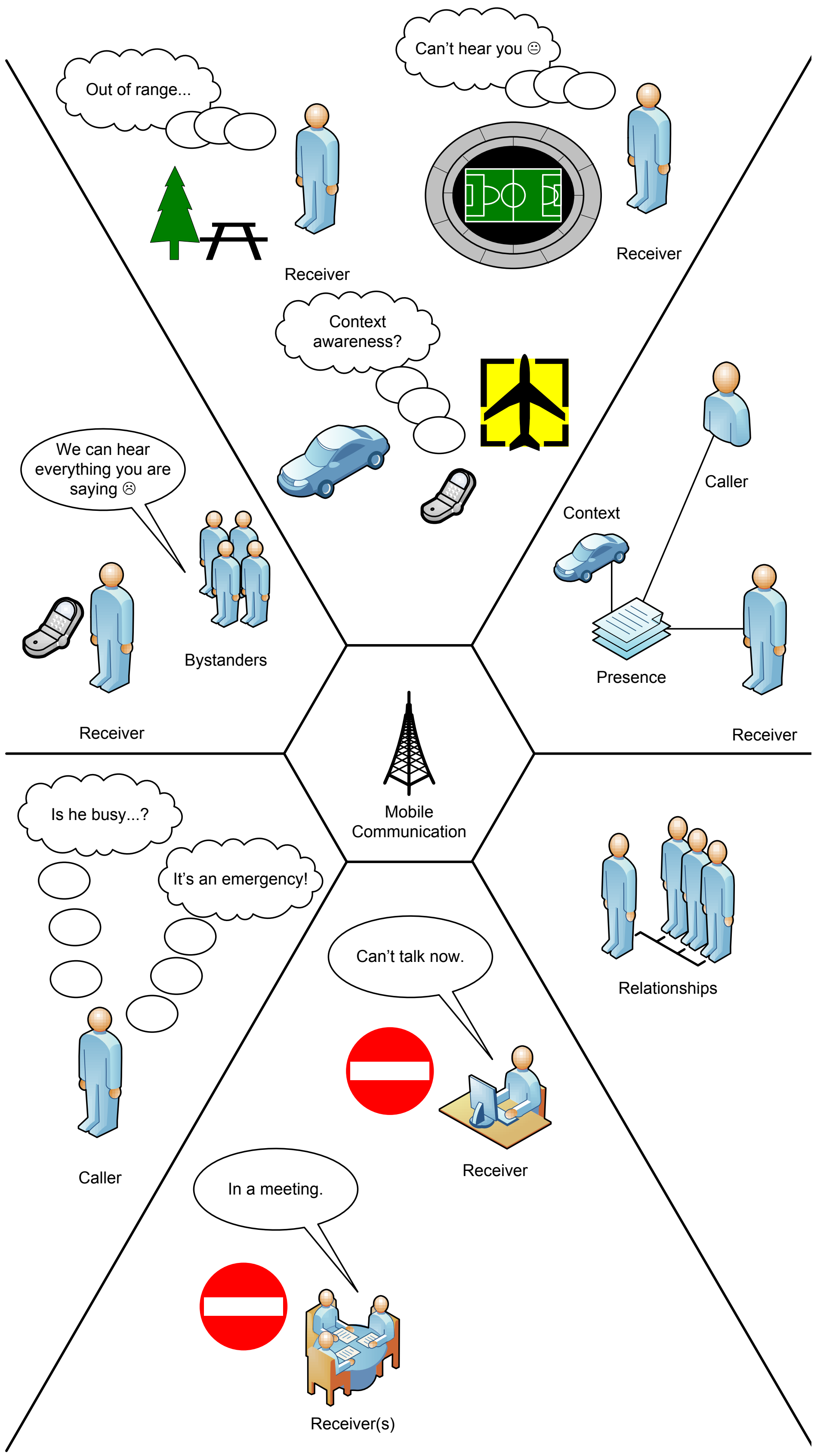

Figure 5.1: Considerations in the Problem Domain

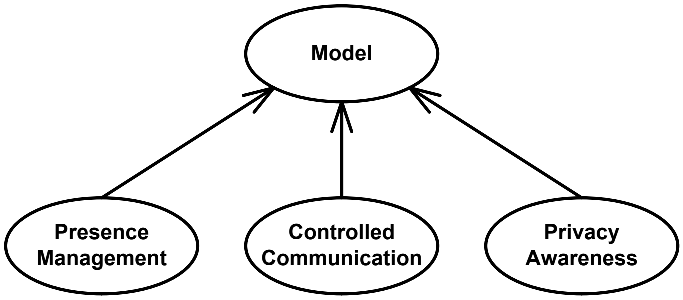

Figure 5.2: Functional Scope of the Model

5.2 Functional Scope of the Model

This research follows a design science methodology which is an iterative process by nature. Design, implementation and evaluation follow each other and are repeated until a satisfactory solution is reached. As part of the design process Vaishnavi and Kuechler Jr. (2008) suggest several patterns to produce an effective solution. The model uses a hierarchical design pattern in which the complex whole is divided into smaller and simpler parts, thus aiming to reduce the complexity of the system.

As Figure 5.2 shows, the model provides three fundamental functions: allowing users to manage their availability for communication through presence, controlling incoming and assisting outgoing communication and providing a privacy aware environment. Each of these functions play an equally important role in the model and together they represent the solution to the research problem.

The model differs from previous research in its use of presence standards to indicate availability for communication. While previous projects have used different types of information to indicate availability, as discussed in Section 2.1.3, the information was not based on presence elements. There are several advantages to using presence: it is robust, provides many core features, is extensible and has been proven over time. These benefits are especially relevant to mobile communication where a multitude of different entities interact with each other.

Presence can be used to provide availability information before a call is initiated. While presence does not directly control communication the information can be used to minimize interruptions. It also provides valuable cues in terms of availability which can save users from making unanswered calls. It would also be possible to generate dynamic communication decisions on the basis of presence information.

A final and important function of the model is to provide privacy awareness. Because of the potential to contain sensitive information the flow of presence needs to be controlled. This includes limiting access to such information and reducing the level of detailed contained therein. By including privacy-aware mechanisms presence can be used to its full potential without concern over unauthorized information access. The model uses a unique privacy approach based on social relationships, as elaborated upon in Section 4.4.2. While previous research has dealt with the concept of privacy, as discussed in Section 2.4, none have used social relationships as a method to control presence sharing.

Each of the following three chapters will divide these fundamental functions into smaller parts to create more detailed definition of the model. However, in line with the design science methodology, the conceptual vocabulary of the model needs to be defined first. The entities within this vocabulary are also known as constructs. The model is based on various constructs from the research field and related technologies. In addition the subject areas also define some of the construct relationships. The model's constructs and their relationships are discussed in the next section.

5.3 Constructs and their Relationships

The proposed model is a combination of several elements and entities from mobile communication, presence systems and privacy theories (in particular those related to social relationships). The following subsections form the foundation of the model by defining the key constructs and their relationships. Together these entities provide a description of the model and its intended purpose.

The Unified Modeling Language (UML) is used in a standards-based approach to present these constructs and their relationships. According to Miles and Hamilton (2006), UML is a formal language using an easily understandable notation in which each element has a strongly defined meaning to avoid confusion. It is also comprehensive which allows all aspects of the solution to be modelled, and scales well for complex systems.

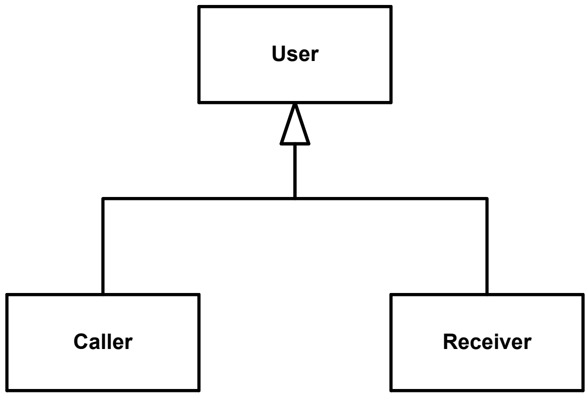

Figure 5.3: Callers and Receivers

5.3.1 User

As the model focusses on communication and personal privacy the user is a central construct in the model. Users represent the human entities in the model and can be broken down into two primary types: callers and receivers. Figure 5.3 illustrates this generalization.

It can be observed that the roles of caller and receiver correlates closely to the roles of watcher and presentity in presence systems. For example, a caller (or watcher) is usually interested in the situation of the receiver (or presentity) before initiating a call.

A caller is the user who initiates or places a call, while the receiver is the user who is the target of the call. A user can exist independent of communication – for example, the receiver does not necessarily have to answer the call, the call can be cancelled by the caller or receiver, the call can be diverted to voicemail or put on hold. However, in all cases an interaction between users is implied. Another important point is that a user can act as a caller or receiver at different points in time, depending on the communications context.

The model limits simultaneous interactions to two users – a caller can only call a single receiver at a time and the receiver can only communicate with a single caller. However, in mobile communication it is possible for a caller or receiver to be connected to multiple users simultaneously.

5.3.2 Context and Presence

Regardless of whether users are acting in the role of caller or receiver they are behaving within a certain situation or context. In the case of a caller

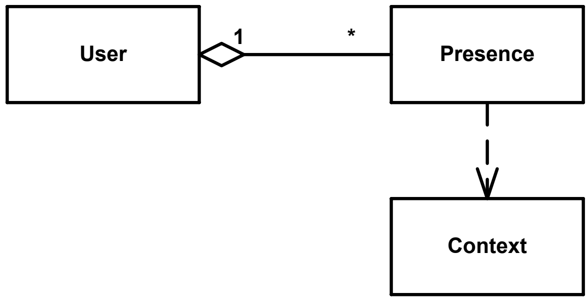

Figure 5.4: Context and Presence

the context influences aspects such as the urgency of the call. In the case of a receiver the context determines the availability for communication. Thus both the caller and receiver are dependent on their context. A change in context likely leading to a change in communication behaviour. The absence of context information is an obvious deficiency of current mobile communications. To clarify the intended meaning which context may have for a user the model abstracts this information in the form of presence. This relationship is shown by Figure 5.4.

In practice a user's current availability may be indicated by various items of presence information. However, each item of presence has a unique meaning to the user in question and is only applicable to that user. As presence is a comprehensive technology not all information is applicable as availability cues in mobile communication. The model recommends limiting presence data to only the most useful availability cues in order to prevent confusion and to allow fast call decisions to be made.

5.3.3 Relationships

As presence may contain information of a sensitive nature it is natural that users may want to control or limit access to it. The model prescribes a privacy implementation based on social relationship groups. This is a unique approach in the context of previous work in this domain. Theory suggests that each user operates within multiple spheres of relationships. This is illustrated in Figure 5.5.

Close relationships usually indicate a willingness to share more personal information. As relationships become less intimate concern over the privacy of personal information grows and thus the willingness to share such information decreases. However, as relationships are complex and unique from

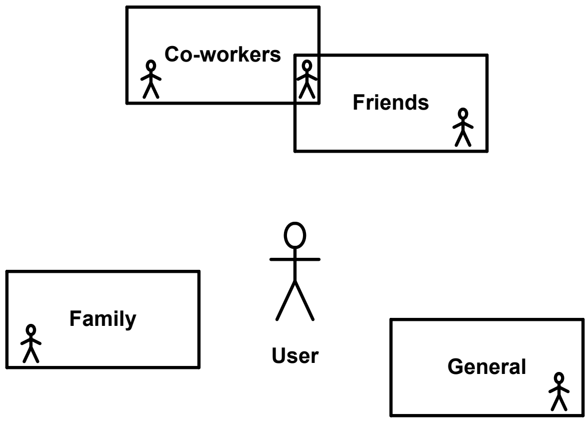

Figure 5.5: Relationships

one user to another, no strict mapping of privacy requirement to relationship can be made.

Unfortunately relationships are not always the only consideration when deciding if personal information should be shared. Certain situations may force users to share information irrespective of their personal feelings. An example may be a company forcing employees to remain logged on to a communications system and displaying their availability status. These situations are acknowledged by the model but are not addressed in this research.

5.3.4 Presence Authorization Policy

To determine the level of presence visible to others each user maintains a presence authorization policy. A policy is unique to a user and must be created and maintained by that user. Figure 5.6 illustrates these points.

A policy contains the mapping between identities and the level of presence access they are authorized to. This allows a user to reduce the granularity of presence information for certain recipients or remove access to presence completely. A policy provides its owner with privacy protection and thus forms an integral part of the model. While policies are defined by presence

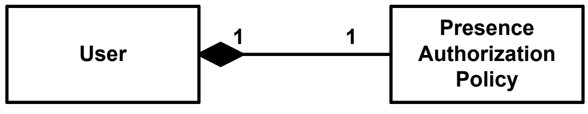

Figure 5.6: Presence Authorization Policy

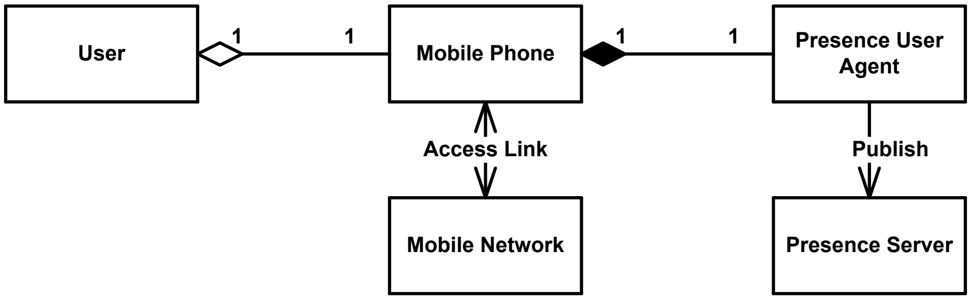

Figure 5.7: Presence User Agent as part of the Mobile Phone

standards no known implementations currently exist. Thus the model differs from previous research in its use of presence authorization policies to control access to mobile communications availability information.

5.3.5 Presence User Agent

Another element which forms an integral part of the model is a presence user agent (PUA). This software agent is tightly coupled to the user and runs on the mobile phone. In mobile communication a user is closely linked to a mobile phone. The mobile phone provides access to the mobile network and can be either a regular phone, smartphone or personal digital assistant. For the purposes of the model a user is associated with a single mobile phone even though a user may own multiple mobile phones in the real world. The model is also not concerned with the type of access link between the mobile phone and the mobile network; for example, a connection can be established using SMS, USSD, cellular radio or WLAN. The ability to use any access link is a strong point of the model. The relationship between the user, mobile phone and network and PUA is shown in Figure 5.7.

The PUA has two key functions: monitoring the user's context for changes in presence and publishing presence updates to a presence server. A user can have multiple PUAs on various mobile phones or networks each collecting and publishing presence. However, for the purposes of the model the PUA is defined as a single element residing on the mobile phone.

5.3.6 Presence Server

A presence server is the entity responsible for collating user presence and making it available to interested parties according to the owner's privacy

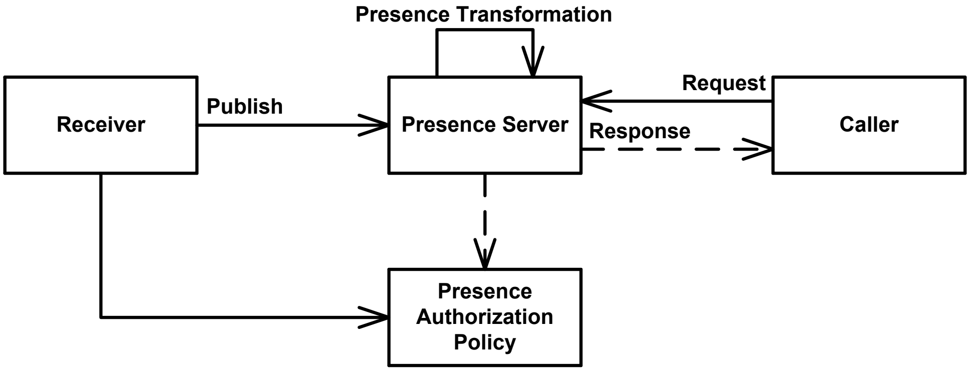

Figure 5.8: Presence Server

preferences. Thus it forms the final link between the constructs presented thus far. Figure 5.8 illustrates the presence server's relationship with other entities.

The presence server is responsible for collating all sources of user presence and serves as a single point of access to such information. As requests for presence are received from callers, the server determines the correct response based on the receiver's presence authorization policy. This policy is co-located on the server and serves as input for presence transformations. Transformed presence documents are returned to callers, containing the receiver's privacy filtered presence state. For the purposes of the model a single presence server is assumed to handle all server functionality.

Having defined the basic constructs and their relationships the thesis can now proceed to discuss each aspect of the model in more detail. A short summary of the following chapters is presented next.

5.4 Model Layout

As defined in Section 5.2 the model has three fundamental functions. Each of these functions form the foundation of the following chapters. In order these chapters discuss:

1. monitoring for context changes, obtaining presence from context and publishing presence updates,

2. controlling communications by presenting availability cues to callers, automating call handling and allowing callers to express relevant communication context, and

3. maintaining a presence authorization policy to protect the privacy of presence information.

Figure 5.9 illustrates the influence of the caller or receiver on each chapter and how they connect to each other. While not intrinsically part of the model, this figure serves as a navigational aid in understanding the presentation of the model in the following chapters.

The first part of the model focusses on the receiver's situation. This includes the monitoring and conversion of context into presence by the PUA. Thereafter presence needs to be published to a presence server. The availability of presence directly influences the control of communication by presenting availability cues to callers.

The second part of the model shows how presence information is used to control communications. For callers this entails the consumption of receiver presence before making a call. In addition it is possible to transmit communication context to increase the chances of a busy receiver answering a call. Receivers also have the option of handling calls automatically depending on their presence state.

The third part of the model shows how concerns over privacy can be handled by using presence authorization policies. Such a policy can protect the receiver's presence information by transforming it before being shown to callers. In addition it is argued that the management of policies is easier and natural through the use of relationships to define presence access.

The next section will provide some concluding remarks to summarize the current chapter's main points.

5.5 Conclusions

The current chapter provided a conceptual overview of a prescriptive model to answer the research problem. It presented a high-level look at the model's fundamental functions, constructs and their relationships. This provides the background necessary to understand the following chapters. The next chapters will define each of the fundamental functions in more detail, providing a detailed and complete view of the model.

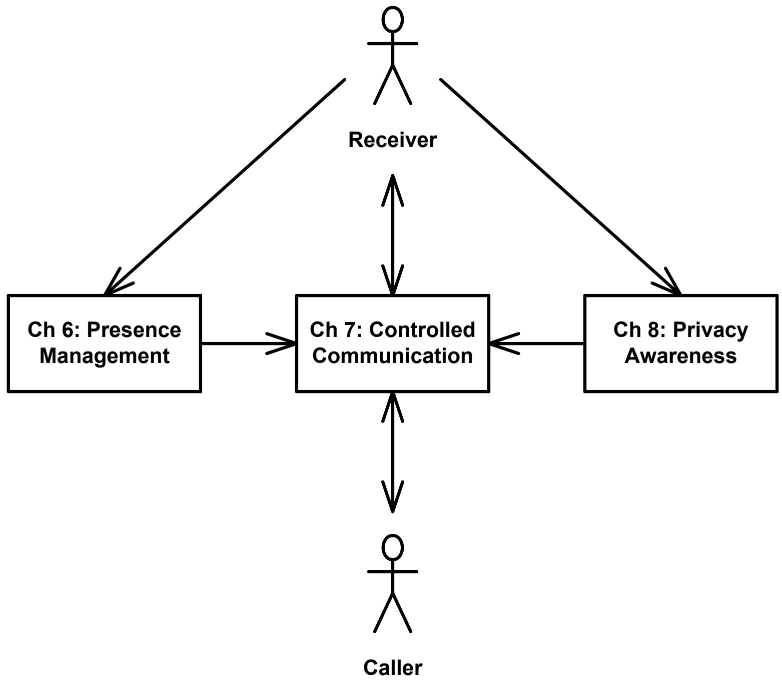

Figure 5.9: Model Presentation in the Context of the Thesis

Chapter 6

Presence Management

A user's current situation greatly influences the ability to accept a call. Context is an effective way to model the current situation as it directly describes several facets of the user's current situation. However, because there is no standardized way to model context it can be interpreted in several ways. Thus it is useful adding another data layer on top of context – presence – which presents the same data in a standard format with a much more specific meaning.

One of the core features of the model is the use of context and presence. While common on the Internet as part of instant messaging systems, presence is not part of mobile communication. Yet it can serve a valuable role in solving the issue of unwanted interruptions to the receiver of a call by indicating availability to communicate. This makes the model different from previous research in the domain.

Context forms the basis for presence. The user and the mobile phone are sources of information which can be used as the basis for context. But users have their own interpretation of the environment and the resulting context. Thus a defined view is needed for the standardized use of such information. Presence is used in the model as a provider of such a standard view of the current user state.

This chapter examines the interaction between context and presence, and how presence is communicated from the user to the network. First, the sources of context and their monitoring is discussed. Next presence is examined, including how updates are published. Thereafter the conversion of context to presence is discussed. Finally, the composition of multiple presence sources into a single view is defined.

6.1 Context

The context data which describes the user's availability is one of the most important parts of the model. The user, phone and environment all provide a rich source of such data. However few, if any, mobile communication applications currently use this information to provide a better user experience. Changes in the situation will affect the availability for communication and thus needs to be monitored. The next subsections will describe sources of context data used in the model and how these sources can be monitored for changes.

6.1.1 Information Sources

As discussed in Chapter 2, several context sources can provide valuable availability information. However, interaction is usually based on the user's situation or interaction with the environment.

Definition 1 (Context Sources). *Context is obtained from the user's environment and situation. Such data shall be obtained through operating system software which reports data from hardware, external sensors and personal information.*

The model allows any context to be used in determining the user's availability. However, previous research has identified context cues which prove useful and are easier to obtain (see Section 2.2). In light of this the model recommends that at least the following context be used:

- location, obtained through the most efficient means available to the user, such as the mobile cellular radio or external sensors,
- activity, especially physical and social activity, based on the user's calendar information, and
- device usage information determined by the user's current call state.

A user's context can be obtained using various sources. Because the phone is usually with a user throughout the day it is seen as a rich source of context information. The various ways in which the phone can sense context is modelled in Figure 6.1.

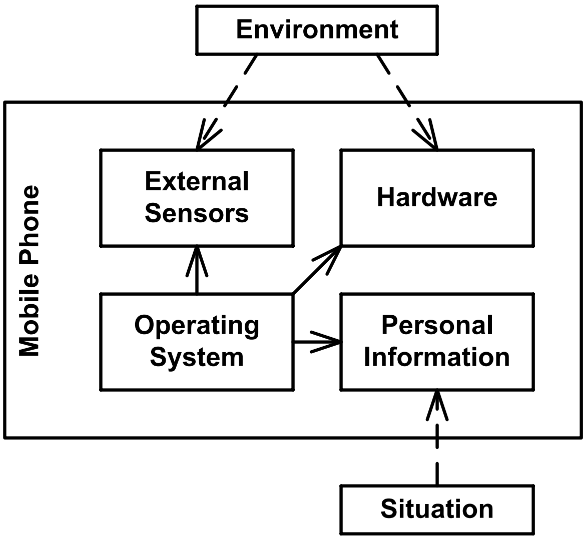

Figure 6.1: Context Sources on the Mobile Phone

The phone interacts with its environment through the available hardware components. This can include the mobile cellular radio for communications, microphone and speakers for speech and sound interaction, a camera for capturing video and pictures and a battery. Each of these components can suggest something about the current context of the user. The mobile cellular radio not only performs communications functions but can also be used to determine a mobile phone's location relevant to a mobile cellular coverage area. The benefit of this method is that it works anywhere in a reception area, both indoors and outdoors, at the cost of accuracy. The battery can provide information on remaining power and thus how much talk-time remains. The camera and microphone allow environmental information, such as light and noise levels, to be obtained. Such information could be used to determine the appropriateness of a conversation.

External sensors usually complement a device by adding missing functionality such as a global positioning system (GPS) or bluetooth component. A GPS component allows the accurate determination of a mobile phone's location outdoors, while bluetooth provides an indication of nearby units such as other devices or sensors.

Access to hardware and external sensor data is provided through software layers built into the mobile phone operating system. Because of its tight integration with the platform the operating system can report data fast and accurately. Thus it is useful to integrate into the operating system to retrieve data.

In addition the operating system can also expose personal information such as a user's calendar or contact list. Such information is often influenced by the situation of the user. As users become more familiar and dependent on mobile phones they store an increasing amount of information on them. In addition many mobile phones synchronize with desktop computers and online services which increase the accuracy of information. Such information, such as a user's calender of appointments or contact list can inform the operating system of what the user is currently doing or with whom the user is communicating. The presence user agent (PUA) is responsible for discovering available context sources and collecting information from them. The model does not prescribe how this should happen – the PUA can create an event listener for state changes or poll for updates.

Definition 2 (Minimum Set Of Context). *The minimum set of context consists of location, activity and device usage. The presence user agent is responsible for discovering and collecting context information.*

As context is dynamic and changes frequently it needs to be relevant to be useful. The next section will examine how context changes can be taken into account.

6.1.2 Monitoring for Changes in Context

A change in the environment or situation can be triggered at any time and can be sensed by the phone using external sensors or the built-in hardware. This flow of information is illustrated in Figure 6.2.

When the phone detects such a change it notifies the operating system by sending the relevant data to be processed. This shall be done asynchronously so that multiple context changes can be monitored. The operating system processes the data and abstracts it to context data. This data is then sent on to the PUA for further processing.

Context is relevant in determining whether a user is available to communicate. However, as raw data it is difficult to interpret and use. In addition some information sources could contradict each other or need to be merged to make sense. A conversion from context to presence needs to take place which provides a clearer indication of availability. This will be discussed in the next section.

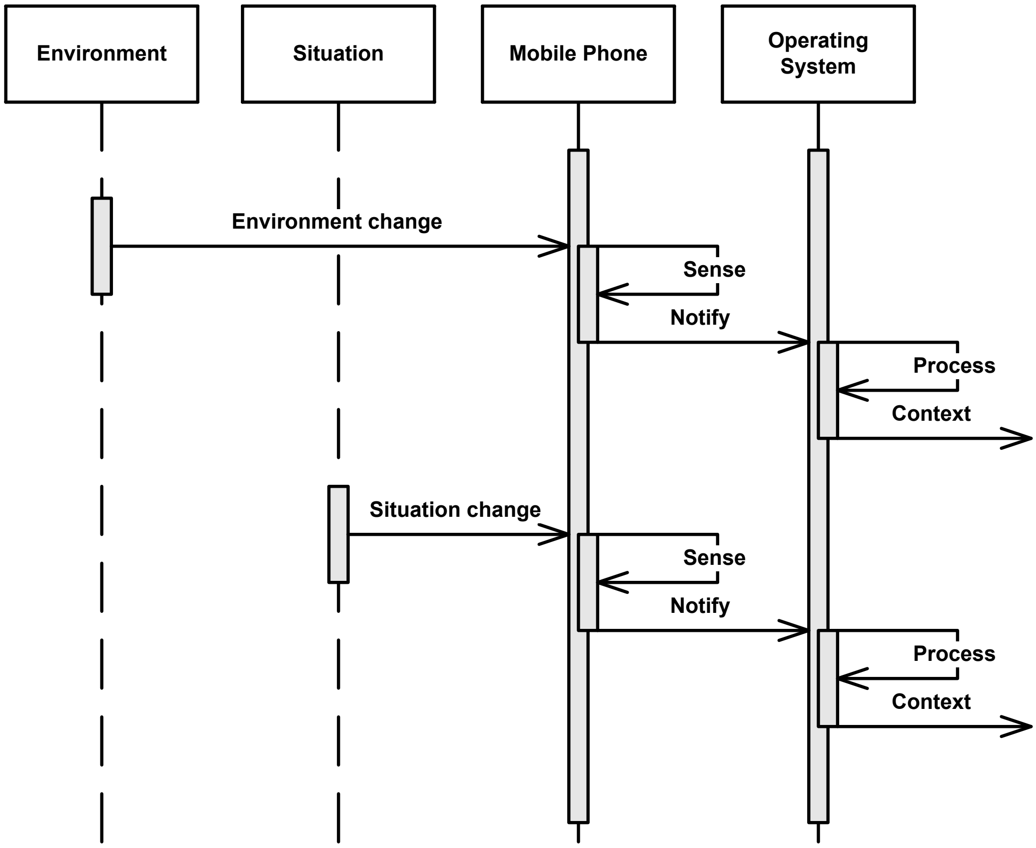

Figure 6.2: How Context Changes

6.2 Making use of Presence

Presence standards are comprehensive and define many data elements to describe the user's state. Thus it is necessary to define which elements will be used by the model and how context will be converted into these elements. Using presence elements relevant to mobile communications is a key differentiating aspect of the model. This makes the model different from previous research which used various unstandardized items of information, as discussed in Section 2.1. This use of presence is the focus of the next two subsections.

6.2.1 Presence Data

The model converts context data to presence information to make use of it. Unfortunately a direct conversion is not always possible as context portrays low level data which needs to be abstracted into presence information. Thus some form of reasoning needs to be applied in the conversion process. However, first the relevant presence data must be identified. Presence can include a large set of information from which the model uses a subset to portray the availability of the receiver. This subset is based upon the SIMPLE presence standard, as defined by Rosenberg (2006), and includes the following elements:

- basic status,
- activities,
- sphere (location),
- user-input, and
- an optional note.

The above items represent dynamic information about a single user. The *basic status* element is a binary indicator of availability for communications. It can have values of either ‘closed’ or ‘open’. Closed means that any communications is likely to fail or not reach the receiver. As an example, if a user is already involved in a conversation a subsequent call would be forwarded to voicemail and the basic status should be set to closed. Thus many different context cues may have an effect on this value.

The *activities* element describes what the user is currently doing, expressed as an enumeration of descriptive elements. A user can be engaged in multiple activities at the same time. Because they can be derived from calendar information it is recommended that the following enumeration be used as possible values for activities:

- appointment: The user has a calendar appointment, without specifying exactly of what type. This activity is indicated if more detailed information is not available or the user chooses not to reveal more information.
- breakfast: The user is eating the first meal of the day, usually eaten in the morning.
- dinner: The user is having his or her main meal of the day, eaten in the evening or at midday.
- holiday: This is a scheduled national or local holiday.

- lunch: The user is eating his or her midday meal.
- meal: The user is scheduled for a meal, without specifying whether it is breakfast, lunch, or dinner, or some other meal.
- meeting: The user is in an assembly or gathering of people, as for a business, social, or religious purpose. A meeting is a sub-class of an appointment.
- performance: A performance is a sub-class of an appointment and includes musical, theatrical, and cinematic performances as well as lectures. It is distinguished from a meeting by the fact that the user may either be lecturing or be in the audience, with a potentially large number of other people, making interruptions particularly noticeable.
- travel: The user is on a business or personal trip, but not necessarily in-transit.
- vacation: A period of time devoted to pleasure, rest, or relaxation.

The activity values are not intended to be a complete set. Rather than prescribing specific values the model allows users to customize activities with values applicable to their context. This provides necessary flexibility in the wide range of contexts in which mobile communications take place.

The *sphere* element designates the current state and role that the user plays. For example, it might describe whether the user is in a work mode, at home, or participating in activities related to some other organization. Common values for sphere include ‘work’ and ‘home’, as well as ‘unknown’. Spheres allow the user to easily turn on or off certain rules that depend on what groups of people should be made aware of the user’s status. For example, if the user is a Boy Scout leader, he might set the sphere to ‘scouting’ and then have a rule set that allows other scout masters in his troop to see his presence status. As soon as he switches his status to ‘work’, ‘home’, or some other sphere, the fellow scouts would lose access.

The *user-input* element records the usage state of the mobile phone based on human user input, such as using the keypad or voice. The element can assume one of two values, namely, ‘active’ or ‘idle’, with an optional *last-input* attribute that records when the last user input was received. An optional *idle-threshold* element records how long the receiver will wait before reporting the service or device to be idle, measured in seconds. If the element wants to indicate user input activity it sends an active indication when the user has provided user input within a configurable interval of time, the idle-threshold. If the user ceases to provide input and the idle-threshold has elapsed, the element is marked with an idle indication instead, optionally including the time of last activity in the last-input attribute. The user-input attribute can be omitted if the receiver wants to indicate that the device has not been used for a while, but does not want to reveal the precise duration. The model prescribes that the user-input element be explicitly used to indicate current call activity. This information can be especially useful to a caller, as it is unlikely that a person will take two calls at the same time.

Lastly an optional *note* element can be used as a free text indication of status. This allows the receiver to specify a custom message with specific availability information.

All the above elements, except for the note, can also can also have a *from* and *until* attribute. This describes the absolute time when the element assumed this value and the absolute time until which this element is expected to be valid. While not limiting information only to these elements, the model recommends that these attributes be included for the activities element.

Definition 3 (Presence Elements). *The basic status, activities, sphere, user-input and optional note elements are used to indicate user presence.*

Now that the context and presence elements used by the model have been defined the model will address the issue of transforming context into presence.

6.2.2 From Context to Presence

While not mandated by presence standards, the model dictates that the PUA is responsible for transforming context data into presence. This makes logical sense as the PUA is closely linked to the context sources. This transformation process is illustrated in Figure 6.3.

The PUA accepts context data from the phone's operating system and transforms it into presence. Three factors are of importance in the transformation process: it should be fast, have minimal resource requirements and be accurate. Many strategies exist to effect this transformation such as key-value, object-role, spatial and ontology-based models (Bettini et al., 2010).

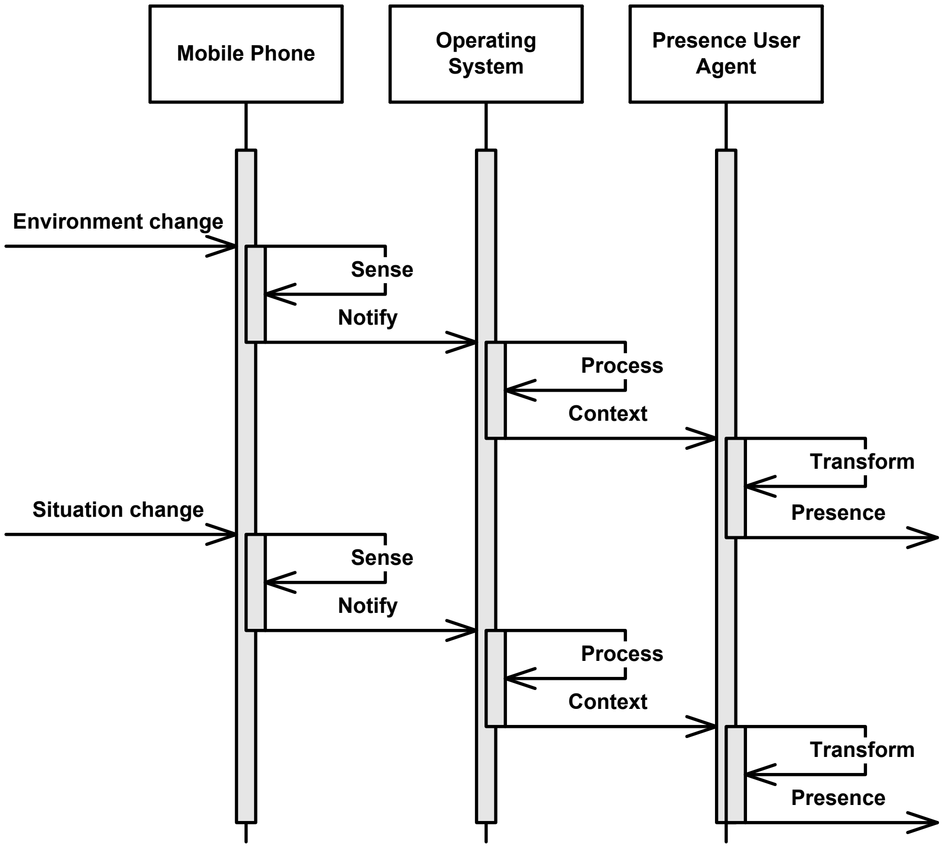

Figure 6.3: From Context to Presence

However, these strategies can be complex and present research areas in their own right. Thus the model opts for a simpler approach which still achieves the goal.

The model proposes that an event-condition-action (ECA) pattern be used to transform context into presence information. While being easy to understand and process, this pattern allows combining several conditions to create rules. For the purposes of the model these rules suffice to convert context into presence. ECA rules have the following format:

- the *event* specifies the triggers that invoke an action,
- the *condition* is a logical test that determines whether the action should be executed, and
- the *action* determines how to react.

The structure of an ECA rule can be defined as "ON event IF condition DO action". To complete the rule structure it is also necessary to define each

Table 6.1: ECA Transformation Rules

Event	**Condition**	**Action**
calendar event	no additional information	activities = 'appointment' + set from/until time
calendar event	metadata available	activities = metadata + set from/until time
location change	location known	sphere = location
location change	location unknown	prompt user + sphere = location
incoming call	call accepted	user-input = active + set last-input
outgoing call	call accepted	user-input = active + set last-input
call ended		user-input = idle

part of the rule. This research defines a minimal set of rules which the user can extend. The important qualities of these rules are that they influence the applicable presence elements and are derivable on a mobile phone. Table 6.1 presents the default ECA transformation logic for the presence elements recommended by the model.

The table shows how context influences the relevant presence elements. The table is not intended to be the complete set of supported events. Rather the model suggests some default rules with the intention that users are able to define additional rules to extend the capabilities of the PUA. For example, if location specific hardware was available this could be used to provide more accurate positioning information. The logic focusses on taking relevant calendar context data and converting it into the appropriate activity presence element. In addition the current location is used to set the sphere, while the call state sets the user-input. The receiver can also add basic status information of open or closed to indicate absolute availability despite any of the above elements.

Definition 4 (Context Transformation). *An ECA pattern is implemented by the presence user agent to transform context into presence information.*

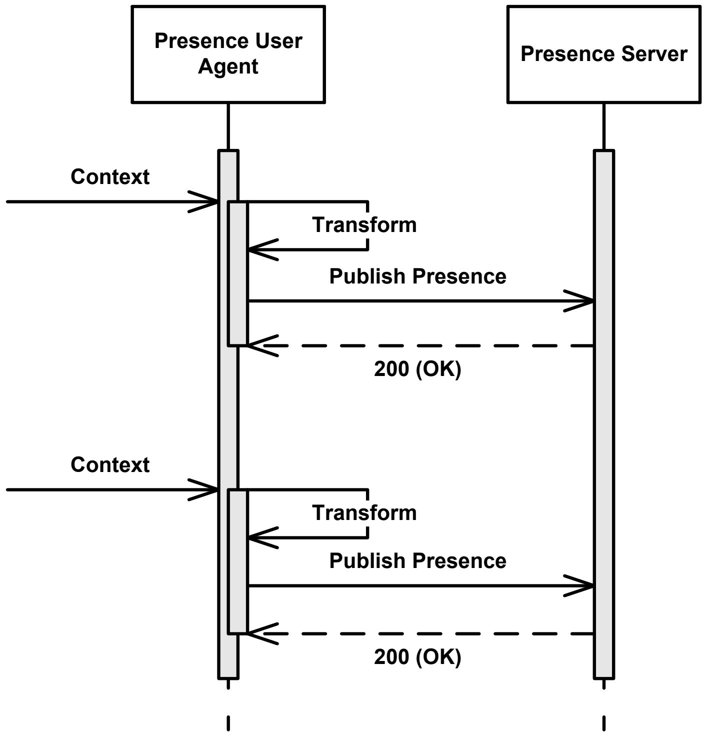

Figure 6.4: Publishing Presence

After the information has been transformed into presence it can be published to the presence server. This process of transmitting the data can be done using GPRS/3G, WLAN or any other available connection to the presence server. This is discussed in the next section.

6.3 Publishing Presence Updates

The model follows the standard publish architecture of presence standards, as described in Section 4.3.2. It is briefly discussed here for completeness.

Figure 6.4 illustrates the various entities involved in the publication of presence. After each change in context the PUA performs a transformation of the data into presence. This is then published to the presence server immediately. If the presence server received the data successfully it responds with an OK message. A lack of success message could indicate a break in network communication. The model does not prescribe how the PUA should handle any failed attempts – the publication could be attempted again or postponed until the next context change.

A final step in processing the presence data once it reaches the presence server is composition. This combines multiple presence documents into a single unified view. This is discussed in the next section.

6.4 Presence Composition

A final consideration in this part of the model is how multiple presence updates should be handled. Almost all existing presence services involve a single PUA which maintains a complete presence state for a user. This allows for a simple model where the PUA publishes a full presence state to the presence server.

However, the presence state for a user may be derived from multiple inputs. In this case the complete presence view for a user is composed of the presence state from each source. The composition of presence is a complex process with several steps. In the process of combining input data into a presence document the following steps may be needed: discarding (conflicting) information, deriving presence information, resolving conflicts and merging information (Shacham et al., 2007). As this does not form a fundamental part of what the model aims to achieve it is prescribed that the model only uses a single PUA which always publishes a full presence state to the presence server. However, in the future this could be a valuable extension of the model.

Definition 5 (Presence Composition). *Presence composition uses the published presence in its current format. Each user has a single presence user agent which maintains the complete presence state.*

After successfully processing context into presence a user's availability can be updated. At this point the presence document is ready to be consumed by watchers.

6.5 Conclusions

Figure 6.5 presents a summary of the main constructs and interactions described in this chapter. As can be seen, the focus is on obtaining relevant context data, converting this data into presence and publishing it to the network. The PUA is the main component responsible for these actions,

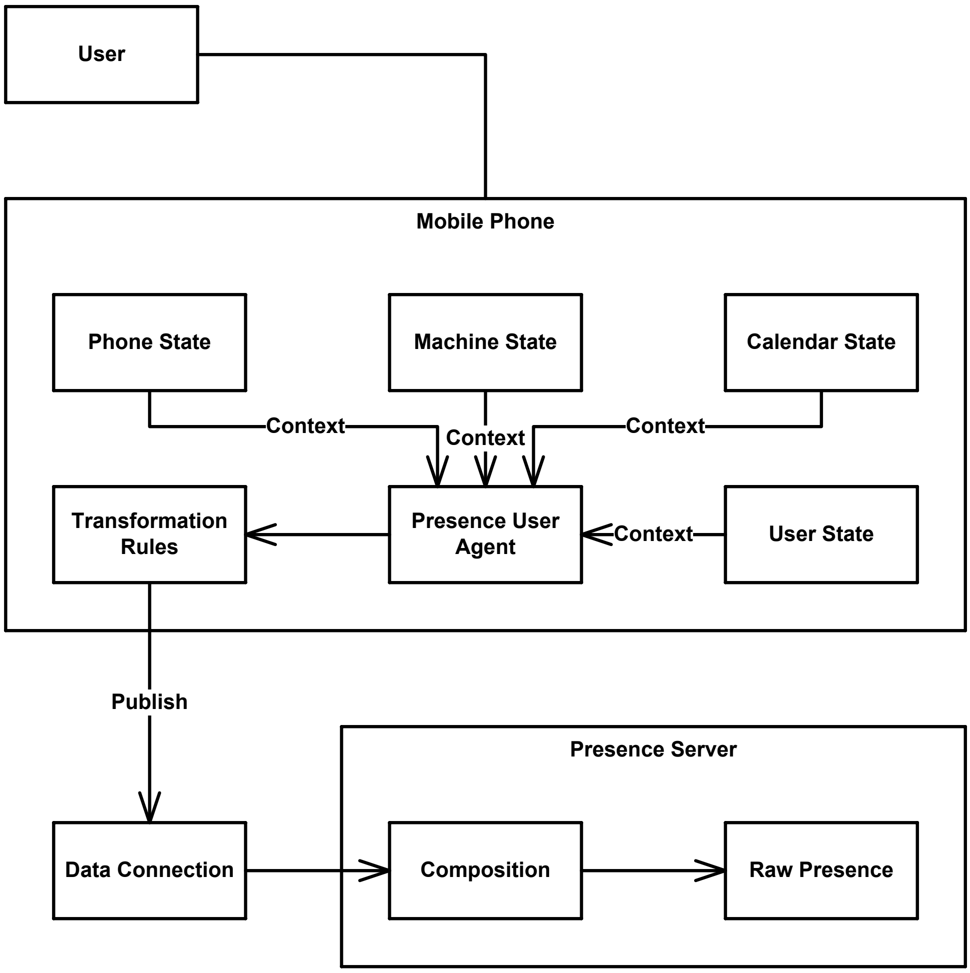

Figure 6.5: Model Summary

collecting, transforming and sending presence to the presence server for composition.

The process of collecting context data, converting it to presence and publishing the results is a fundamental part of the model. This chapter has defined the constructs and relationships applicable to this process.

Although largely based on existing standards, the model presents a new interpretation of presence in mobile communications. Users often suffer from disturbances which presence can help minimize. However, it is uncommon to see it used as such in mobile communication.

The next chapters build on these constructs by defining how this information is used in the management of communications. While the presence publication process is important it is of little value without distribution to callers. The next chapter models the process of using presence for availability cues in the flow of making a call.

Chapter 7

Presence for Call Management

The model uses presence information in the call process to reduce unwanted interruptions and limit unsuccessful call attempts. Where the previous chapter defined the process of obtaining presence information the current chapter defines the use thereof. A receiver's presence is of no use if callers cannot access this information. The caller benefits from an initial indication whether a communication is appropriate while the receiver may experience fewer interruptions because of considerate callers.

As mentioned in Chapter 2, three main approaches to managing communication have been used: receiver-oriented, negotiated and caller-oriented. The model uses mechanisms from each of these approaches to provide a holistic solution to the research problem.

This chapter presents the part of the model which allows a caller to retrieve presence and use it in the call management process. First the concept of a user is expanded to include the necessary information for the functioning of the model. Thereafter the retrieval of presence and display thereof is discussed. The ability to indicate caller context is examined next, before finally moving on to discuss receiver options for automatically handling a call.

7.1 Users

Mobile communication provides a link between users in the network. Chapter 5 defined two types of users: callers and receivers. Establishing a communication link between users is primarily dependent on the caller knowing the network identity of the receiver. The identity of a user could be based on:

- a cryptographic shared secret or a certificate,
- a network address based on standard presence systems,
- the International Mobile Subscriber Identity (IMSI) or International Mobile Equipment Identity (IMEI), or
- the Mobile Station Integrated Service Digital Network (MSISDN) identifier, more commonly known as a phone number.

The model defines a telephone number as the primary identifier of a user. Because of its human readability a phone number is a common method for users to identify each other in mobile communications networks. It is also globally unique which ensures that communications can be routed correctly between users. It is possible that a user can have multiple phone numbers; in such cases each number acts as a pseudonym for the user, providing a measure of privacy to the user. As the model implements privacy protection in a different manner such cases are not supported - each number is treated as a separate identity. This allows a user to be identified by a phone number.

The model also accepts that a phone number is used by a single user. It is acknowledged that in real life many users could share a single phone number, as in the example of a public pay-phone. In addition the survey in Chapter 3 also indicated that the type of users likely to benefit from this model would potentially not share their phone number. However, it is practically infeasible to differentiate between users in this situation and thus it is left to the user to ascertain the association between a phone number and another user.

Definition 6 (User Identity). *A user shall be identified by a phone number following the tel URI scheme. This phone number must be globally unique. The model assumes that each user is associated with a single phone number and that no other users share that number.*

For communication purposes a phone number shall be based on the tel URI scheme defined by Schulzrinne (2004). This also allows messages to be routed across standard presence systems. The tel URI syntax can be summarized as follows:

- the phone number shall be preceded with the string *tel:*,

- thereafter the number should be presented in global format, if possible, starting with a +,
- the number shall consist of the country code followed by the national area code and then the rest of the number,
- the - character should be users as a visual separators to aid readability of the number,
- for example, tel:+1-201-555-0123.

Whereas a caller needs the phone number of the receiver to initiate a call the receiver often seeks to know the phone number of the caller as well. This can be useful in deciding whether to accept the call. However, the provisioning of a caller's phone number is a value-added service known as Caller ID (Petersen, 2002, pp. 161–162), and several factors can influence the availability of such information. First, a receiver must be subscribed to the service which usually coincides with a fee. Second, the caller must not be using the call block feature which allows anonymity by withholding the caller's telephone number from the receiver (Petersen, 2002, p. 158). Third, the phone number is often removed from the receiver's view by underlying systems in the process of routing a call. While the network still uses the phone number the receiver is unable to access it.

The model prescribes that the term 'unknown' be used as a substitute for the phone number if callers do not identify themselves. Thus three states of identity can exist:

- the phone number is available and known to the receiver,
- the phone number is available but unknown to the receiver, or
- the phone number is withheld or removed.

Definition 7 (Unknown Identity). *In the absence of a phone number the term 'unknown' shall be substituted as the identity of a caller.*

When initiating a call a user acts as the caller, while the recipient of the call is known as the receiver. It is possible that a user who is already involved in a call also receives a call from another caller at the same time. Some network providers may allow multiple callers and receivers to be connected at the same time, but this falls outside the scope of the model. The model enforces a simple relationship where only one caller and receiver can be involved in a call simultaneously – a caller can only place a call to one receiver at a time and a receiver can only accept a call from one caller at a time. Thus, as soon as a user is involved in a call he or she becomes unavailable to other users until that call is completed.

Definition 8 (Single Connections). *A user can only be involved in a single connection (call) at a time and shall be unavailable to other users until that call is completed. A user can act as a caller or receiver.*

In addition to using a phone number as user identity it is recommended to associate a name with a user. This name can be displayed to the receiver, instead of the caller's phone number. This allows the receiver to identify the caller quickly. Each associated name need not be unique and can be associated with more than one phone number (or user). Such information is usually stored in a list referred to as an address book or contact list. This list provides an indication of whom a user communicates with frequently.

Having defined the user in more detail the next section looks at how a caller obtains the presence of a receiver with the aim of making a more informed decision about the appropriateness of a call.

7.2 Receiver Availability

Having access to a receiver's presence presents a caller-oriented approach to call management. Before placing a call such information can be used to communicate in the most effective manner. The two considerations in this regard is how to obtain presence and how to display it to the caller. These issues will be examined in the following subsections.

7.2.1 Polling for Presence

Presence systems are usually based on a publish/subscribe model where a caller would get constant presence updates of any user on their contact list. However, the model argues that presence is only needed at the point of communication and thus uses a simpler method for obtaining information – fetching presence information as needed.

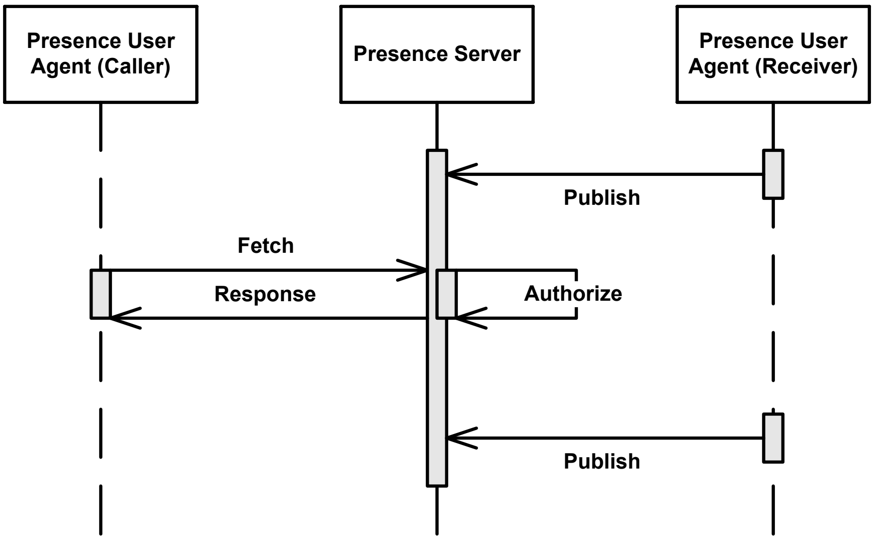

Figure 7.1: Fetching Presence from the Presence Server

The presence user agent (PUA) functions as a presence fetcher, as discussed in Section 4.3.1. In this case the PUA initiates a presence request on behalf of the caller. This process is illustrated in Figure 7.1.

The caller's PUA fetches presence from the receiver's presence server when the caller is interested in the information. Once again the model is also not concerned with the type of link between the mobile phone and the mobile network; for example, a connection can be established using SMS, USSD, GPRS/3G or WLAN. The presence server authorizes the request and returns the resulting response to the PUA. It is important to note that the caller does not receive further presence updates when the receiver's state changes.

Definition 9 (Presence Polling). *Polling for presence information is performed by the presence user agent only when needed. Any type of network connection can be used for this purpose.*

Using a fetching approach reduces the amount of network traffic received on the phone and thus improves battery life. It is also debatable whether the caller needs to be updated constantly of the receiver's presence – it can be argued that this information is only useful at the moment when a call is about to be placed. Thus it can be fetched at the appropriate time. A possible drawback of the approach is that the process may add additional time to the establishment of a call.

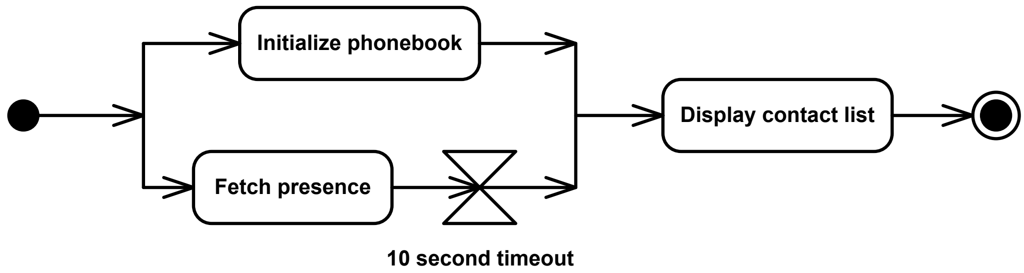

Figure 7.2: Displaying Presence in the Phonebook

After receiving presence this information is used by the caller PUA. This is discussed in the next section.

7.2.2 Integration with the Contact List

When callers open their phonebook they usually do not see any presence about the people in their contact list. While they can call any person in the phonebook it is a guess as to whether the call will disrupt the receiver. The model's goal is to provide the caller with a better indication of the receiver's presence from within the phonebook. This allows a caller to make a communication decision before wasting time on the call establishment process.

Figure 7.2 illustrates how the model accomplishes this. When the caller opens the phonebook application the PUA sends a request for presence to the presence server while the user interface is initialized. This aims to keep the waiting time for the caller to a minimum. If no presence is received within ten seconds the phonebook is displayed without any presence. This value is in line with the findings from Chapter 3. Alternatively, if the request completed in time the caller will see a list of contacts along with their presence. This presents the caller with an initial view of receiver availability.

While having access to the receiver's context is useful a caller may also want to present context information to the receiver, with the hope of improving the chances of a call being accepted. A mechanism to achieve this is discussed in the next section.

7.3 Caller Cues

Presenting a caller cue to the receiver is reminiscent of the negotiated approach to call management. While there is no communication back and forth

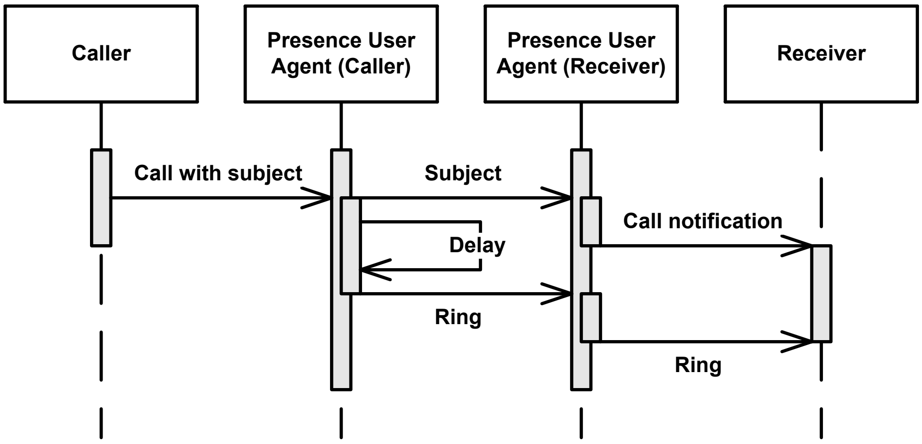

Figure 7.3: Presenting the Call Subject to the Receiver

it allows a two way sharing of context prior to the establishment of a call. In some cases a receiver may only be willing to accept certain calls of high importance.

The model allows such a cue to be transmitted in the form of the subject of the call. While it is also possible to use information such as the priority of the call this is a more subjective metric which may be harder to judge. The call subject allows easy evaluation by the receiver and also presents more useful information.

The subject must be contained in a short message indicating the reason or topic of the call. Much of the usefulness depends on the caller's accurate description of the subject. It is recommended to keep the message length below 140 characters to allow the additional option of transmission by SMS.

Definition 10 (Caller Cue). *A caller cue provides the subject of the call, as additional information, before the receiver answers the call.*

While the process adds extra time to the establishment of communication it gives callers a greater chance of reaching a busy receiver. Conversely it allows receivers to accept important calls while still being able to filter many disruptive ones. Figure 7.3 presents the flow of information.

The caller starts the communication process by choosing the receiver, entering the subject and calling the receiver. The call is then managed by the PUA which first sends the subject to the receiver PUA. A small delay is inserted before continuing the call, which give the receiver more chance to evaluate the incoming call's subject. The receiver is now notified of the incoming call with subject before the phone rings and can make the decision to accept or reject the call with the additional knowledge about the call.

The process aims to allow more calls through to the receiver when the calls are judged to be important. It is still up to the receiver to allow the call. Because the process may still be open to abuse by unscrupulous callers the receiver can also benefit from automatic call handling. This is discussed in the next section.

7.4 Receiver Call Profiles

The final approach to call management is receiver-oriented. This allows the receiver to decide beforehand how incoming calls will be handled. No further interaction is required as the PUA automatically evaluates each incoming call. While this seems convenient the danger exists that important calls will go unnoticed. Therefore this approach should be combined with the previous and not be used exclusively all the time.

Depending on the receiver's context a call profile can be created to automatically handle incoming calls. The model prescribes two profiles: blocking calls or changing the call notification to vibrate. Blocking calls will immediately reject any incoming calls regardless of the identity of the caller or the subject of the call. A vibrate call notification will treat the call as normal with the exception that a ring tone will not sound. Although dependent on the environment to allow the vibration to be noticed, this provides a less intrusive way to be notified of incoming calls.

Definition 11 (Call Profile). *A call profile can block or change the notification type of incoming calls.*

A receiver can either set an explicit call profile or allow it to be set by context changes. The same transformations used by the PUA to convert context to presence can be used for this purpose. Table 7.1 presents the default rules for call profiles.

Depending on the event and conditions several types of actions are possible. Actions are enriched from merely affecting presence by also allowing the call profile to be set automatically. By changing the mobile phone ring tone

Table 7.1: ECA Call Profile Rules

Event	Condition	Action
calendar event start		profile = vibrate
calendar event end		profile = normal
incoming call	call accepted	profile = block
outgoing call	call accepted	profile = block
call ended		profile = normal

to vibrate the user is protected from unwanted disturbances during scheduled calendar events. After an event has finished the ring tone can be switched back to normal. Additional calls can also be handled automatically when the receiver is already engaged in a conversation.

By combining all call management approaches the model provides a comprehensive approach to the research problem and uses the constructs in a unique manner.

7.5 Conclusions

Figure 7.4 summarizes the additional constructs and relationships defined in this chapter (new entities are indicated through shading). It extends the functionality already described in Section 6.5.

The model defined several additional functions of the PUA: fetching and displaying presence information, communicating the subject as a caller cue and implementing call profiles. As discussed in this chapter, the main goal of these functions is the use of presence to manage calls more effectively.

The process of call management is complex, but can be assisted with the intelligent use of presence information. This chapter has identified and implement three approaches to provide additional controls to manage calls. While not new in themselves the combination of approaches forms a unique and comprehensive solution to the research problem.

The final part of the model addresses the important topic of information privacy and authorizing access to presence. The sensitive nature of presence makes this a critical part of the model. The mechanisms needed to ensure privacy awareness are examined in the next chapter.

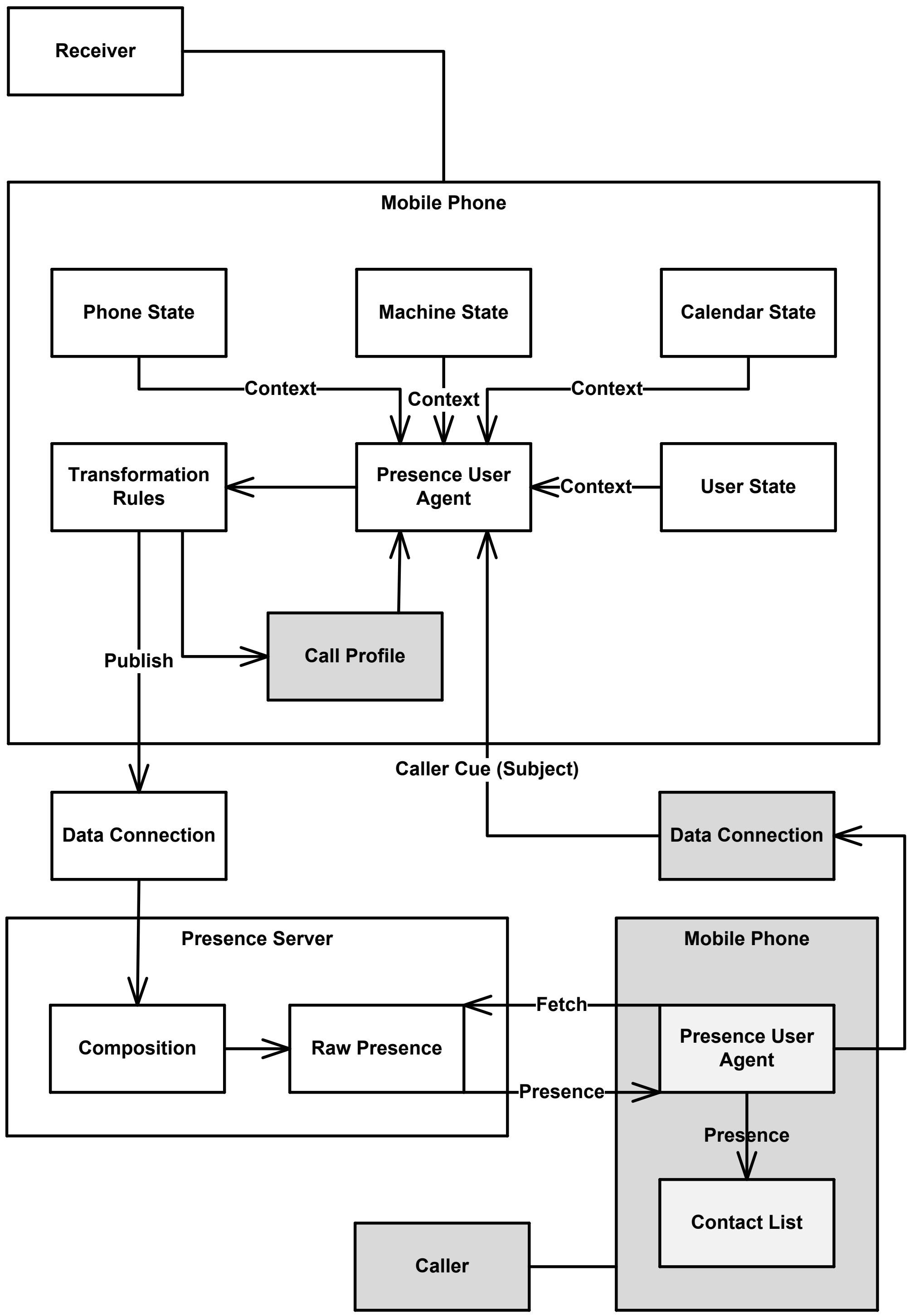

Figure 7.4: Model Summary

Chapter 8

Presence Authorization

Authorization is the most important part of the model and a key function in presence systems. In this context the term presence authorization means controlling access to presence information. This is done using authorization policies, also known as authorization rules, which specify what presence information can be given to which callers, and when.

An authorization policy is a framework which defines a set of rules for transmitting sensitive presence information, with the permission of the receiver, to callers. The document format needed to express these rules is also known as a presence authorization document. According to the primary objective stated in Chapter 1, the model aims to use established standards. In this regard the current chapter extends the common policy document format defined by Schulzrinne et al. (2007) using a combination of rules from Rosenberg (2007b) and unique social relationship groups.

The chapter starts by defining the constructs and interactions specific to this part of the model. It then describes the sequence of operations as implemented by the model. Next, the data model examines the parts of an authorization rule, with the model supporting extensions specifically for the mobile communications domain. The support of social relationship groups is discussed next, which contributes a unique extension by the model. Finally, a permission combining algorithm explains the processing of authorization rules.

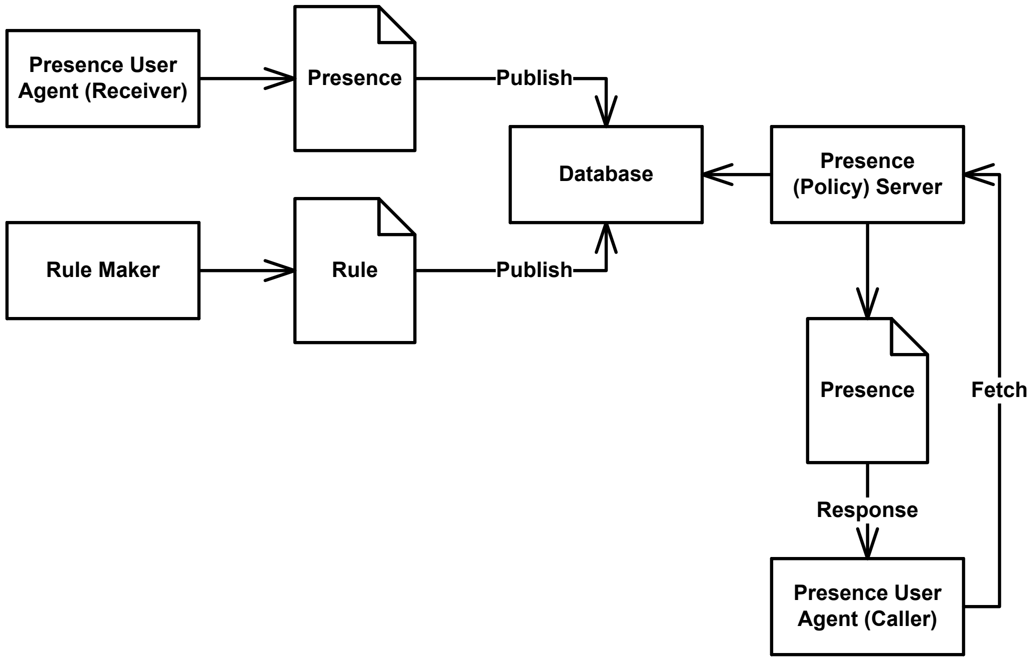

Figure 8.1: Authorization Policy Framework

8.1 Authorization Policy Framework

The authorization policy framework defines several entities which form the underlying model architecture. These entities are illustrated in Figure 8.1.

The architecture is built on the previously defined presence fetching mechanism in which a caller requests presence information about a receiver in order to judge the appropriateness of communication. The receiver is an entity known from previous chapters. This is the user about whom presence information has been requested. The receiver's presence user agent (PUA) publishes presence information as changes to the data occurs.

The rule maker is a new entity that creates the authorization rules which restrict access to presence information. The model defines the rule maker and receiver to be the same user. However, in other scenarios it is possible for these entities to be separate users, such as a personal assistant managing the communications of a boss. The rule maker publishes authorization rules to a database repository via the PUA. The model defines each rule to be representable as a row in a relational database. This allows efficient implementation which can leverage standard database optimization techniques.

Definition 12 (Rule Maker). *The rule maker is responsible for creating and maintaining authorization rules. The rule maker and receiver is the same user.*

The presence server, also known as the policy server when dealing with authorization policies has access to the database containing the presence information as well as the authorization rules. The presence server is responsible for checking incoming presence requests against authorization rules and returning the relevant presence notification to the caller.

Lastly, the caller entity is also known from previous chapters. This is the user who requests access to the presence information of the receiver. The model only allows read requests to be issued by callers. Now that the model architecture has been defined the next section moves on to describe the sequence of operations carried out by each entity.

8.2 Mode of Operation

In order for an authorization policy to be applied effectively the entities discussed in the previous section need to interact in a specific manner and sequence. This mode of operation is detailed in Figure 8.2.

The diagram shows that the receiver, who is also the rule maker in the model, should first create authorization rules which will protect the privacy of presence information. These rules are published to the database by the PUA. If this is not done initially the receiver may end up transmitting too much presence information, with a loss of privacy, or denying requests for presence information and negating the purpose of the model. In addition to the above rules the receiver will also publish regular presence information updates as context changes.

At any point in time the presence server may receive a request from a caller for the presence information of a receiver. This request must contain an authenticated identity for the caller which is used to search through the authorization rule set. The presence server retrieves the rules and presence information from the database and searches for all matching rules which apply to the caller. At this point other information, such as the location of the receiver, may also trigger extra rules to apply to the request. All matching authorization rules are combined according to an algorithm which is described in detail later in this chapter.

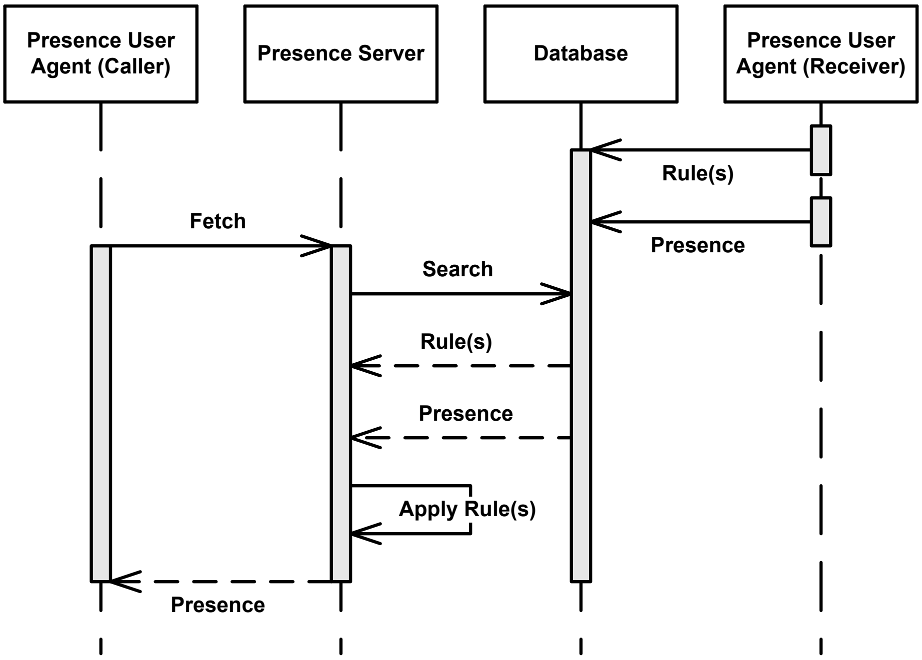

Figure 8.2: Mode of Operation

Finally the presence server uses the rule set to determine whether the caller is authorized to access the receiver's presence information, refusing the request if needed. The combined rules are applied to the presence information, resulting in a presence document which ensures receiver privacy based on the identity of the caller and current context of the receiver. Presence information may be filtered by removing or changing the level of detail of elements. The resulting presence information is returned to the caller.

In the mode of operation described above the presence server acts as a passive entity, waiting for a request before performing any processing and delivering a response. This mode of operation is also employed by other well-known protocols, such as HTTP and FTP. The model prescribes this mode, above an active push of presence information from the presence server to the caller, for the following reasons:

- It promotes simplicity. Message flow, and potentially cost, is reduced and there is no need for a caller to create additional rules to throttle pushed presence information.

- For the application of having availability cues before a call it is only necessary to have access to presence information at the time of the call, and not every time the receiver's context changes. Practice suggests that a particular combination of caller and receiver usually have a low frequency of calls over time.

Definition 13 (Passive Presence Operation). *The presence server shall use a passive mode of operation which waits for presence requests before sending information to callers.*

Having described the entities and their interaction it is now time to look at the data model. The next section describes the composition of authorization rules.

8.3 Data Model

Every user maintains an authorization policy, or rule set, consisting of zero or more authorization rules. The structure of this policy is defined by presence standards (Schulzrinne et al., 2007). Rules only provide permission to access presence information, rather than denying access. This has several consequences:

- The ordering of rules do not matter, making it much easier to update rule sets and to process them efficiently.
- Removing a rule can never result in more presence information being returned.
- To make a policy decision all rules need to be processed – the overall permission granted is the union of all rules.

A rule set can be stored at the presence server and conveyed from the rule maker to the presence server as a single document, in subsets or as individual rules. Each rules consists of three parts: conditions, actions and transformations. The actions and transformations part of a rule is also known as a permission. Figure 8.3 illustrates the data model of an authorization rule. Each part is discussed in more detail in the following sections.

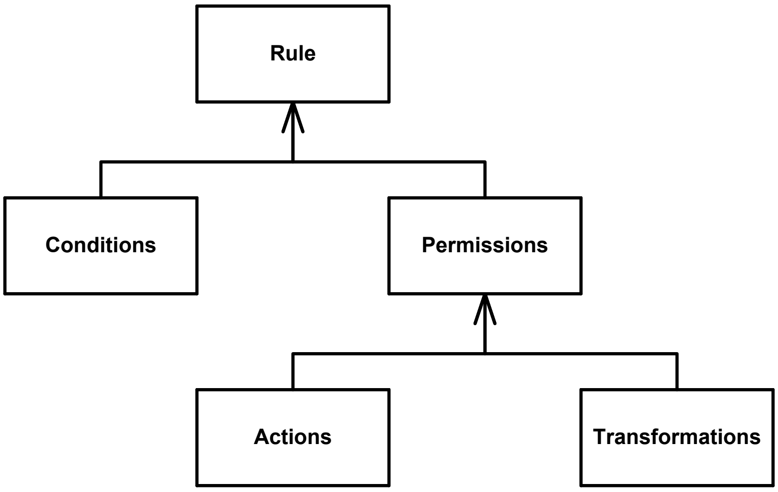

Figure 8.3: Structure of an Authorization Rule

8.3.1 Conditions

The conditions part of a rule is a set of expressions which determine whether the rule is applicable to a request for presence information. Each expression evaluates to either true or false. When a caller requests data about a receiver the presence server goes through the rule set and evaluates the expressions in the conditions part of each rule. If all the expressions in a rule evaluate to true the rule is applicable to the request.

Usually conditions are associated with the context of the request, but can also include external variables. The model prescribes the following conditions for application in the mobile communications domain:

- Matching a social relationship group defined by the receiver. The use of groups is intended to simplify the use of authorization rules by applying permissions to a group of users rather than individually. This is discussed further in Section 8.4.

- The authenticated identity of a single user. This is intended for users who are known to the receiver but do not form part of any group.

- Any authenticated identity. This is a catch-all for other callers, known or unknown to the receiver. If multiple identity conditions are specified for a single rule the result is combined using a logical 'or' operator. It is important to note that in the absence of any identity a rule will apply to any user, authenticated or not.

- The current state, or sphere, as specified by the receiver. If more than one sphere is specified the result is combined using a logical 'or' operator. The sphere only matches if the receiver is currently in the specified state.

- A validity period expressed by a starting and ending time. This condition is true if the current time falls between the specified range, with multiple ranges again being combined using a logical 'or' operator. This allows rules to be invalidated automatically without further interaction between the rule maker and presence server.

Definition 14 (Authorization Rule Conditions). *Authorization rules can be matched on social relationship groups, single identities, sphere or validity period conditions.*

In summary, a rule may have a number of conditions which need to be met before the permission parts of a rule are executed. While conditions are the 'if' part of rules, actions and transformations are the 'then' part that determine which operations the presence server must execute before returning presence information to the caller.

8.3.2 Actions

If a rule is applicable the actions specify the operations that the presence server must execute. A typical action in response to a request is to allow the request to proceed or to block it. The model prescribes the following actions for handling incoming requests:

- A block action tells the presence server to reject the request for presence information. It has an integer value of zero which is also the default value. It is not strictly necessary for a rule to include an explicit block action, since the default in the absence of any action will be block. However, it is included for completeness.

- A polite-block action indicates that the presence server must reject the request with a response that the receiver is unavailable. It has a value of twenty.

- An allow action tells the server to allow the request, proceed with further operations and respond with the appropriate presence information. It has a value of thirty.

Definition 15 (Authorization Rule Actions). *Authorization rules can invoke blocking, polite-block or allow actions for a request.*

The numeric values assigned to actions are used to determine the appropriate action when combining rules. This will be explained in more detail in Section 8.5. The model uses the actions part of a rule to determine how a presence request should be processed. This differs from transformations in that no data is modified in the process.

8.3.3 Transformations

Assuming the actions allow a request to proceed, the transformations part of a rule specifies how the receiver's presence information is presented to the caller. Transformation specify presence server operations that can lead to a modification of the data requested by the caller. This is especially useful to reduce the granularity of information, such as taking exact location coordinates and broadening it to the city level.

Every transformation is a positive permission, and specifies an item of information that is allowed for transmission, by the receiver. When the presence server processes a request it creates a union of all the transformations across rules that match. For this union the data type plays a role, as specified by the combining algorithm in Section 8.5.

The model prescribes the following transformations for presence information:

- Providing the current activity of the receiver. If the transformation value is true this data is returned to the caller if present, if false it is removed.

- Providing the current sphere of the receiver. Similar to activity, if the transformation value is true this data is returned to the caller if present, if false it is removed.

- Providing the last usage state of the mobile phone. This is an enumerated integer type which can have the following values: false or zero indicates that the element is removed, bare or ten returns basic information without any idle times, threshold or twenty returns basic information with an idle-threshold and full or thirty includes all information.
- Providing a receiver specified note containing text. If the transformation value is true this data is returned to the caller if present, if false it is removed.
- Providing all presence attributes. This is effectively a macro which expands to provide the maximum permissions of the data elements above.

Definition 16 (Authorization Rule Transformations). *Authorization rules can apply activity, sphere, usage state and note information transformations.*

It can be seen that transformations specify the information returned if any authorization rules match conditions and the actions allow it to proceed. The model does not prescribe all rules defined in standards documents as many of them are not applicable in the mobile communications domain. However, since it is intended that users only manipulate their presence authorization rules from a single client this is not a problem, because a user will not deal with different subsets of data.

It remains to explain how multiple matching rules are combined by the presence server. However, before that the next section first discusses the social relationships group condition in more detail.

8.4 Social Relationship Groups

The users in a contact list often express a social relationship with the user in charge of it. Relationships affect our knowledge of another user as well as our availability for communication with a user. Relationships also influence our perception of privacy and will determine how much information we are willing to share.

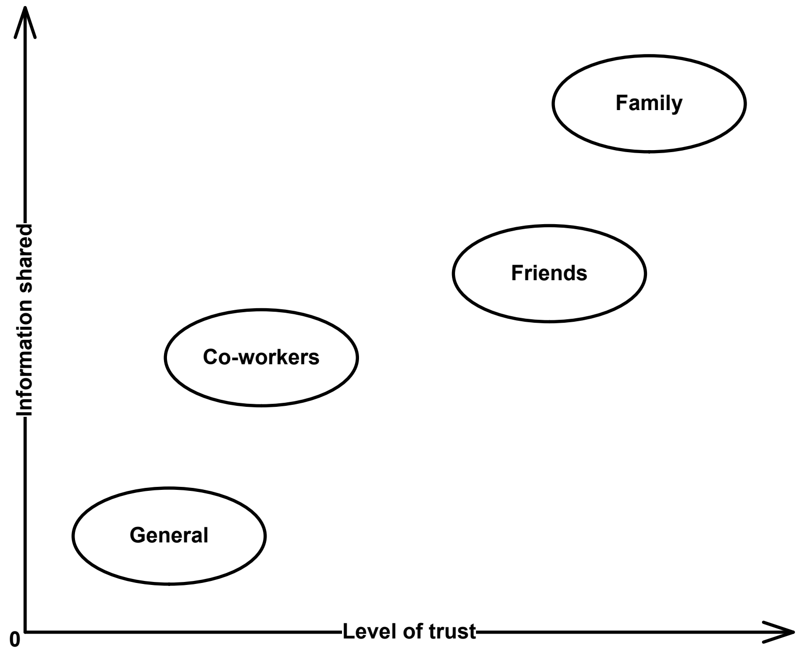

Figure 8.4: Social Relationship Groups

A receiver can have a relationship with zero or more callers and vice versa. It is acknowledged that many types of relationships exist in the real world and it is left up to users to define the relationships which apply to them. However, it is common to identify general groups of relationships which apply to almost all users. As example the model uses four groups: general, co-workers, friends and family. This is illustrated in Figure 8.4.

In general each group along the horizontal axis, from left to right, indicates a greater level of trust. In turn this should lead to an increasing willingness to share presence information, illustrated on the vertical axis from bottom to top. The figure presents an illustrative example of how a user may classify relationships. However, the model acknowledges that group dynamics can be complex and completely different from one user to the next. Despite of this the underlying concept of classifying users and relationships into groups remains valid.

Definition 17 (Social Relationship Groups). *Callers can be classified into a general, co-workers, friends or family group depending on their social relationship with the receiver.*

The general group is a catch-all for callers outside the other groups or those unknown to the receiver. Co-workers often share information during working hours and this group may be enforced by company policy. The friends group classify callers in a close relationship, even after working hours. The family group is intended for relatives. Groups provide an easy way to classify the users in a contact list and to manage presence information access requests for a large amount of callers. It is intuitive and usually takes place on a subconscious level, without explicitly stating the relationship. It may also be possible to add groups based on mistrust, such as people whom a user dislikes but needs to share information with. This extension adds a further dimension to the relationship graph.

The model prescribes that callers be assigned into one or more groups. This is based on the fact that relationships are dynamic and could change according to criteria such as time, location and activity, to name a few. This makes it possible for two users to be in multiple groups relevant to each other. The model also acknowledges that a relationship may change over time. It is the responsibility of the user to maintain the accuracy of such (dynamically changing) groups.

Definition 18 (Group Membership). *Callers must be assigned to at least one social relationship group. The receiver is responsible for keeping groups up to date.*

The existing standards which define authorization policies do not allow for user defined groups as part of the conditions of an authorization rule. Thus the model uniquely extends these rules by adding the group condition. This provides a mechanism by which a receiver can define a privacy list by groups of contacts and apply presence authorization accordingly.

The model defines a new type of identity condition which represents a group. This element is called a *list* to conform to existing standards (Rosenberg, 2007a). A list must have a unique name to identify it and zero or more entries which represent the users who belong to it. The entries can be specified locally or in a separate document, also known as a resource list.

If a resource list is used the location of the document must be given. Unfortunately existing standards define a resource list's name element as optional which conflicts with the model's mandatory and unique name requirement (Rosenberg, 2007a). To overcome this problem the model proposes the following: if an authorization rule contains a link to a list and the specified

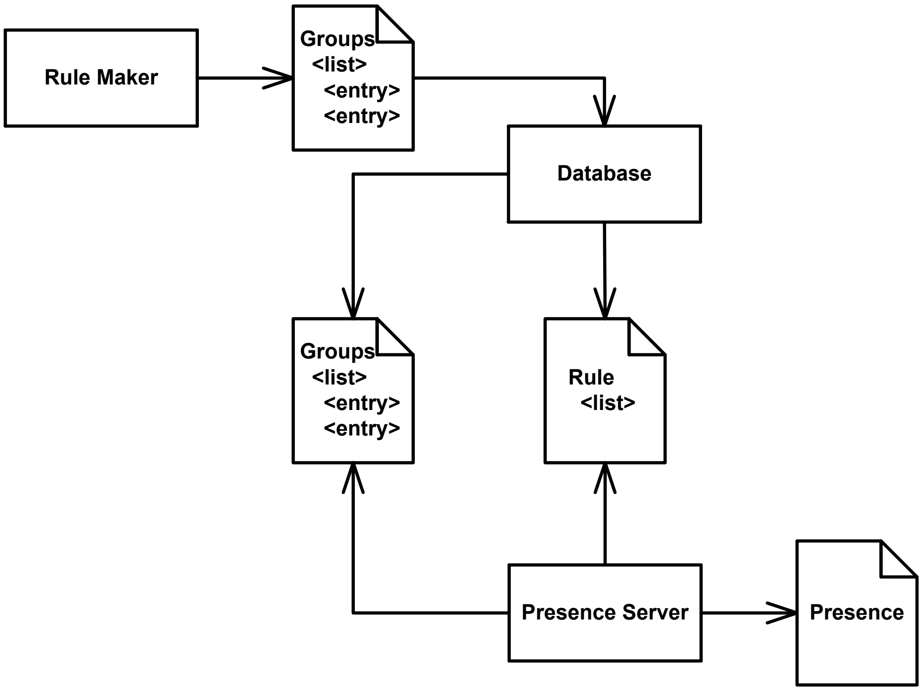

Figure 8.5: Groups Extension to the Presence Authorization Framework

name attribute does not exist in the resource list the server shall ignore that identity condition. This may result in the absence of any identity condition in which case other conditions will apply to any user.

Definition 19 (Resource List). *A list element shall be used to represent group information inside standard resource lists.*

A rule maker would create and maintain a group document which is stored alongside presence information and authorization rules. When examining a rule set the presence server would identify whether a group (list element) exists as an identity condition. If a group does exist the presence server imports all the corresponding callers as specified in the document, before applying the necessary actions and transformations to the presence information. This is illustrated in Figure 8.5. Appendix C contains an abbreviated XML Schema to define this structure.

Having defined the use of groups as authorization rule conditions the next section concludes this chapter by explaining how authorization rules are processed using a combining algorithm.

8.5 Presence Processing

A receiver may create multiple rules to protect the privacy of presence information. To determine the relevant presence for a caller these rules need to be evaluated and combined before any data is returned. When a presence server receives a request for presence information it is matched against the rule set. A rule matches if all conditions of a rule evaluate to true. All rules where the conditions match the request form the matching rule set. The actions and transformations in the matching rule set are combined using an algorithm called the combining rules.

Each type of action or transformation is combined across all matching rules. Each action or transformation is also combined separately and independently. The combining rules generate a combined permission. The combining rules depend only on the data type: for boolean permissions, the resulting permission is true if and only if at least one permission in the matching rule set has a value of true, and false otherwise. For integers the resulting permission is the maximum value across the values in the matching set of rules.

In the previous sections the model defined the following data types for the prescribed actions and transformations:

- Boolean types: the provide activity, sphere and note transformations. If an element's value is true it will be reported to the caller.

- Integer types: the request actions and the last usage transformations. The element corresponding to the maximum integer value will be reported to the caller.

An example of a set of rules is shown in Table 8.1. Every rule is identified by a unique 'id' value. Next each rule has several conditions including identity, sphere and a validity period. A single request action is present which will determine whether the request is allowed. Finally two transformations are shown for the receiver's activity and usage state. In addition let's assume validity periods specified as:

- A1 = 2010-12-24 08:00:00

- A2 = 2010-12-24 17:00:00

- B1 = 2010-12-24 18:00:00

Table 8.1: Authorization Rule Set

Conditions					**Actions**	**Transformations**	
Id	**Identity**	**Sphere**	**From**	**To**	**Request**	**Activity**	**Usage**
1	Family	Work	A1	A2	30	false	20
2	Friends	Work	A1	A2	0	false	10
3	Family	Work	A1	A2	30	true	30
4	Family	Home	B1	B2	30	false	0

- B2 = 2010-12-24 22:00:00

As an example of the intended use of a rule, rule 1 applies to callers in the family group while the receiver's sphere is set to work and the time is between ranges A1 and A2. In this case the actions would allow a request and the transformations would provide no activity information and basic usage state information with an idle-threshold.

To illustrate the combining of rules let us assume that an entity in the family group requests presence information before making a call. The presence server receives the request between the time ranges A1 and A2. The example also assumes that the receiver's current sphere value is set to work.

First it is necessary to determine which rules match by evaluating the conditions part of each rule. Rule 1 matches since all conditions evaluate to true. Rule 2 does not match since the identity only applies to the friends group. Rule 3 matches since all conditions evaluate to true. Rule 4 does not match since the sphere and validity periods both evaluate to false.

Next we evaluate the actions and transformations of rules 1 and 3 to determine the combined permissions. Each column is treated independently. The combined value of the request actions is set to thirty, the maximum value between the two rules, which means the request will be allowed. The combined value of the provide activity transformation is set to true which means that activity information will be available. The combined value of the usage state transformation is set to 30, the maximum value between the two rules, which means that full information will be included. Table 8.2 summarizes the resulting presence permissions for this request.

From Table 8.1 it can be seen that it is also possible to use individual identities as rule conditions. However, the model does not recommend mixing individual identities with groups as this can complicate the user's

Table 8.2: Resulting Presence Permissions

Request	Activity	Usage
30	true	30

understanding of permissions and may lead to less privacy. The emphasis is on rule simplicity and clarity, which reduces misunderstandings and false assumptions as to when rules apply.

Definition 20 (Rule Clarity). *Authorization rules should remain simple and clear by not mixing individual and group identities.*

As the output of this research is an abstract model, the instantiation of authorization rules is not of primary concern here. However, it is acknowledged that this is a challenging aspect to implement, and could be difficult for the user to understand and manage. In the next chapter a possible instantiation is provided. This concludes the definition of the constructs and relationships which form the model. The next section presents a brief overview of the complete model.

8.6 Model Summary

Figure 8.6 presents a final summary of the model's constructs and relationships as described across the past three chapters. New entities are shown using shading.

The first part of the model described the publication of presence. To achieve this it was necessary to define what information, or context, is considered important about the user. Once context has been identified the next step was to detail how this context can be captured as presence information. These constructs allow a receiver to present an availability state to potential callers.

The second part of the model dealt with call management and the use of presence. Three approaches were described relating to both the caller and receiver. These approaches allow the receiver to indicate availability for communication, the caller to presence call cues and the automatic handling of calls for the receiver. These mechanisms allow the caller to make an informed

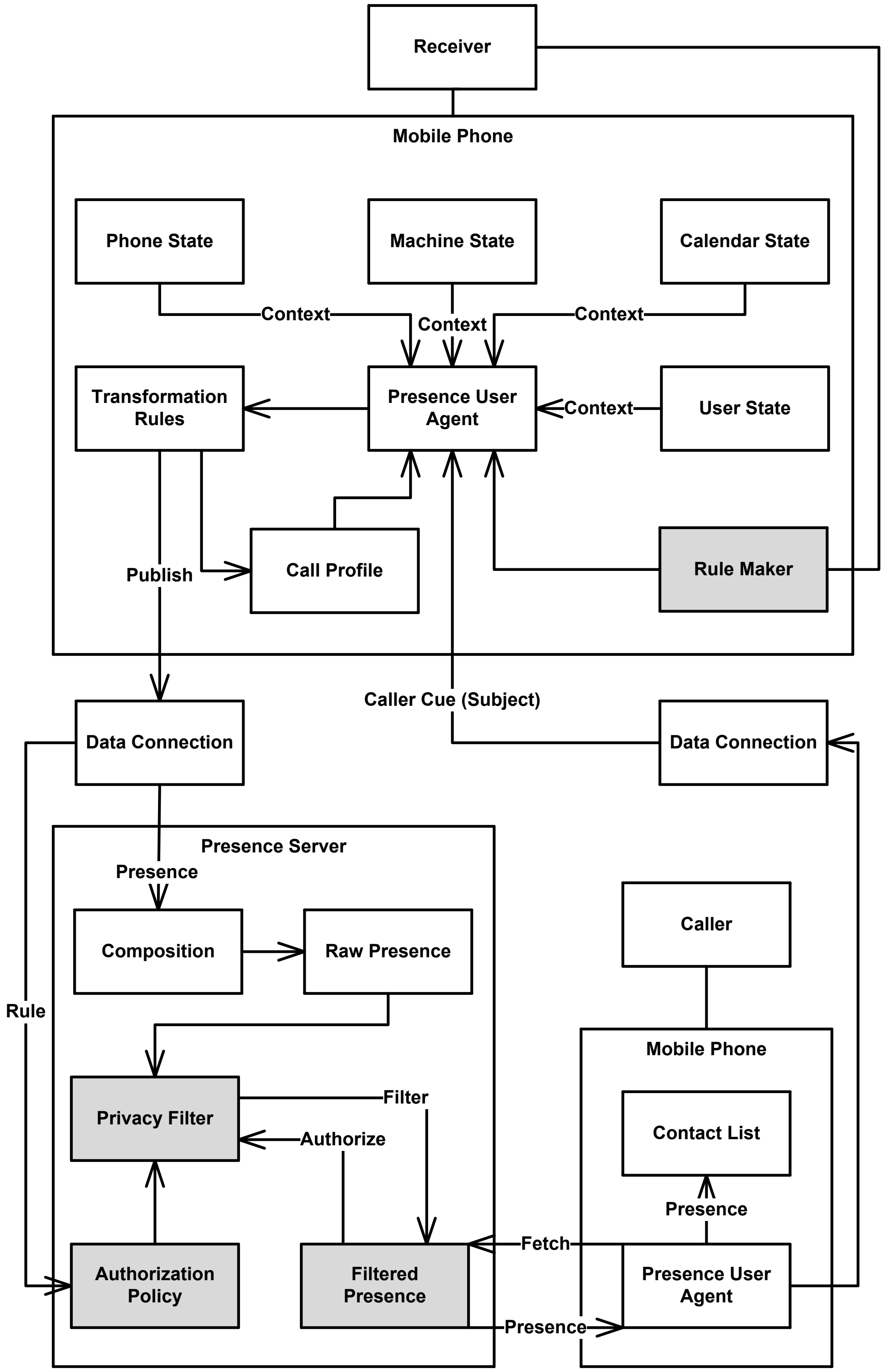

Figure 8.6: Model Summary

decision about whether to proceed with an intended call. Such knowledge can not only minimize receiver disruptions, but also save the caller from fruitless attempts to contact an unreachable target.

The third part of the model discussed how the receiver can maintain privacy of presence information. Requests for information can come from multiple sources, not all of them trusted. Thus the use of groups based on social relationships is prescribed. This allows for authorization rules which closely relate to trust relationships in the real world. Authorization rules allow requests to be evaluated according to several conditions and can filter the final presence document before it is returned to a caller.

In addition to the model summary above, Table 8.3 provides the complete set of definitions defined by the model.

8.7 Conclusions

Achieving privacy is one of the most challenging exercises when dealing with any kind of information. It is an even more critical factor when dealing with personal information such as presence. Existing standards have laid the foundation for an authorization policy framework which can be used to effectively address the privacy issues.

However, not all users are technically advanced, especially when dealing with the mobile communications population. Thus there is a need for simplicity with the aim to make authorization rules understandable to a wide audience. The model contributes to this area by adding the concept of social relationship groups, something intuitive to most users, and allowing rules to be linked to such groups.

Part II defined the complete model and explained the mechanisms underlying it. Chapter 5 provided the foundation for the next chapters by defining the constructs and relationships used in the model. Thereafter Chapters 6, 7 and 8 defined the complete model, with each chapter building on the previous.

In Part III the model will be evaluated. In Chapter 9 a prototype based on the model will be examined. This prototype will provide an instantiation of the underlying constructs to show the practicality of the model. Thereafter Chapter 10 will examine how the model can be extended to next generation mobile networks.

Table 8.3: Model Definitions

	Definition
1	**Context Sources.** Context is obtained from a user's environment and situation. Such data shall be obtained through operating system software which reports data from hardware, external sensors and personal information.
2	**Minimum Set Of Context.** The minimum set of context consists of location, activity and device usage. The presence user agent is responsible for discovering and collecting context information.
3	**Presence Elements.** The basic status, activities, sphere, user-input and optional note elements are used to indicate user presence.
4	**Context Transformation.** An ECA pattern is implemented by the presence user agent to transform context into presence information.
5	**Presence Composition.** Presence composition uses the published presence in its current format. Each user has a single presence user agent which maintains the complete presence state.
6	**User Identity.** A user shall be identified by a phone number following the tel URI scheme. This phone number must be globally unique. The model assumes that each user is associated with a single phone number and that no other users share that number.
7	**Unknown Identity.** In the absence of a phone number the term 'unknown' shall be substituted as the identity of a caller.
8	**Single Connections.** A user can only be involved in a single connection (call) at a time and shall be unavailable to other users until that call is completed. A user can act as a caller or receiver.
9	**Presence Polling.** Polling for presence information is performed by the presence user agent only when needed. Any type of network connection can be used for this purpose.
10	**Caller Cue.** A caller cue provides the subject of the call, as additional information, before the receiver answers the call.
11	**Call Profile.** A call profile can block or change the notification type of incoming calls.
12	**Rule Maker.** The rule maker is responsible for creating and maintaining authorization rules. The rule maker and receiver is the same user.
13	**Passive Presence Operation.** The presence server shall use a passive mode of operation which waits for presence requests before sending information to callers.
14	**Authorization Rule Conditions.** Authorization rules can be matched on social relationship groups, single identities, sphere or validity period conditions.
15	**Authorization Rule Actions.** Authorization rules can invoke blocking, polite-block or allow actions for a request.
16	**Authorization Rule Transformations.** Authorization rules can apply activity, sphere, usage state and note information transformations.
17	**Social Relationship Groups.** Callers can be classified into a general, co-workers, friends or family group depending on their social relationship with the receiver.
18	**Group Membership.** Callers must be assigned to at least one social relationship group. The receiver is responsible for keeping groups up to date.
19	**Resource List.** A list element shall be used to represent group information inside standard resource lists.
20	**Rule Clarity.** Authorization rules should remain simple and clear by not mixing individual and group identities.

Part III

Evaluation

Chapter 9

Instantiation

This chapter presents an implementation of the model as an additional output of the design science research approach. The instantiation is intended to show the feasibility of the model and allow limited evaluation of its practical use.

The instantiation includes all the aspects comprising the model and is based on existing presence standards. Design science allows for an instantiation to precede the complete understanding of the underlying model. In this way it can help to show problematic or incomplete areas in the model. The instantiation in this chapter has been used in this way, thus being influenced by but also influencing the model during its development.

The chapter begins by discussing the chosen development platform. Thereafter the use of context and derived presence is reviewed, along with the communications architecture. The method of maintaining presence authorization rules is presented and how this influences requests for presence information. Finally, integration of presence with the phone's address book is reviewed.

9.1 Development Platform

The broad range of platforms available for mobile phones makes it impossible to develop a standard application across platforms. Taking this into account it was decided to choose a platform based on other factors, such as ease of development, hardware and software support and familiarity. The Windows Mobile 5.0 platform from Microsoft (2005) provided all these factors.

Windows Mobile 5.0 provides a platform on which developers can build rich mobile applications. An advanced application programming interface (API) allows full use of all the latest device features, while Microsoft's Visual

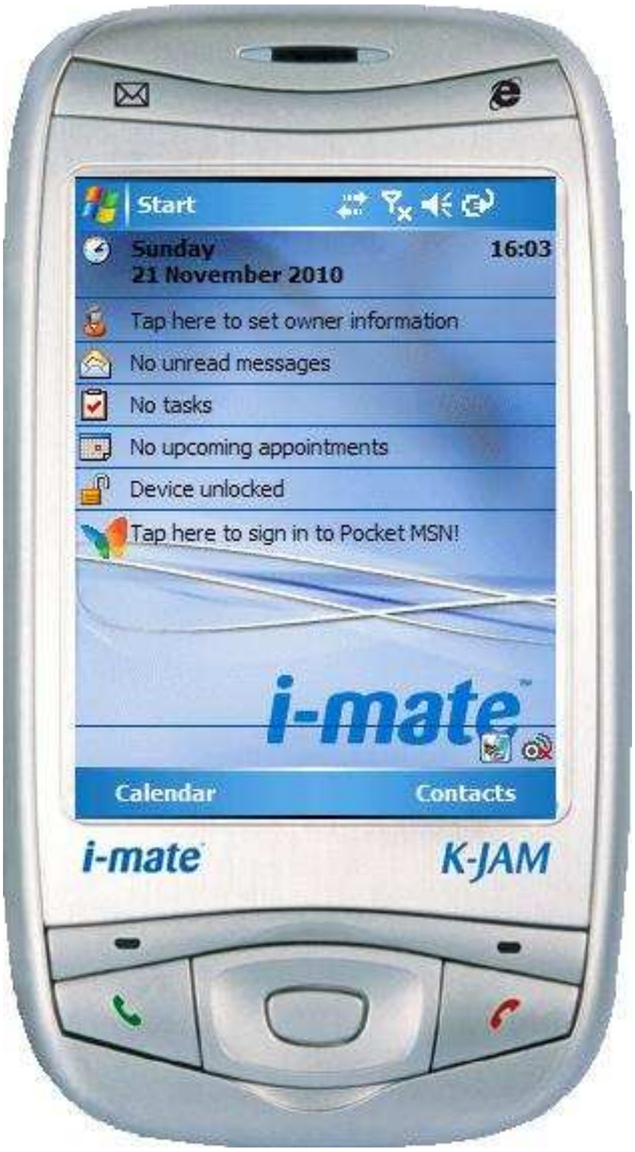

Figure 9.1: I-Mate K-JAM Mobile Phone

Studio development environment allows rapid application development. In addition, the .NET Compact Framework provides developers with standard functionality which speeds up the development cycle.

The physical device chosen was the I-Mate K-JAM, also known as the HTC Wizard 200. This device supports the Windows Mobile 5.0 platform while also exposing several useful hardware features, such as touch screen, slide out keyboard and mobile connectivity. An example is shown in Figure 9.1.

Lastly, two external developer libraries were used to provide standards compliant presence functionality. The SIP.NET library from Independentsoft

(2010) provides basic SIP functions which can be used to support a SIMPLE presence implementation. In addition the MatriX XMPP SDK from AG Software (2010) provides a comprehensive library for the XMPP protocol. Both these libraries were used as trial versions.

9.2 Presence Management

One of the major functions of the instantiation is detecting context changes in the user's environment and updating the presence state. This can be complex as it requires interacting with various low-level hardware sources as well as user data. Luckily Windows Mobile 5.0 includes an API which simplifies this interaction, known as the State and Notifications Broker API (SNAPI).

SNAPI provides notifications whenever the device state changes which allows an application to respond to events, such as context changes. In addition SNAPI also provides extra information about a state change when applicable. In support of the model this is used to detect user task status and device usage.

9.2.1 Monitoring Context Sources

The instantiation monitors several context sources: device location, calendar appointment, call state and a user-defined note.

Unfortunately the K-JAM does not have GPS hardware for determining location. However, using low-level interaction with the device hardware it was possible to determine location by the cell tower ID the phone is connected to. This information is not exposed by the SNAPI and thus polling is used every 30 seconds to determine if a change in location has occurred.

By accessing the radio interface layer it was possible to retrieve the following information about a user's location: cell ID, location area code, mobile country code and mobile network code. While this information may be hard to understand by itself it allows the user to attach a descriptive name to a location which can be recalled whenever the user reaches the same location again. Unfortunately the limitations of mobile networks need to be kept in mind – a user may be connected to the same cell tower from a fairly broad area and thus an exact position can not always be assured. However, in cities cell towers usually cover smaller areas and it is possible to distinguish between general location such as work and home.

User activity is monitored by checking for calendar event changes. To operate correctly it requires that the user set the subject of each calendar appointment with applicable start and end times. This information is retrieved as a transition between events occur and is used in conjunction with ECA rules to set the user's presence. Device usage is also reported through the SNAPI and is converted to presence through a similar set of ECA rules.

The following code monitors for a change in calendar state:

```
SystemState calendarState;
...
calendarState = new SystemState(SystemProperty.CalendarHomeScreenAppointment);
calendarState.Changed += new ChangeEventHandler(calendarState_Changed);
```

This shows how SNAPI allows a subscription to state changes. By using the above code a function can be called every time the user's home screen appointment changes. A similar format was used to determine call state changes.

Finally, the user was allowed to enter a free text message indicating which is used as an additional presence indicator. In conjunction with the above context sources this provides a caller with much more information than would normally be available before making a call.

9.2.2 Context to Presence

To allow a standardized view all context information was converted into presence. The SIMPLE presence standard was used as the main implementation, but other standards are equally viable. Figure 9.2 shows a class diagram for the presence structure used.

Presence consists of a basic status, activity which includes a from and until time, sphere, user-input and a note. Where applicable the presence were also defined.

The ECA rules to convert context to presence used SNAPI events, if/else conditions and assignment actions, as shown in the code sample below.

```
void callState_Changed(object sender, ChangeEventArgs args)
{
    if ((bool)SystemState.GetValue(SystemProperty.PhoneCallTalking))
    {
        Presence.UserInput = Presence.UserInputValues.Active.ToString();
        Presence.BasicStatus = Presence.BasicStatusValues.Closed.ToString();
    }
```

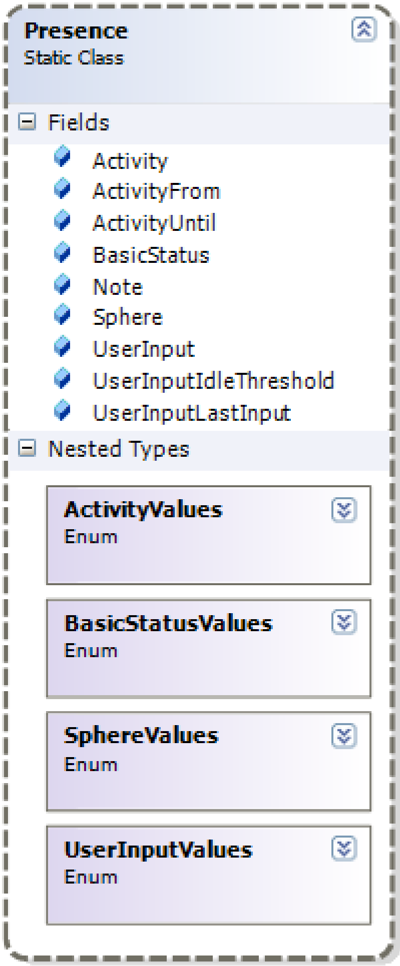

Figure 9.2: Presence Class Diagram

```
    else
    {
        Presence.UserInput = Presence.UserInputValues.Idle.ToString();
        Presence.BasicStatus = Presence.BasicStatusValues.Open.ToString();
    }
}
```

A comparison of state values allows the instantiation to determine when an active call is in progress. In these cases the user-input state is set to "active" and the basic status is set to "closed", indicating that the user is unavailable for further communication.

9.2.3 Publishing Presence

Communication is done using standard presence protocols. Messages are sent from the presence user agent on the mobile phone to a presence server located on the network. These messages need to be formatted appropriately and sent across a valid data connection.

For sending SIMPLE presence messages the SIP.NET library was used. This library support a basic set of SIP communication methods. Because the model uses rich presence extensions the library had to be extended to support this functionality. An example of a presence update is given below.

```
<xml version="1.0" encoding="UTF-8"?>
   <presence xmlns="urn:ietf:params:xml:ns:pidf"
     xmlns:dm="urn:ietf:params:xml:ns:pidf:data-model"
     xmlns:rpid="urn:ietf:params:xml:ns:pidf:rpid"
     entity="pres:alice@example.com">

  <tuple>
    <status>
      <basic>closed</basic>
    </status>
  </tuple>

  <dm:device>
    <rpid:user-input idle-threshold="600"
      last-input="2011-01-01T11:30:00+05:00">idle</rpid:user-input>
  </dm:device>

  <dm:person>
    <rpid:activities from="2011-01-01T12:00:00+05:00"
       until="2011-01-01T13:00:00+05:00">
       <rpid:appointment/>
    </rpid:activities>
    <rpid:sphere>work</rpid:sphere>
  </dm:person>

  <note>Please don't disturb!</note>

</presence>
```

The example shows an update for user Alice with a basic status of closed. One can see that her last device usage was on 2011-01-01 at 11:30. Her current activity is set to appointment with a sphere of work. She has also included a note. From all this one can deduce that now would not be a good time to call her.

With mobile data communication speeds increasing and costs dropping the type of communications link is becoming less of an issue. However, it is possible to use various methods of network connectivity. WLAN provides always on, high-speed connectivity but is usually confined to a specific location. However, in the future a ubiquitous WLAN connection is very likely which will provide convenient access to data. As the K-JAM support WLAN communication this method is a possible configuration for the instantiation.

Other methods of communication include using the mobile network itself. Current 3G networks provide high-speed access with only a slight delay while connecting to the network. The benefit to the user is that wherever voice communication is possible data transmission is supported as well, thus supporting the user to always maintain an accurate presence state.

On the network it is also possible to use text messages (SMS). This allows a caller and receiver to communicate with each other directly, with the presence server and presence user agent located on the same device. This provides additional privacy safeguards, as the user is fully in control of data storage. However, SMS messages can be insecure, unreliable and have length restrictions which limit the use of standard protocols.

Another possible network technology is unstructured supplementary service data (USSD) which allows the creation of a session-based communications channel between the subscriber and a service on the network. It is supported by all mobile phones and messages have a length restriction similar to SMS messages. USSD has several advantages as a bearer technology. Firstly, messages can be initiated both out of or during a call. This makes simultaneous voice and data communications possible. Secondly, messages are sent directly to the receiver allowing an instant response. Thirdly, USSD services on the home network are accessible while roaming in another country, and unlike SMS there are no charges from roaming partners (Henry-Labordère and Jonack, 2004, p. 191).

Finally, in a controlled environment a device may also be physically connected to a computer using the USB port. This was also a commonly used method during the development of the instantiation as it saves data transfer costs across the mobile network.

As the model prescribes a complete presence state to be sent with each change presence composition on the server consisted of replacing the previous presence information. This is an easy approach but is not optimal if a user is paying for data transmission. In such a case it would be better to send only the presence information that has changed.

9.3 Presence for Call Management

The goal of the model is to provide call management features to enable more efficient communication. The next sections will examine how the instantiation achieved this, bearing in mind that functionality was not always available

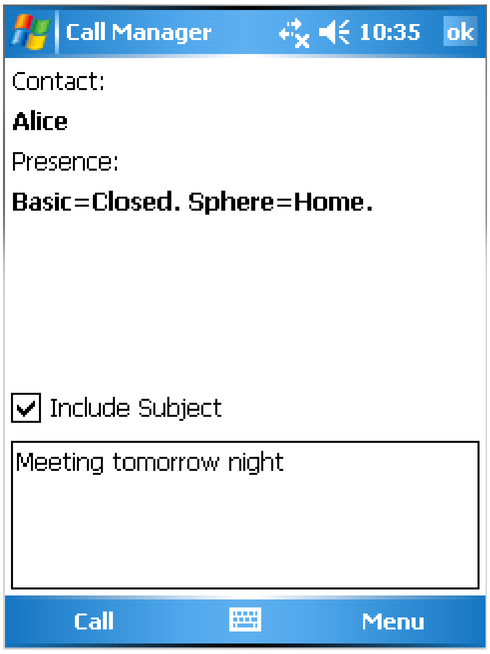

Figure 9.3: Adding a Subject to a Call

through managed code.

9.3.1 Indicating Receiver Availability

The model's main focus is on indicating receiver availability. In this regard the instantiation uses the phone's built-in contact list to select the contact to call. Thereafter available presence information is retrieved from the presence server and returned to the caller presence user agent (PUA). Information is displayed to the caller as shown in Figure 9.3.

From the information given the caller may opt to present an additional cue to the receiver in the form of the subject of the call.

9.3.2 Presenting Caller Cues

To increase the chance of a busy receiver accepting a call the caller can add a subject for the call. This is presented to the receiver immediately before or on the call. The receiver can then use this additional information to decide on an action. Figure 9.3 shows the caller interface.

After choosing a contact the caller sees any available presence information. If the caller wants a subject can be added before making a call to that receiver. The subject is sent to the receiver as an instant message and inter-

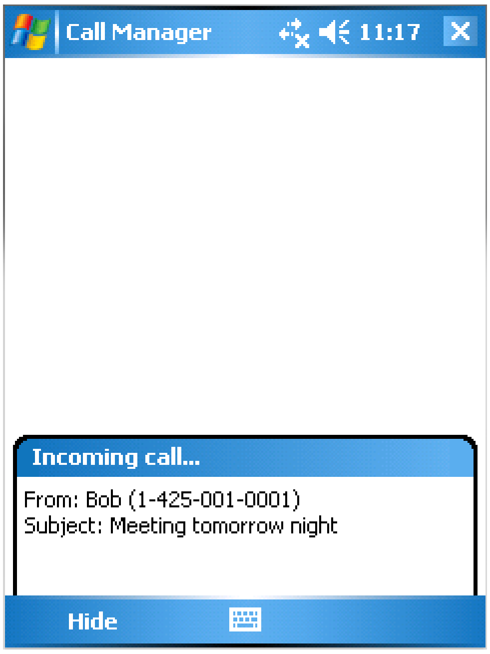

Figure 9.4: Receiving a Call Subject

cepted by the presence user agent. The receiver will receive a notification as indicated in Figure 9.4.

An informational message is displayed to the receiver indicating the caller name, number and subject of the incoming call. The receiver can still choose to ignore or answer the call.

9.3.3 Call Profiles

Changes in context can also be used to enable a device profile, such as switching the phone to silent mode. The receiver can specify a unique set of presence or individual presence data to enable a profile. The instantiation allowed the use of vibrate mode and automatic call rejection as possible options.

Vibrate mode is enabled through calls to system libraries. The following code sample shows the function to change the ring type to vibrate.

```
public static void EnableVibrate()
{
    SoundFileInfo sfi = new SoundFileInfo();
    sfi.sstType = SoundType.Vibrate;

    uint ret = SndSetSound(SoundEvent.All, sfi, true);
}
```

The function makes use of native functions embedded in the operating system. Specifically the "SoundType" value allows switching between normal, vibrate and silent ring types.

Call rejection can be achieved in several ways. It is possible to use the telephony API which also provides the most control over the call flow. However it is simpler to simulate disconnecting a call, and this was the approach used in the prototpe. The following code shows how this is achieved.

```
const int VK_F4 = 0x73;
const int KEYEVENTF_KEYUP = 0x0002;

[DllImport("coredll.dll", SetLastError = true)]
private static extern void keybd_event(byte bVk, byte bScan,
    int dwFlags, int dwExtraInfo);

public static void CancelCall()
{
    keybd_event(VK_F4, 0, 0, 0);
    keybd_event(VK_F4, 0, KEYEVENTF_KEYUP, 0);
}
```

Again use is made of a native operating system function. It simulates the pressing of the call ended key, which effectively disconnects the call. This code can then be called from the appropriate SNAPI event notification for incoming calls, as shown below.

```
SystemState incomingCall;
...
incomingCall = new SystemState(SystemProperty.PhoneIncomingCall);
incomingCall.Changed += new ChangeEventHandler(incomingCall_Changed);
...
void incomingCall_Changed(object sender, ChangeEventArgs args)
{
    if (cancelCalls == true)
    {
        Functions.CancelCall();
    }
}
```

The "PhoneIncomingCall" system state value indicates whether there is an incoming (ringing) call. If this is the case a boolean variable, indicating whether the call profile is active, is checked. If active the call is cancelled immediately without further interaction from the user.

9.4 Presence Authorization

One of the key functions of the model is to protect the privacy of presence information. Several interactions take place to ensure that the privacy of presence information is maintained. This includes allowing the user to maintain an authorization policy, the PUA transmitting the policy to the presence server, storing the data and processing incoming presence requests according to authorization rules. The next sections will explore these interactions in more detail.

9.4.1 Data Model

The instantiation uses the data model depicted in Figure 9.5. The Presence tables contains the user's current presence state and is updated every time the PUA performs a publish action. The AuthorizationRules contain the rules for protecting the privacy of presence information and can be updated by the user. The User, Group and SocialRelationship tables form a whole which classifies incoming requests according to the relationship between the users.

Authorization rules form a critical part of the instantiation. As each rule is a positive grant of information each additional rule may expose more information about the receiver. Figure 9.6 illustrates the data containing in the AuthorizationRules table.

Each row in the table presents a unique authorization rule, identified by the "uid". The first four columns form the conditions upon which rules are matched. The IdentityCondition depends on the incoming caller's identity and group classification while the other conditions depend on the receiver's status and preferences. As per Section 8.3.2 the Action columns determines which rules are to be executed. Finally the transformation columns determine which data will be sent to the caller and what filtering will be done on that data.

An authorization rule can be assigned to a group as a whole or to an individual contact. The model recommends using only groups to make it easier to determine the information available to callers. The next section will examine how social relationship groups are integrated into the model.

9.4.2 Social Relationship Groups

The instantiation allows a user to indicate the social relationships within a contact list. This is done by assigning each user to one or more groups. Figure 9.7 illustrates the user interface for this step.

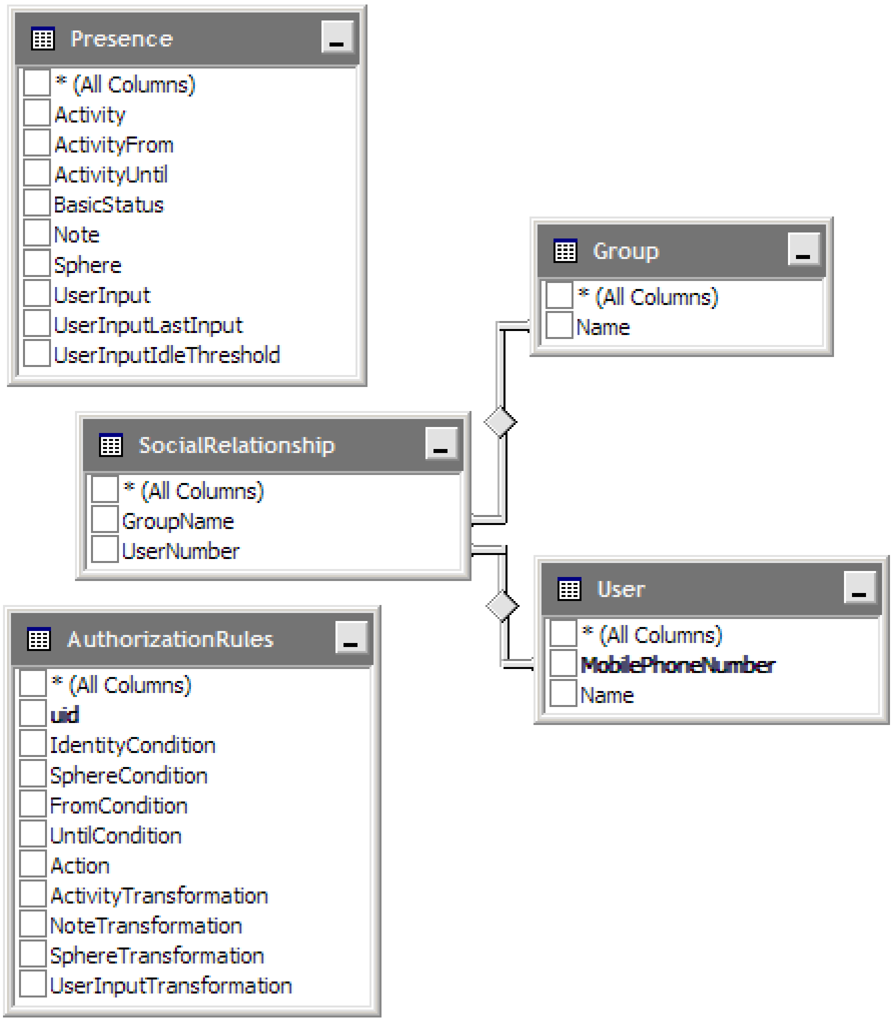

Figure 9.5: Data Model

uid	IdentityCondition	SphereCondition	FromCondition	UntilCondition	Action	ActivityTransformation	NoteTransformation	SphereTransformation	UserInputTransformation
2	Family	Home	*NULL*	*NULL*	30	False	True	True	30
3	Friends	Home	*NULL*	*NULL*	30	True	True	True	30
4	Friends	Work	*NULL*	*NULL*	30	False	True	True	30
5	Co-Workers	Work	*NULL*	*NULL*	30	True	False	True	0
6	General	Work	*NULL*	*NULL*	20	False	False	True	0
8	General	Home	*NULL*	*NULL*	30	False	False	True	10
9	General	*NULL*	2011/01/01 18:00:00	2011/12/31 20:00:00	0	False	False	False	0

Figure 9.6: Authorization Rules Table Data

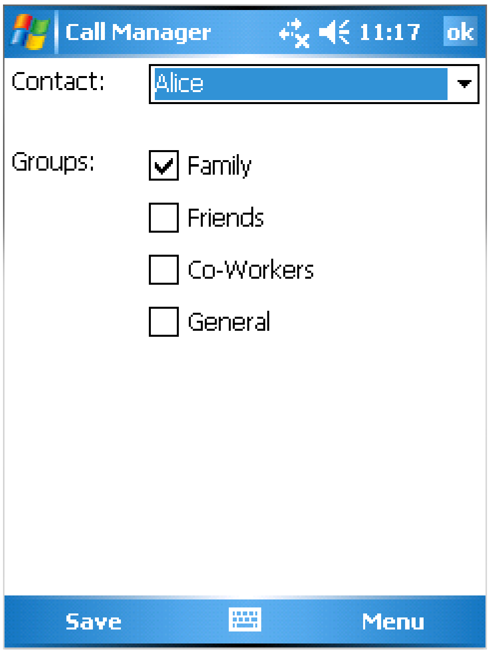

Figure 9.7: Setting Social Relationships

A user may belong to zero or more groups. The instantiation does not enforce the assignment of a group but it is recommended to use at least the "General" group to ensure that presence information is still transmitted. If a contact does not fall into any group no presence information is returned to that user.

The information is sent to the presence server by the PUA in a SIMPLE resource list message, as illustrated below.

```
<?xml version="1.0" encoding="UTF-8"?>
<resource-lists
  xmlns="urn:ietf:params:xml:ns:resource-lists"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <list name="Family">
      <entry uri="sip:alice@example.com">
          <display-name>Alice</display-name>
      </entry>
      <entry uri="sip:bob@example.com">
          <display-name>Bob</display-name>
      </entry>
  </list>
  <list name="Friends">
      <entry uri="sip:carol@example.com">
          <display-name>Carol</display-name>
      </entry>
      <entry uri="sip:dave@example.com">
          <display-name>Dave</display-name>
      </entry>
  </list>
```

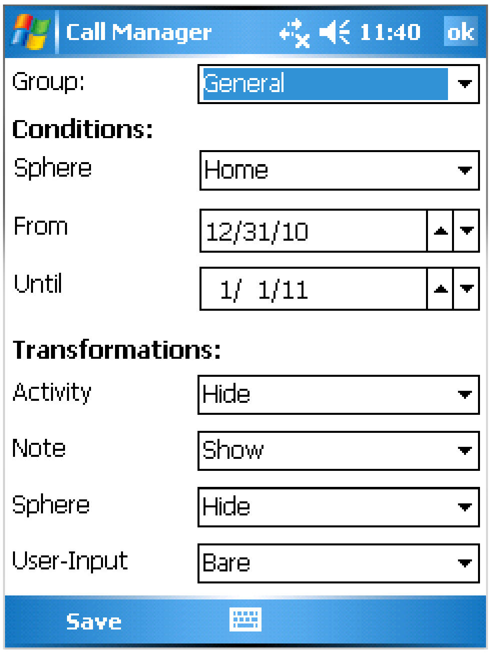

Figure 9.8: Setting Authorization Rules

```
  <list name="Co-Workers">
      <entry uri="sip:dave@example.com">
          <display-name>Dave</display-name>
      </entry>
  </list>
</resource-lists>
```

In addition to specifying groups the receiver must also function as a rule maker and maintain the authorization rules to control which information is available to callers. This is illustrated in Figure 9.8.

The instantiation follows an assignment-by-group approach as this enforces that a group is always available for the identity conditions of a rule. The user has the option to specify other conditions and may also specify the transformation rules. There is no limitation to the amount of rules that can be specified. All applicable rules are used by the presence server to determine the presence information available to callers.

9.4.3 Presence Processing

The presence server examines every incoming presence request to determine the identity of the requestor. The mobile phone number of the caller is used for this purpose, as specified in Section 7.1. The identity is matched against the social relationships created by the receiver. If a request does not fall

Activity	Note	Sphere	UserInput
0	1	1	20

Figure 9.9: Transformation Result

into a pre-existing group the general group is assumed. The SQL to retrieve group information is shown below.

```
SELECT GroupName
FROM   SocialRelationship
WHERE  (UserNumber = '1-425-001-0001')
```

The query yields a result set of all the applicable groups. If no group was found a null value is returned and the general group can be assigned as a substitute value.

Once the group has been determined the presence server examined all applicable authorization rules to determine the presence information to return to the user. Once again an SQL query is used. This query is shown below.

```
SELECT MAX(ActivityTransformation) AS Activity,
       MAX(NoteTransformation) AS Note,
       MAX(SphereTransformation) AS Sphere,
       MAX(UserInputTransformation) AS UserInput
FROM   AuthorizationRules
WHERE  (Action = 30) AND
       (IdentityCondition = @GroupName) AND
       (SphereCondition IS NULL OR SphereCondition = @Sphere) AND
       (FromCondition IS NULL OR FromCondition <= GETDATE()) AND
       (UntilCondition IS NULL OR UntilCondition >= GETDATE())
```

To determine which conditions match the query accepts a group name and sphere parameter. These values are compared against the IdentityCondition and SphereConditions fields respectively. If they match the rule is valid. In addition a from and until validity period is checked against the current date and time. The Action field must contain a value of 30 to allow processing. For all rules that match the conditions the maximum value of the transformations are queried to determine what data will be included in the request. Boolean fields use a value of 1 or 0 for convenience. When running the above query against the data in Figure 9.6, with the values of "Friends" for group and "Work" for sphere, the data in Figure 9.9 is returned.

This indicates that no activity information will be returned, while both the note and sphere will be present. The user-input value of 20 indicates that basic information with an idle-threshold will be returned. These results are in line with the model definition of Section 8.3.3.

The resulting data is returned to the caller in standard presence format. Thus the receiver can control the privacy of presence information by specifying under which conditions information will be available and how that information must be transformed.

9.5 Conclusions

This chapter proves the feasibility of the model through the discussed instantiation. While the instantiation has focused on a specific platform and set of technologies it should be possible to implement the same functionality on any platform which exposes the same device features and developer resources.

By using both SIMPLE and XMPP protocols the instantiation has also shown the generic use of presence standards. In combination with various data access methods this presents a generic and flexible communications platform. The requirement of a data connection can also be solved for older devices by using SMS or USSD communications.

While the chapter presented a complete instantiation no associated user evaluation was performed. Several factors contributed to this decision. Budget constraints limited access to the necessary hardware to perform more widespread testing. While single user tests may provide some information around the usability of the instantiation and usefulness of the model, only widespread testing with a ‘network effect’ can provide a complete evaluation. Multiple callers and receivers using the instantiation would provide a valuable addition to this research and allow for the model to be evaluated more thoroughly.

While implementing the model on a device is convenient it presents a dilemma in the current environment where a multitude of hardware and software combinations exist. It would be much better if the model could be instantiated as a network service which any device can connect to and use. The next chapter will discuss this possibility and how the model can be applied therein.

Chapter 10

The Model in Next Generation Networks

Due to the requirement of research rigour in design science this chapter presents another evaluation of the model. This chapter asks the question of whether the model can be extended to cater for next generation networks in the mobile domain? It examines the entities in these networks and how they could be used for presence-based communications management.

The mobile domain is developing at a rapid pace and technology is continuously unlocking new features and services. In addition, an inevitable convergence with the Internet is taking place. These factors create the possibility to extend the model into next generation networks. A primary example of such a network is the IP Multimedia Subsystem (IMS), which presents a new approach to overcome the inherent limitations of current mobile networks (Camarillo and García-Martín, 2004, p. 6).

This chapter shows how the model developed in Part II can be extended to the IMS. It starts with a brief overview of the network architecture and core components. Next it discusses a service platform which focusses on presence, as examined in Chapter 4. Finally this chapter shows how the model developed through Chapters 5 to 8 fits into this service platform and how it can be used to manage mobile communications.

10.1 The IMS Network Architecture

As people are becoming more mobile in their daily functions, they are demanding access to the services and technologies that have traditionally been

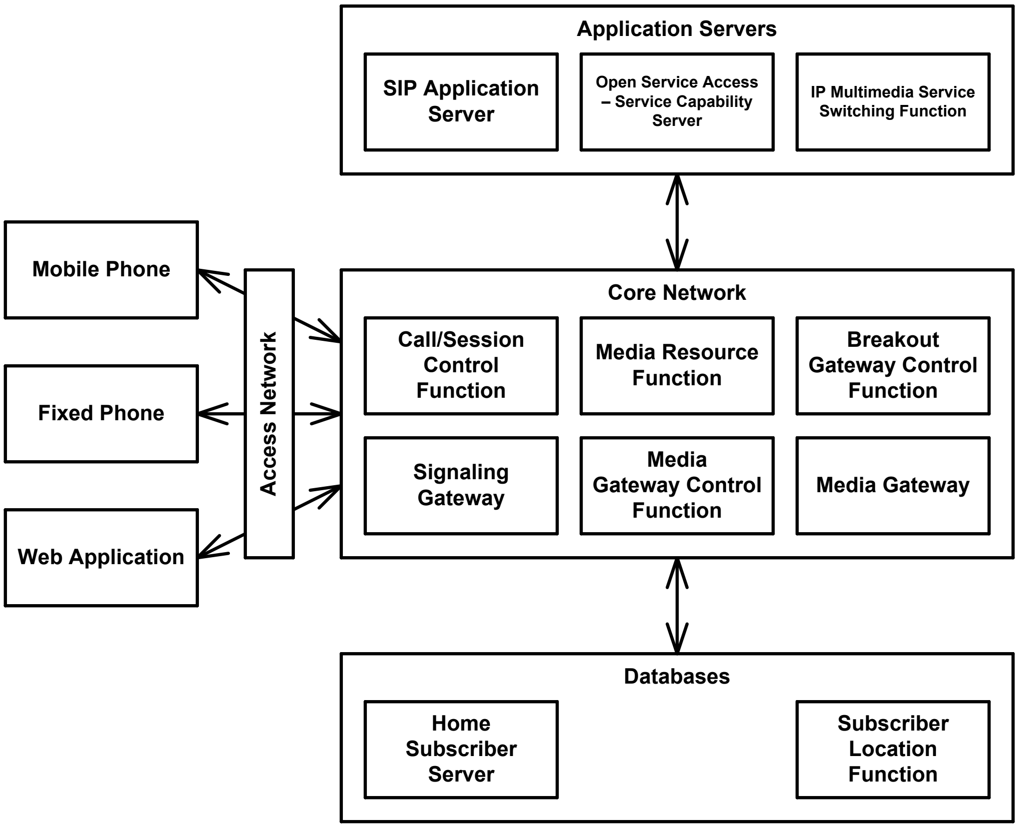

Figure 10.1: The IMS Network Architecture

available to them over the Internet, through a cellular connection. The IMS is heralded as the technology that will make this possible and which will provide "ubiquitous cellular access to all the services that the Internet provides." (Camarillo and García-Martín, 2004, p. 5) A conceptual overview of the network architecture is presented in Figure 10.1.

Users can interact with the network through a variety of terminals. This can include various devices, such as mobile or fixed phones, as well as applications. A terminal is typically connected to the network through a radio link, but other types of access a are also possible.

The network consists of a collection of components, each performing a standardized function. The Home Subscriber Server is the central repository for user information and contains all the subscription data required to manage multimedia sessions. In addition a Subscriber Location Function may be required to map data in large networks which have more than one Home Subscriber Server.

A collection of Call/Session Control Function servers are responsible for signaling. These servers are classified into three categories: Proxy Call/Session Control Function, Interrogating Call/Session Control Function and Serving Call/Session Control Function. Some of the functions performed by these servers include user authentication, message compression, user information retrieval and message routing. A Media Resource Function acts as a provider of media, such as announcements, in the network. The Breakout Gateway Control Function, Signaling Gateway, Media Gateway Control Function and Media Gateway act as gateway servers for sessions addressed to circuit-switched network users.

On top of the core network various application servers reside which host and execute services. These include native Session Initiation Protocol (SIP) Application Servers, which host and execute IP multimedia services, and servers for interfacing with other services such as Open Service Access - Service Capability Servers and IP Multimedia Service Switching Function servers.

The main protocol linking all these components and responsible for establishing and managing sessions (referred to as calls in traditional telephony) is the Session Initiation Protocol (Rosenberg et al., 2002). In addition the Session Description Protocol plays an important role in describing multimedia sessions (Handley, M. and Jaconson, V., 1998).

The principal driving factors behind the development of the IMS are threefold (Camarillo and García-Martín, 2004, p. 7–8). First, it allows a certain quality of service level to be established per session. This is possible because quality of service provisioning can be made during session establishment. Second, operators have the opportunity to charge multimedia sessions more appropriately. This is because operators have knowledge of the actual service that a user is consuming and can create alternative charging schemes which are more enticing to users. Third, it creates a powerful mechanism for providing integrated services to users. Third party service developers can easily create powerful services because the network is based on well known Internet protocols with standard interfaces.

The IMS represents a revolutionary step forward in the merging of the mobile world with the Internet. It promises to take the user experience of mobile services to a whole new level by allowing third party service developers the chance to develop powerful, innovative services and creating a multitude of business opportunities for network operators.

The model developed in this research builds on presence to provide users with a service platform to manage mobile communications. The next section discusses the management of presence in the IMS.

10.2 Presence Management in the IMS

The ability to create advanced services is one of the most important features of the IMS. One of the most significant services that the network will provide is presence. Other services include instant messaging and push-to-talk over cellular. However, presence is a basic service that is likely to become ubiquitous in the future.

10.2.1 Presence as a Service

It is predicted that presence will become a universal service in the future, even in the mobile domain where it is virtually unknown (Camarillo and García-Martín, 2004, p. 303). This will largely be enabled by the presence service in the IMS.

As the IMS leverages existing presence standards the parts of the model dealing with presence publication and composition can be used without modification. IMS presence can provide a much more detailed description of the current user state than currently available in applications, where presence information is limited to user availability. This description can include communication address information, such as email or mobile phone, the terminal capabilities, for example video support, and location information, all distributed in real time to authorized users. Information is transmitted in real time meaning enriched communications and a better end user experience.

In the network presence information is not only available to end-users, but also to other services which can benefit from the information. For example, an answering service which knows when a user comes online can automatically send an instant message notifying the user of pending messages on the server (Camarillo and García-Martín, 2004, p. 323).

Figure 10.2 illustrates the extensions based on the original model. Entities and messages from the original model that are not impacted are hidden from view. The new entities are shown using shading.

In addition to the data and sensors on the mobile phone, services used by the receiver can publish context information. These services can share context with each other and also publish presence information to the presence

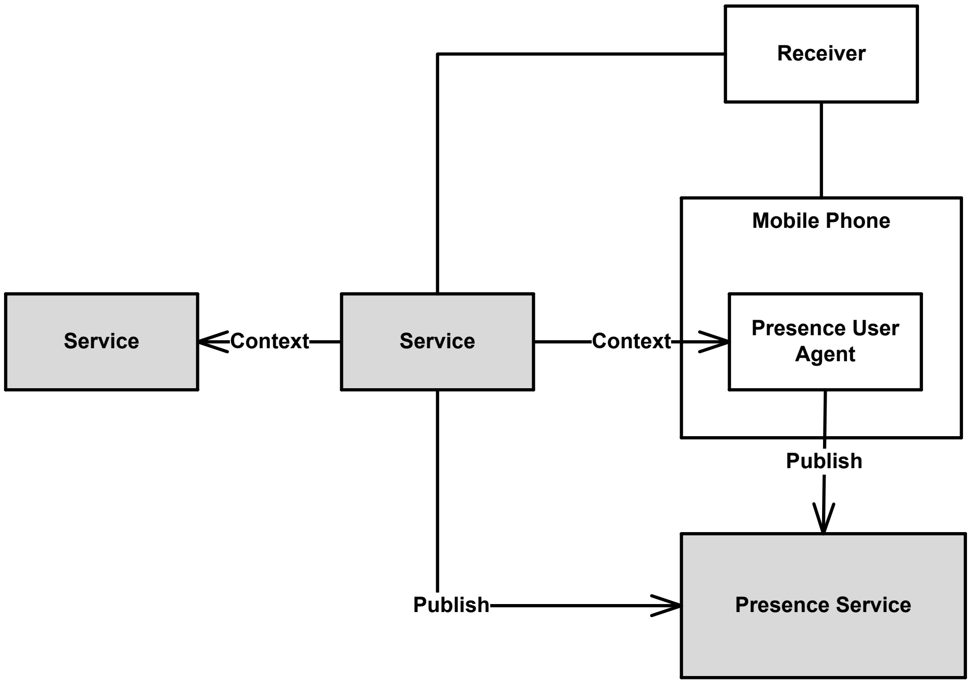

Figure 10.2: Extending the Model with Service Context

service. It should be noted how the original presence server changes to be a presence *service* instead.

Another feature which increases the available context in the IMS is the session protocol, which is examined next.

10.2.2 Session Description Protocol

In the network the Session Description Protocol (SDP) provides a complete description of the session to be established. This description can be further classified into session- and media-level information.

The Session Description Protocol makes session information, such as the subject of the session and the time at which the session is to take place, available. In addition, information about the media requirements for the session, such as port numbers and codecs, can also be retrieved. Figure 10.3 shows an Session Description Protocol message used in the model.

The Session Description Protocol extends the previous caller cue messages which only contained the call subject. The PUA or user can then use this additional information for further decision making regarding the session. When combined with presence and a Session Initiation Protocol message, which contains information such as the user address, routing and security

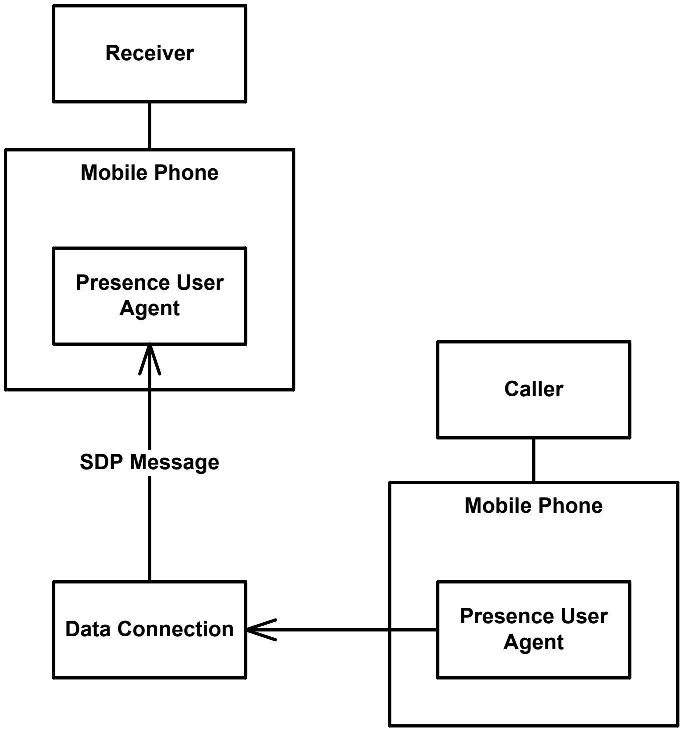

Figure 10.3: Extending the Model with Session Description Information

requirements, this creates a comprehensive set of data which can be used as the basis for a decision model in a communications management service. However, unlike the model a lot of information can be obtained from the network automatically.

Presence should play an important role in making the process of communication more efficient. The next section considers how such call management features can be implemented in the IMS.

10.3 Presence for Call Management in the IMS

The IMS provides an ideal environment for creating innovative services. By leveraging the protocols, standard interfaces and components provided in the network an architecture for managing communication can be created. The model developed in this research provides a foundation for this architecture.

The essential components to enable such a service include the previously discussed Session Description Protocol, Home Subscriber Server, Serving Call/Session Control Function and SIP Application Server. The interaction

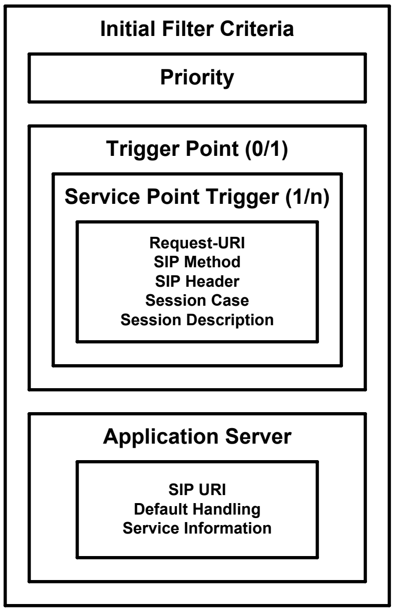

Figure 10.4: Filter Criteria (Camarillo and García-Martín, 2004, p. 164)

between these components are described in the next sections.

10.3.1 Home Subscriber Server

To make a decision on when to provide a particular service to a user, information about the session needs to be combined with some filtering information previously defined by the user and stored on the network. The Home Subscriber Server stores data related to the user in a data structure called the user profile (Poikselkä et al., 2004, p. 101).

Along with various user identities and service profiles the user profile also contains data referred to as filter criteria. The filter criteria determine the services which are applicable to the user's identities in the service profile. Figure 10.4 illustrates the filter criteria structure.

The *priority* field determines the order in which the filter criteria is evaluated by the Serving Call/Session Control Function when multiple criteria exist for a service profile. Thus, multiple services can be invoked for a specific profile. Next, zero or one *trigger points* can be specified. This is a boolean expression filter that determines the criteria for invoking a service and consists of a collection of individual filters called *service point triggers*. The service point trigger can be information stored in different parts of a session request. This includes, for example, the intended recipient (Request-URI), the type of session to be established (SIP Method) and any partial or full match of the session description. Lastly, an *application server* is specified which will process the session request if the conditions described in the trigger points are met.

In effect the Home Subscriber Server contains the decision model determining when a message warrants the involvement of a SIP Application Server which will provide a service to the user. This decision is based on the session information obtained from the Session Description Protocol and is enforced by a Serving Call/Session Control Function.

10.3.2 Serving Call/Session Control Function

When a user registers with the network the user's profile, including filter criteria, is downloaded by the Serving Call/Session Control Function responsible for that user. Furthermore, it is responsible for session control and the routing of information to and from the user.

The Serving Call/Session Control Function inspects every message and evaluates it according to the filter criteria. If the message contains data which correlates to a rule in the filter criteria the Serving Call/Session Control Function will involve the SIP Application Server defined for that rule in the session setup. The SIP Application Server contains the logic to provide a service to the user.

This means that a call can be intercepted by the network without the need for additional services on a terminal. The call profiles on a mobile phone (as defined by the model in Section 7.4) would not be needed as this is handled by the network. Thus in addition to just relaying presence information the service could actively manage communication based on user-defined rules and context data.

Figure 10.5 illustrates how the Home Subscriber Server and Serving Call/ Session Control Function fits into the model. The receiver can update user profile information through a SIP Application Server. This can provide an administrative interface allowing users to manage their own filter criteria. Updated information is distributed to the Serving Call/Session Control Function in real time and can be applied immediately. This information, together with additional information provided by the Session Description Protocol, can then be used by the Serving Call/Session Control Function as input for making call management decisions.

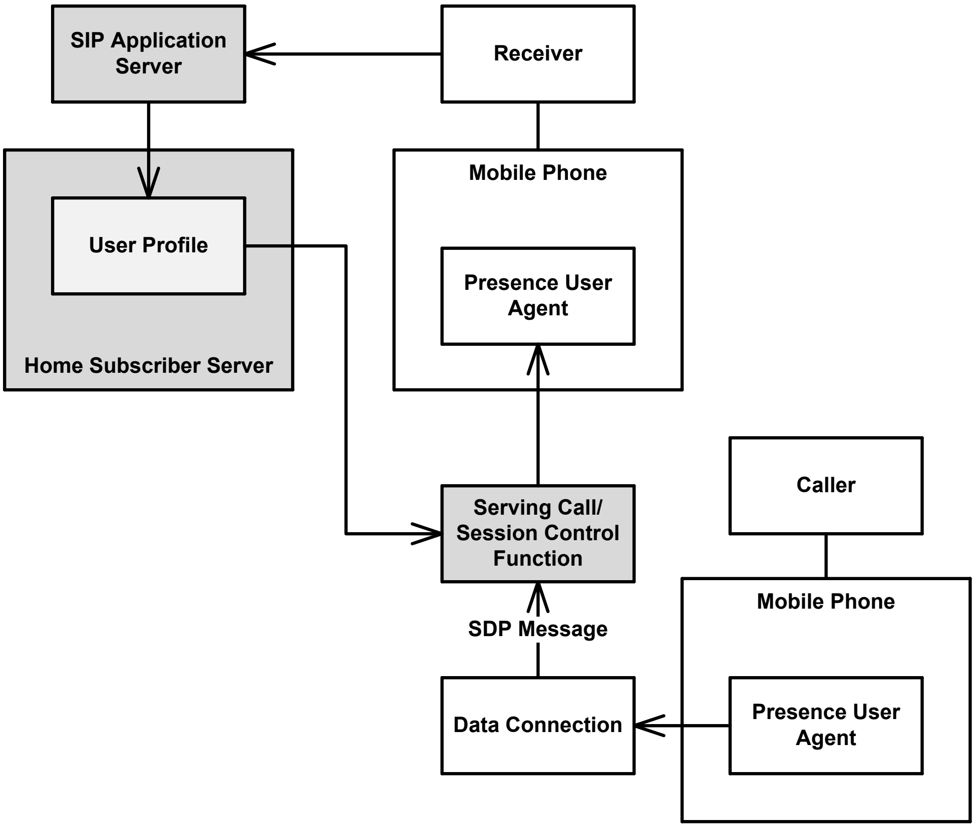

Figure 10.5: Extending the Model with User Profile Information

A typical call management action would be to involve a SIP Application Server to provide a service to the user. This is examined in the next sections.

10.3.3 SIP Application Server

New services developed for the IMS are executed on a SIP Application Server. It is the component responsible for processing the data contained in messages or the Home Subscriber Server and providing a service to the caller, receiver or both of them. An example of a service to the caller is providing presence information, while a service to the receiver could include screening incoming calls.

Figure 10.6 illustrates a SIP Application Server providing a caller service. In this case additional receiver information, beyond presence, may be made available to the caller. This can assist the caller to make a better decision around potential calls. In this scenario a potential concern exists that the receiver's privacy is not protected by authorization rules. Because no authorization policy is used the use of such as service should be considered

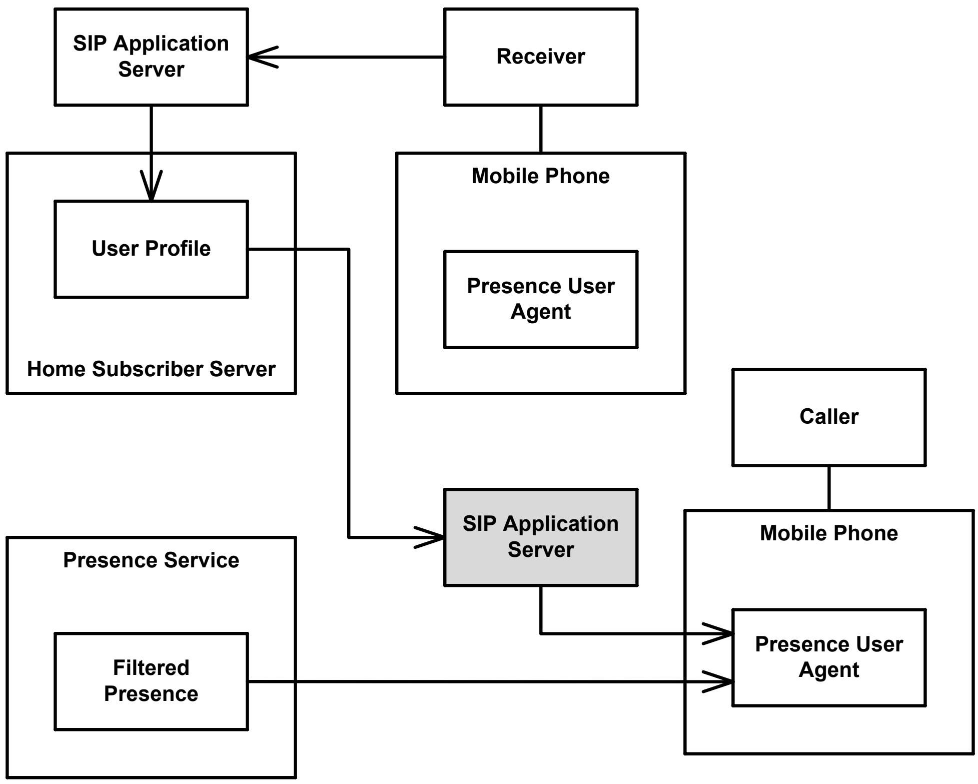

Figure 10.6: Extending the Model with a SIP Application Server

carefully.

A SIP Application Server can be involved in the session in different ways depending on the type of service required. When focusing on a call management service, these specifically include: a SIP Proxy Server and an originating or terminating SIP User Agent.

10.3.4 SIP Proxy Server

When acting as a SIP Proxy Server the SIP Application Server intercepts requests and performs a service before forwarding them on to their next destination. A SIP Application Server acting as a SIP Proxy Server while providing a service to the receiver is illustrated in Figure 10.7.

The Serving Call/Session Control Function decides to involve the SIP Application Server, because of some filter criteria, and forwards the request to the application server. The service executes, after which the request is sent back to the Serving Call/Session Control Function and on to the next node.

As an example, consider a scenario where a receiver is leaving the office to

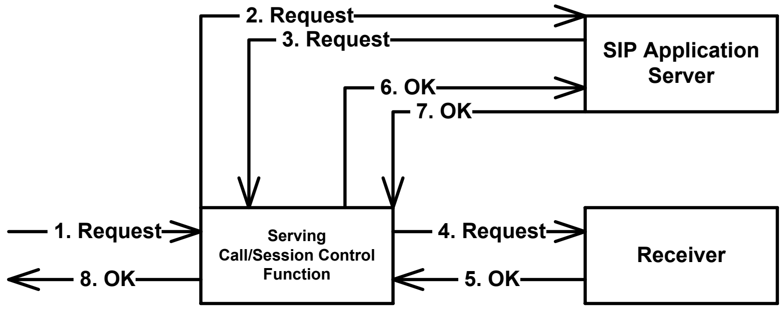

Figure 10.7: SIP Application Server Acting as a SIP Proxy Server

attend a meeting in another city. However, the receiver still wants to receive calls coming through on the office phone. The receiver can configure the filter criteria in the Home Subscriber Server to forward all calls addressed to the office phone to a specific SIP Application Server. When an incoming call is detected, the Serving Call/Session Control Function will examine the session description and forward the call to the SIP Application Server as appropriate (such as when the office phone is being called). The SIP Application Server can then be configured to change the call's destination address to the user's mobile phone before sending the request on, in effect forwarding all calls to the user's current location.

Figure 10.8 illustrates the above scenario. In this case the caller would be unaware of the change in destination but would receive the benefit of getting connected to the receiver. While this service is maintained by the receiver it can be considered a receiver- and caller-oriented approach to call management because it allows callers to reach the receiver in the optimal way. Thus, in addition to indicating availability to the caller, a SIP Application Server can also route calls without the caller becoming aware of a change in the receiver's situation.

It is also possible to create a service which extends the call profile feature of the model. Such a service can be achieved through an originating or terminating SIP User Agent and is examined next.

10.3.5 Originating or Terminating SIP User Agent

In this configuration the SIP Application Server can act as an originating SIP User Agent (such as the caller initiating the session) or, more appropriately for a call management service, as a terminating SIP User Agent. Figure 10.9

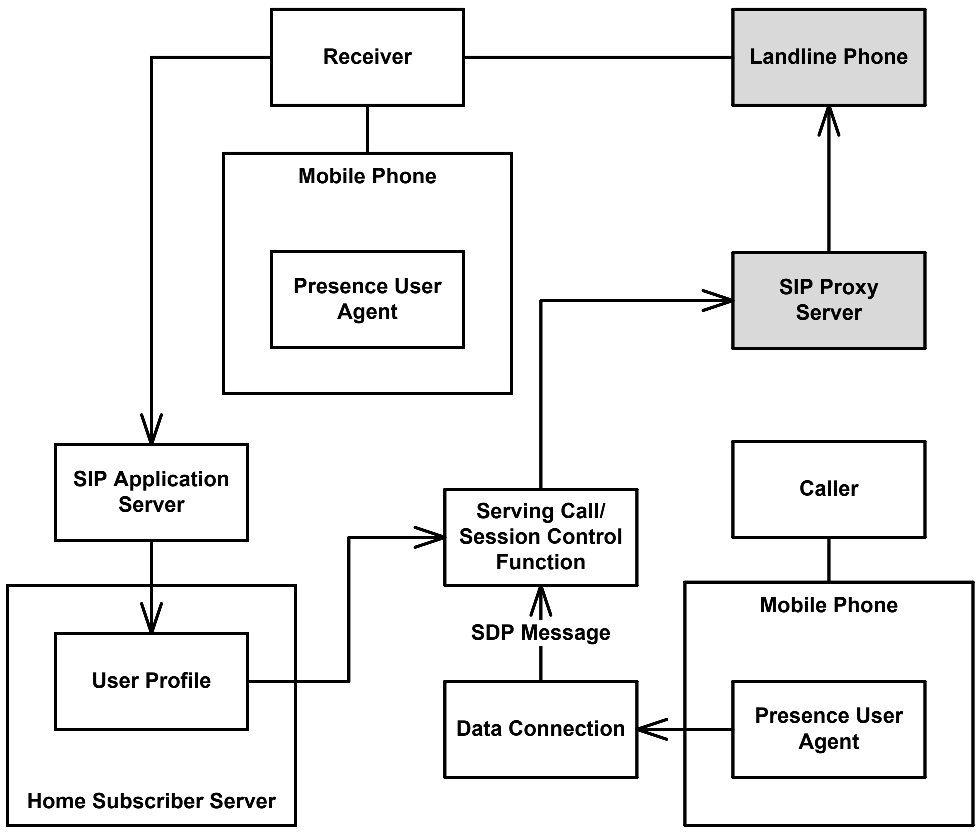

Figure 10.8: Extending the Model with a SIP Proxy Server

illustrates a SIP Application Server acting as a terminating SIP User Agent and providing a service in the terminating call leg (to the receiver). In this scenario the SIP Application Server intercepts requests before they reach the receiver, shielding the receiver from unwanted disturbances.

For example, consider a scenario where a receiver is busy in an important meeting and does not want to be disturbed, except when it is the boss calling. The receiver can configure the filter criteria in the Home Subscriber

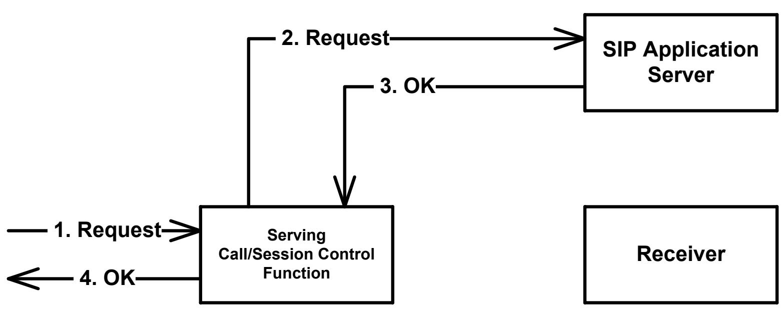

Figure 10.9: SIP Application Server Acting as a Terminating SIP User Agent

Server to forward all calls to a specific SIP Application Server, except when the caller is identified as the boss. When an incoming call is detected, the Serving Call/Session Control Function will examine the session description and forward the call to the SIP Application Server as appropriate (when it is not the boss). The SIP Application Server will then act as a terminating SIP User Agent, blocking the call without interrupting the receiver. In addition, the SIP Application Server could notify the caller to try again later, acting as an originating SIP User Agent and sending a message on behalf of the receiver.

Figure 10.10 illustrates the interaction of this scenario in the model. The dotted-line call profile indicates that the element may be redundant in the IMS. When a call is made to the receiver it can be forwarded to the mobile phone as usual or to a terminating SIP User Agent. In the second case the receiver is shielded from the call by the service. The service can also provide information back to the caller, acting as an originating SIP User Agent. This service acts as a receiver-oriented approach to call management. However, because the service resides on the network it allows more functionality than merely switching call profiles, as discussed in Section 7.4.

An important part of the model is ensuring the privacy of presence information. The IMS also allows new possibilities in this regard. The next section will examine this.

10.4 Presence Authorization in the IMS

Several entities in the IMS provide additional privacy to the user. This extends the functionality available through presence standards.

The information contained by the Session Description Protocol can be used as conditions for authorization rules (as discussed in Section 8.3.1). For example, the type of session can be used to filter requests for presence information when connected to anything other than a voice call.

The filter criteria in the Home Subscriber Server can also be used as part of a presence authorization policy. The filter criteria is similar to the conditions part of authorization rules. However, in addition to specifying when a permission applies the filter criteria can specify specific services to be invoked. This functionality is not possible using existing presence standards. Figure 10.11 illustrates this extension of the model.

In addition services which promote privacy can also be created in the

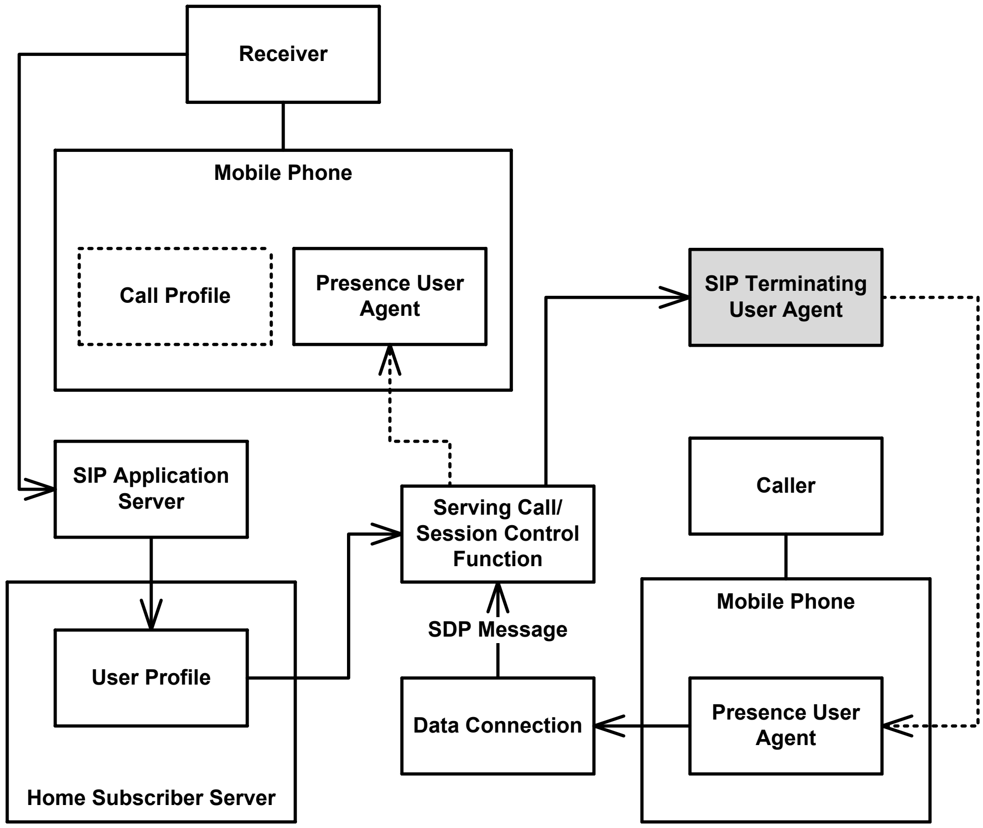

Figure 10.10: Extending the Model with a SIP Terminating User Agent

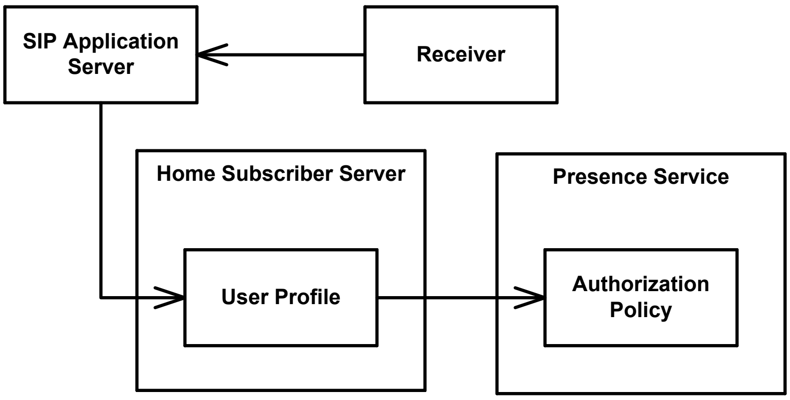

Figure 10.11: Extending the Model with User Profile Information as part of an Authorization Policy

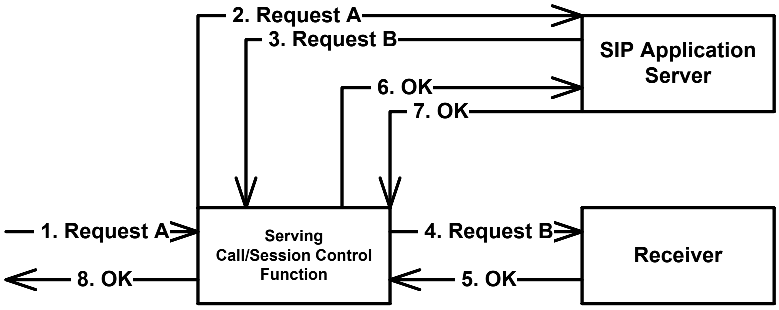

Figure 10.12: SIP Application Server Acting as a SIP Back-To-Back User Agent

IMS, such as implementing a SIP Back-to-Back User Agent.

10.4.1 SIP Back-to-Back User Agent

In general a SIP Back-to-Back User Agent can be seen as two SIP User Agents connected by some application logic. It functions similarly to a SIP Proxy Server, receiving requests and forwarding them to another destination. However, a SIP Back-to-Back User Agent is allowed to perform additional actions, such as modifying any message header field or the Session Description Protocol information and generating requests. This action is not allowed by a SIP Proxy Server. In fact, a SIP Back-to-Back User Agent could be used in the scenarios described above, but the additional complexity required in this server configuration does not always make it desirable. Figure 10.12 shows a SIP Application Server acting as a SIP Back-to-Back User Agent that is providing a service to the receiver.

Messages travelling to and from the receiver can be intercepted and modified by the service before being sent to their final destination. For example, a user could be provided with additional privacy protection by using a SIP Back-to-Back User Agent. The user can configure the filter criteria to direct all communications through a SIP Application Server. The SIP Application Server can then obfuscate fields that reveal information about the user, including header fields and Session Description Protocol information. This includes information that a SIP Application Server acting as a SIP proxy server can not modify, such as the user's IP address. In this way the receiver can not see any information related to the caller and vice versa. This is similar to the transformations part of authorization rules as defined by the

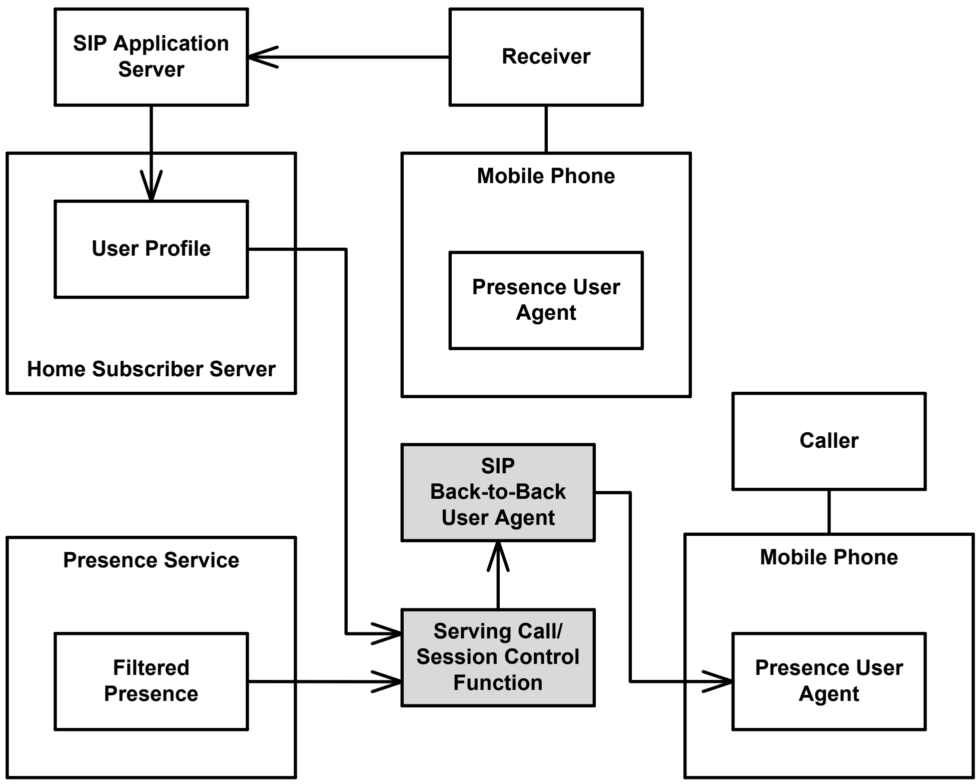

Figure 10.13: Extending the Model with a SIP Back-to-Back User Agent

model in Section 8.3.3.

Figure 10.13 illustrates a SIP Back-to-Back User Agent as part of the model. Such as service can provide additional privacy to a receiver by filtering out information sent to callers.

As a result of the filter criteria it is also possible that several application servers are involved in the session setup. In this case the order in which services are executed become important. This is in contrast with authorization rules where the ordering of a rule in a rule set is irrelevant, as explained in Section 8.3.

By utilizing the functionality and user information described above, a communications management service seems highly feasible in the IMS. Next, a summary of the key points discussed in this chapter is given.

10.5 Conclusions

This chapter examined the question of whether the model can be extended to cater for next generation networks in the mobile domain. The IMS was

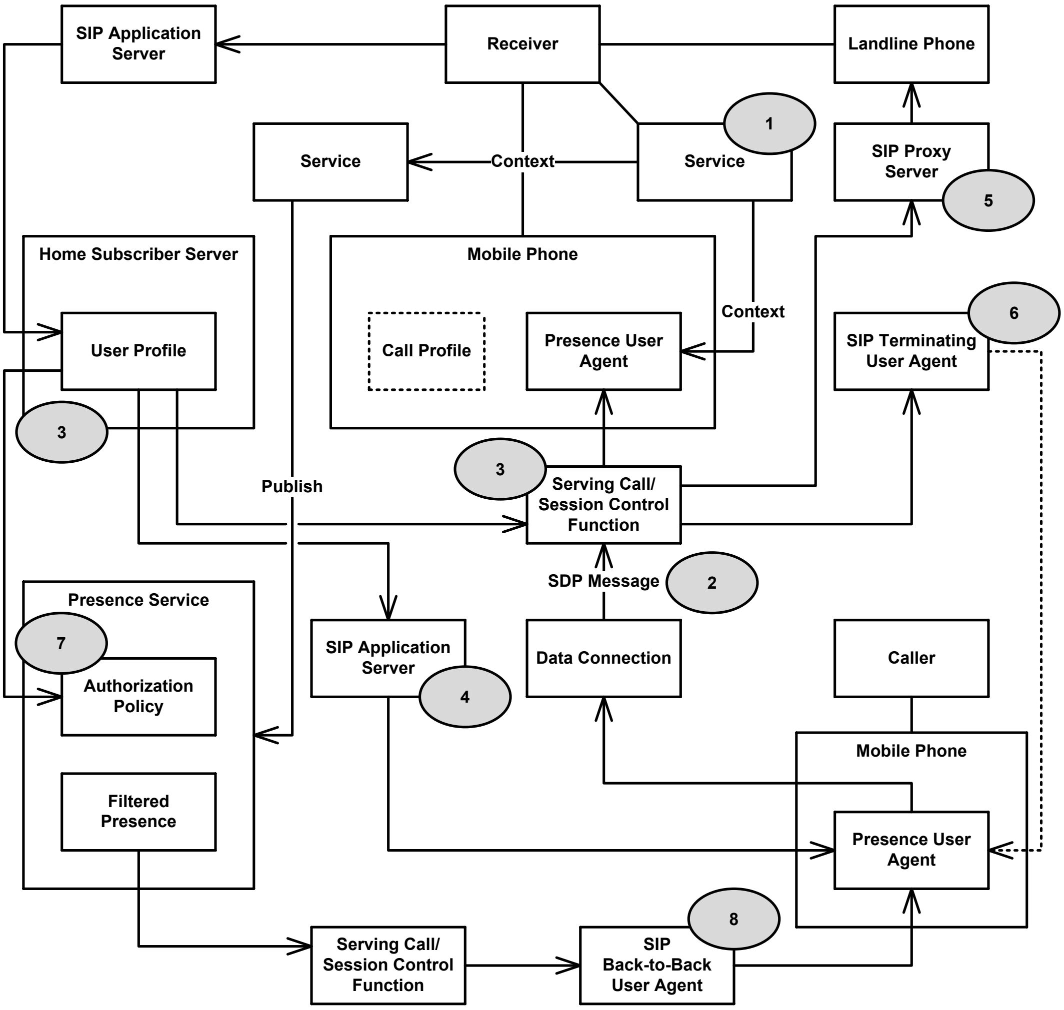

Figure 10.14: IMS Extended Model Summary

used as the foundation for this examination, with a quick overview of the basic network functionality.

The chapter argued that the model is indeed extensible in the IMS. It did this by discussing each aspect of the model, as defined through Chapters 5 to 8, in parallel to the IMS's entities. It was shown that presence services, call management features as well as presence authorization rules could all be extended.

This chapter concludes by providing a compact summary of the mappings and extensions that were discussed. Figure 10.14 provides an illustrations of all the entities that were discussed, with numeric labels indicating the points of extension. Below each extension point will be briefly discussed:

1. IMS services used by a user can provide additional context information

to a PUA or other services. A service can also publish information directly to a presence service. This service replaces the presence server previously used by the model.

2. Session information exposed by the Session Description Protocol further extends the information available to entities in the IMS. Together with service context this extends the information available to make communications management decisions.

3. The user profile information located on the Home Subscriber Server can be used for decision making by services in the IMS. The Serving Call/Session Control Function is the main entity that makes use of this information. The user profile can be updated via a SIP Application Server.

4. User profile information can also be used as an availability indicator to callers. However, in this case the omission of authorization rules must be taken into consideration as a potential loss of privacy can occur.

5. Services can assist users in managing communications. A SIP Proxy Server can be used to manage incoming communications and provide forwarding to a new endpoint. Thus the receiver can remain connected for communication using a single identity on the network.

6. The model call profile construct can be extended by using a SIP Terminating User Agent. This allows communications to be screened and routed in a preferred manner, including blocking calls. The caller may also be informed of any action and given additional options to proceed.

7. The IMS extends privacy features by allowing user profile information to be used as part of presence authorization rules.

8. Services can provide additional filtering of presence or session information before sending it to a caller. An example is a SIP Back-to-Back User Agent.

From the above it can be concluded that the model fits well into the IMS network and can be extended to meet the needs of communications management. The extensions can be based on key concepts in several parts of the research model. Thus the model looks promising for providing value in next generation networks. While the IMS creates the opportunity for implementing innovative features specifically relating to communications management, it is important to remember that such services are only available to users in the IMS.

This concludes Part III of the thesis. It has evaluated the model in two ways: by implementing and discussing an instantiation and by evaluating extensibility in next generation networks in the mobile domain. The next and final part of this research will summarize the research contribution and propose avenues for future work.

Part IV

Epilogue

Chapter 11

Conclusion

This research identified a tangible problem in mobile communications. Such systems are used by billions of people and continue to become a greater part of our lives – which makes the problem relevant and worthwhile to address. Problems which affect such a large group can be both stimulating and daunting at the same time. A worthwhile solution has the power to change lives but there are many difficulties to solve along the way. This chapter summarizes the research contribution and reflects on the relevance and value of the study.

This chapter revisits the research problem, showing how the study has built on existing work as discussed in the background. It then looks at the research objectives, both primary and secondary, and whether they have been met successfully. A similar comparison is made with the research design principles to measure the effectiveness of the study. Next the limitations of the research are acknowledged. Finally some recommendations for future research is made.

11.1 Revisiting the Problem Statement

In the first chapter the problem addressed by this research was clearly defined. In essence it centered on providing more information to users of mobile communications. Without such information mobile calls can become disruptive to the receiver as well as frustrating to callers who cannot connect a call successfully. The information shared between users can be referred to more specifically as context and in principle describes the activity of a person. What made this research unique is not only its focus on context sharing, but also the focus on sharing such information in a privacy-aware manner.

The second chapter showed how other research projects have addressed similar problems. Three main approaches for sharing context emerged from previous work and were discussed in detail. The types of context information which are considered useful were also discussed and compared across research projects. This background created a solid grounding from where this research could continue its specific focus. Most importantly it was seen that no other projects has focused on the privacy aspects of sharing context between mobile subscribers. This illustrated the relevance of this research.

The third chapter also expanded on existing research by examining whether the research area is actually perceived in a similar way by real-world users. Many other research projects simply focused on solving a technical issue as perceived by the researchers without confirming the relevance to the user. However, the survey which was conducted and reported on does show that users are experiencing problems and perceive value in the intended research output. Many important suggestions were also captured which could assist researchers in this field in the future. In this way this research has also contributed to the existing knowledge base of the field.

Finally, the fourth chapter considered the available standards to leverage in addressing the problem. Presence was shown to be a useful form of context which could be leveraged in a solution model. In this regard various concepts and interactions were discussed in the domain. As privacy plays an essential role in this research the final part of chapter four also examined privacy theories. It defined the understanding of privacy used in this research to be one mainly based on context and relationships. In this regard social relationships were examined and it was shown that the way in which they are managed has important consequences for privacy.

Together these chapters frame the research problem and domain in which it resides. Next the research objectives, and whether they have been addressed successfully, are discussed.

11.2 Meeting the Desired Objectives

The primary objective of this research – a prescriptive model for controlling disruptions in mobile communications using established presence standards – was met by the model defined throughout Chapters 5 to 8. The model was defined from high-level constructs and definitions to low-level implementation details which explained it as a whole. As the model was based on presence standards, while incorporating privacy and social relationship theories, it can be concluded that it successfully implements privacy-aware presence management. However several secondary objectives were also identified as important to address.

The perception and practice of users were investigated through the survey discussed in Chapter 3. Suggestions for addressing issues were identified not only by users who completed the survey, but also by assimilating knowledge from related research in the area. Thus the second chapter also contributed to fulfilling this objective.

Of course the primary contribution of Chapter 2 was investigating current theories and existing research regarding mobile communications management. The literature was synthesized using a concept-centric approach, which should also assist future researchers with a compact overview of the research domain.

Chapter 4 addressed the next objective by reviewing presence standards. By examining all existing presence standards and comparing them, a complete overview of presence functionality was given. It could also be seen how mobile-focussed standards differ from others and what technical aspects are of importance in the mobile domain. The review in this chapter also helped to identify candidate standards for the implementation of a prototype system based on the model.

Lastly, aspects regarding privacy were also examined in Chapter 4. The privacy risks in presence standards were discussed, as well as the mechanisms available for the protection of informational privacy. To facilitate a concise model a definition of privacy based on context and relationships was also discussed. This allowed the model's core functionality to be defined in a clear and consistent way.

The model could also not be considered complete without evaluation. In this regard two efforts were made. First Chapter 9 demonstrated a practical implementation of a mobile communications management system based on the model. While showing that the model is practically feasible it also showed that various communication channels could be leveraged for presence messages. Thus the system was complete and robust. Chapter 10 also showed how the model could be extended for future communication networks which merge the Internet and mobile networks. It was shown how the various aspects of the model could map to elements in such networks, allowing the model's use to be extended.

While the research objectives have been met, the research design used in this project, design science, also imposes certain criteria for evaluating projects. The next section examines how this research met design science principles.

11.3 Meeting Design Science Principles

As discussed in the first chapter, design science establishes seven guidelines for effective research. This section examines the guidelines in turn and how this research satisfies each.

1. Design as an artifact. This research produced several artifacts including constructs, a model and an instantiation. As the model was novel and not similar to existing literature it can be said to be innovative. It's use of presence and privacy concepts also makes it purposeful for the research objectives.

2. Problem relevance. Existing literature clearly identifies the research problem. Relevance was further confirmed by the user responses to the conducted survey.

3. Design evaluation. The functionality, completeness and usability of the model was confirmed by an instantiation. This software system demonstrated the operation and feasibility of the model as an artifact.

4. Research contributions. Novel contributions were made by the conducted survey, model and use of presence standards in a new application domain. It was also shown that the issue of contextual privacy had not been examined yet by research efforts in the mobile communications domain. The research contributions are examined in more detail in the next section.

5. Research rigour. The model was defined in detail and presented formally using a standard modelling language. The research also made effective use of the knowledge base and evaluated the artifacts within appropriate environments.

6. Design as a search process. The circumscription process between problem awareness, suggestion, development and evaluation was used as

part of the research methodology. Thus the general design cycle enabled a search process for an effective solution.

7. Communication of research. Several papers at all stages of the project communicated research ideas and results. Papers were presented at International conferences which addressed both technical and managerial audiences.

It can be concluded that, from a design science perspective, the research was executed appropriately and produced acceptable results. Next the research contributions of the thesis are reviewed.

11.4 Research Contributions

The main artifact delivered by this research is a prescriptive model. The model adds new knowledge in the areas of privacy-aware presence management and mobile communications. The model uses existing principles and standards to provide a comprehensive and interoperable solution to the research problem. It was shown that the model can be used as the basis for managing availability and improving mobile communications. As part of the model development knowledge contributions were made in several ways.

Existing knowledge about the problem area was extended through a quantitative analysis of mobile communications management. This analysis was done using a novel survey, and useful empirical data for future research was collected. In addition to the survey, a synthesis of related work should prove useful to future researchers in this area, because it provides a concept-centric overview of the available literature in this domain.

The examination and use of presence standards, as a foundation for the model, provided a comparison of the main presence technologies available today. This comparison identified the strong and weak points of each standard. A focus on privacy features identified several shortcomings in standards which, if addressed, can help to improve and make these standards more complete.

An analysis of privacy theories identified social relationships to play an important part of the model. This is a unique perspective as social relationship theories have not been used in this research area before. The use of relationship groups not only makes logical sense but also assists in the management of presence information. A novel contribution to presence authorization rules was made by incorporating social relationships as a condition for revealing sensitive information.

Finally, the evaluation of the model demonstrated its extensibility in the IP Multimedia Subsystem. As a next generation mobile network, which is based on presence standards, this was an excellent test of the generality of the model. Thus the model presents a solid foundation for future services to be developed.

These contributions have been communicated to a large and diverse audience through the publications in Appendix A. Next the limiting factors present in this research are acknowledged.

11.5 Research Limitations

Research limitations were introduced by the scope of the work as well as the process that was followed in producing the main research output. This section discusses the impact of these factors in more detail.

Chapter 1 indicated that the focus of the model was on the evaluation of mobile communications, in the context of telephony. Thus the research objectives in Section 1.6 centered on the issue of disruption in mobile communications without the influence of other channels. However, it is recognized that other communication channels exist which could add additional, context specific, requirements to the model.

In its use of existing information and standards, the model could also be extended. Only context which was easily obtainable from the mobile phone was considered, as discussed in Section 6.1.1. In addition, defined presence standards with limited extensions were used. These decisions served to keep the model practical, but limits the theoretical contribution and extension of presence information in the context of mobile communications. For example, external sensors on a person could provide information for determining availability for communication and further extend presence standards.

The research process also provides scope for further enhancements. The choice of presence elements, such as in Section 6.2.1 and 6.2.2, were used as examples without being empirically derived. By conducting user tests the most useful information could be established while making further contributions to presence standards.

The research is limited by the lack of user trials of the instantiated model. While the instantiation demonstrates the feasibility of the model it makes no claims as to its usefulness or usability. It is acknowledged that user trials would be a valuable addition in this regard.

The instantiation in Chapter 9 focused on showing a link to the formal model, without specific attention to user interface design. Thus the resulting prototype may contain some usability issues. However, due to the lack of user trials possible issues did not come to the fore in this research.

Several factors made such trials infeasible in the current research. Extensive time would be required for formal usability testing as a sufficient network effect is needed – usefulness may not be apparent with only a few users. This in turn leads to extensive budget requirements, as specific devices (or different device implementations) and enough users would need to be commissioned. However, this addition can provide valuable information to the research community.

While not without its limitations, this research has identified several interesting topics worthy of further investigation. Next some suggestions for future research in this area is made.

11.6 Further Research

In the discussion of the research area, in Chapter 2, it became clear that negotiating communication is a viable option to enhance mobile communications. As this was not the focus of this research it was omitted from the model. However, it could prove an interesting and practically valuable avenue to pursue. Not only would users be interested in such additional functionality, but mobile network operators would also welcome additional streams of network use. Indeed, modern networks should already make some of the problematic issues redundant.

While technology drives innovation, the user is also a critical factor in this research domain. More work on understanding user requirements and behaviour is needed so that effective solutions can be built. Because of the diverse population using mobile technologies it is debatable whether a universal solution will ever be possible. However some needs, such as avoiding unwanted disruptions, are universal and worthwhile investigating from the perspective of the end-user. An evaluation of the model in a practical setting will provide valuable data regarding the usefulness of the model and undoubtedly highlight further areas of investigation.

Regarding the use of presence, an interesting approach is allowing different presence states to be presented. Thus, instead of just limiting access and level of detail, a person would be able to present differing presence information to contacts. It would be interesting to determine if this is more effective in managing communications. While certainly being more flexible, it would be difficult to manage on a per-contact basis. In addition social and moral perceptions could be studied as well. For example, two callers could discover that they are each receiving different presence information. What would they make of that and how could that affect their social relationship with the receiver?

The presence information revealed may also affect the social relationship between the receiver and caller. It is possible that by revealing certain presence information a relationship can be strengthened, according to the theories of self-disclosure. This presents an interesting area for further research, crossing the borders between multiple disciplines.

Technological evolution will also affect future research in this domain. Implementing presence as part of more applications or services should be investigated. It would seem as if many of the advantages of the technology are still going unused in today's instant messaging applications. Exposing rich presence features will allow new services to be developed, which in turn will lead to new user interaction studies and many other possible research areas. This will expand the current research focus beyond mobile telephony.

Social networks are also making presence-like information more ubiquitous, with people becoming more comfortable sharing personal information. Can this information be used to manage mobile communications more effectively? Integration with social networks and services seem like a new and promising avenue to be considered.

11.7 Epilogue

The author trusts that this work has reawakened concern about the past and current situation in mobile communications. It is hoped that users will not just helplessly accept the situation, but actively seek out solutions which will benefit them as well as others in their vicinity. The author would like to conclude this research by expressing the hope that this work will stimulate further research in the subject area.

Part V

Appendices

Appendix A

Summary of Publications

	Publication Details
1.	Ophoff, J. and Botha, R. (2005). Revisiting Reachability Management as a Multilateral Security Mechanism, *ISSA 2005: Peer-reviewed Proceedings of the ISSA 2005 New Knowledge Today Conference*, Information Security South Africa, `http://icsa.cs.up.ac.za/issa/2005/index.htm` **Impact:** Highlighted the problem area and questioned the effects of modern technology.
2.	Ophoff, J. and Botha, R. (2006). Architectural Considerations for Reachability Management in Mobile Communications, *INC 2006: Proceedings of the Sixth International Network Conference*, University of Plymouth, pp. 331–338 **Impact:** Compared the strengths and weaknesses of architectural implementations.
3.	Ophoff, J. and Botha, R. (2006). Privacy-enhancing Call Management in an IP-based Infrastructure, *ICWMC 2006: Proceedings of the Second International Conference on Wireless and Mobile Communications*, IEEE Computer Society, pp. 42–47

Continued on next page ...

Impact: Proposed a call management solution in the IP Multimedia Subsystem.

---

4. Ophoff, J. and Botha, R. (2008). Mobile Communications: User Perception and Practice, *South African Computer Journal*, Vol 40 June 2008, pp. 63–73
   Impact: Presented new data on the perception of mobile communications disruption.

---

5. Ophoff, J. and Botha, R. (2008). Comparing presence standards: IMPS, SIMPLE, and XMPP, *ZA-WWW 2008: Proceedings of the 10th Annual Conference on World Wide Web Applications*, Cape Peninsula University of Technology, `http://www.zaw3.co.za`
   Impact: Analysed prominent presence standards.

---

6. Ophoff, J. and Botha, R. (2008). Unstructured Supplementary Service Data: A Forgotten Technology for Mobile Services?, *WCITD 2008: Proceedings of the 2nd IFIP International Symposium on Wireless Communications and Information Technology in Developing Countries*, International Federation for Information Processing, `http://www.cs.uct.ac.za/Research/DNA/microweb/WCITD2008/index.php`
   Impact: Described the use of USSD as data bearer for call management.

---

7. Ophoff, J. and Botha, R. (2009). Presence on the Web: How to make it more useful, *ZA-WWW 2009: Proceedings of the 11th Annual Conference on World Wide Web Applications*, Cape Peninsula University of Technology, `http://www.zaw3.co.za`
   Impact: Exposed the privacy shortcomings in presence standards and presented a solution implementation.

Copies of these papers are provided on the accompanying CD.

Appendix B

Survey: Questions and Answers

	Question (multiple options)	Abbreviated answers
1.	What is your gender?	Male; Female
2.	Please indicate your age range.	24 or under; 25–34; 35–44; 45–54; 55 or older
3.	Please indicate your primary job title. If you selected other please specify.	Various
4.	Please indicate your primary job function. If you selected other please specify.	Various
5.	Do you have a personal secretary?	Yes; No
6.	On average, how many meetings do you attend per week?	Open-ended
7.	How frequently are you "unavailable" while expecting an important call?	Never; Sometimes; Often
8.	How many mobile (cellular) phones do you use?	0; 1; 2; 3 or more
9.	How frequently do you use the following phone features? (Bluetooth; Caller ID; Short Message Service (SMS); Voice Mail)	Never; Have used, but rarely; Couple of times per month; Couple of times per week; Daily

Continued on next page . . .

10.	How often do you change your Voice Mail message?	Never; Once in a while; Regularly
11.	How frequently do you use your phone for: (Browsing the Internet; Sending/receiving Email; Connecting your PC to the Internet)	Never; Have used, but rarely; Couple of times per month; Couple of times per week; Daily
12.	How often do you switch off your phone to avoid disruption? (Business environment; Social occasions)	Never–Always
13.	Do you use the Caller ID feature of your phone to identify the caller before answering an incoming call?	Not aware of feature; No, I am never concerned with the caller; Yes, sometimes I don't answer because of caller information
14.	Do you use the Caller ID feature of your phone to hide your identity for outgoing calls?	Not aware of feature; No, I never hide my identity; Yes, I hide my identity depending on the person being called; Yes, I always hide my identity
15.	How do you feel about the following statement: "The use of the Caller ID feature can be seen as a threat to your privacy."	Strongly Disagree; Disagree; Neutral; Agree; Strongly Agree
16.	To what extent are you concerned about the misuse of your personal information by people you call?	Not concerned; Slightly concerned; Extremely concerned
17.	To what extent are you concerned about the misuse of your personal information by your network operator?	Not concerned; Slightly concerned; Extremely concerned

Continued on next page . . .

18.	Do you ever question the accuracy of the information provided by your network operator, e.g. billing statements?	Never; Sometimes; Always
19.	How often do you upgrade your mobile phone?	Never, only when it stops working; As soon as my contract allows it; When I need a feature my current phone does not support; Whenever a new model comes out with the latest technology
20.	How often do you download and install applications, games, ringtones, etc. on your mobile phone?	Never; Once in a while; Regularly
21.	Assuming the following information is available for incoming calls, how would you perceive its usefulness? (Caller identification; Subject of call; Priority of call) If any other information would be useful please specify.	Useless–Useful
22.	Please indicate your willingness to specify the following information for outgoing calls. (Caller identification; Subject of call; Priority of call)	Unwilling–Willing
23.	Would you be willing to wait longer for your call to get connected if information such as caller identification, subject of conversation, priority of call, etc. were available?	No, I do not want to wait any longer; Yes, up to 10 seconds longer; Yes, up to 20 seconds longer; Yes, as long as it takes

Continued on next page ...

24.	In mobile networks SMS messages and data traffic are often billed according to the time of day the message/data is being sent. Do these varying message/data costs influence your decision on when to send messages or generate data traffic?	Never; Sometimes; Always
25.	(Optional) If you would like to receive feedback regarding the results of this survey please enter your email address below. This information will only be used in connection with this survey.	Open-ended

Appendix C

Abbreviated XML Schema

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema targetNamespace="urn:ietf:params:xml:ns:common-policy"
    xmlns:cp="urn:ietf:params:xml:ns:common-policy"
    xmlns:xs="http://www.w3.org/2001/XMLSchema"
    elementFormDefault="qualified" attributeFormDefault="unqualified">
    ...
    <!-- //conditions/identity -->
    <xs:complexType name="identityType">
        <xs:complexContent>
            <xs:restriction base="xs:anyType">
                <xs:choice  minOccurs="1" maxOccurs="unbounded">
                    <xs:element name="one" type="cp:oneType"/>
                    <xs:element name="many" type="cp:manyType"/>
                    <xs:element name="list" type="cp:listType"/>
                    <xs:any namespace="##other" processContents="lax"/>
                </xs:choice>
            </xs:restriction>
        </xs:complexContent>
    </xs:complexType>
    <!-- //identity/list -->
    <xs:complexType name="listType">
        <xs:complexContent>
            <xs:restriction base="xs:anyType">
                <xs:sequence>
                    <xs:any namespace="##other"
                    minOccurs="0" processContents="lax"/>
                </xs:sequence>
                <xs:attribute name="id"
                type="xs:anyURI" use="required"/>
                <xs:attribute name="name"
                type="xs:string" use="required"/>
            </xs:restriction>
        </xs:complexContent>
    </xs:complexType>
    ...
</xs:schema>
```

Bibliography

Aarnio, A., Enkenberg, A., Heikkil, J. and Hirvola, S. (2002). Adoption and Use of Mobile Services: Empirical Evidence from a Finnish Survey, *HICSS '02: Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS'02)-Volume 3*, IEEE Computer Society, Washington, DC, USA, pp. 87–96.

Abowd, G. D. and Mynatt, E. D. (2000). Charting past, present, and future research in ubiquitous computing, *ACM Transactions on Computuer-Human Interaction (TOCHI)* **7**(1): 29–58.

AG Software (2010). MatriX XMPP SDK. Cited: 21 November 2010.
URL: *http://www.ag-software.de/matrix-xmpp-sdk/*

Ammenwerth, E., Buchauer, A., Bludau, B. and Haux, R. (2000). Mobile information and communication tools in the hospital, *International Journal of Medical Informatics* **57**(1): 21–40.

Bardram, J. E. and Hansen, T. R. (2004). The AWARE Architecture: Supporting Context-Mediated Social Awareness in Mobile Cooperation, *CSCW '04: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, pp. 192–201.

Barth, A., Datta, A., Mitchell, J. C. and Nissenbaum, H. (2006). Privacy and Contextual Integrity: Framework and Applications, *2006 IEEE Symposium on Security and Privacy*, pp. 184–198.

Baskerville, R. (2008). What design science is not, *European Journal of Information Systems* **17**(5): 441–443.

Becker, D. (2004). Beware of geeks bearing 'presence'. Cited: 05 Oct 2005.
URL: *http://news.com.com*

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

Bentley, F. R. and Metcalf, C. J. (2008). Location and Activity Sharing in Everyday Mobile Communication, *CHI '08: Extended Abstracts on Human Factors in Computing Systems*, ACM Press, pp. 2453–2462.

Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathan, A. and Riboni, D. (2010). A survey of context modelling and reasoning techniques, *Pervasive and Mobile Computing* **6**(2): 161–180.

Booth, W. C., Colomb, G. G. and Williams, J. M. (2008). *The Craft of Research*, 3rd edn, The University of Chicago Press.

Camarillo, G. and García-Martín, M. A. (2004). *The 3G IP Multimedia Subsystem (IMS) Merging the Internet and the Cellular Worlds*, John Wiley & Sons, Ltd.

Carroll, J. M. (2002). Introduction: Human-Computer Interaction, the Past and the Present, *in* J. M. Carroll (ed.), *Human-Computer Interaction in the New Millennium*, ACM Press.

Carstensen, P. H. and Schmidt, K. (1999). Computer Supported Cooperative Work: New Challenges to Systems Design, *Handbook of Human Factors*, pp. 619–636.

Chaum, D. (1985). Security without identification: Transaction systems to make big brother obsolete, *Communications of the ACM* **28**(10): 1030–1044.

Checkland, P. (1981). *Systems Thinking, Systems Practice*, John Wiley & Sons.

Cohen, J. (1988). *Statistical Power Analysis*, 2nd edn, Academic Press.

Collis, J. and Hussey, R. (2003). *Business Research*, 2nd edn, Palgrave Macmillan.

Consolvo, S., Smith, I. E., Matthews, T., LaMarca, A., Tabert, J. and Powledge, P. (2005). Location Disclosure to Social Relations: Why, When, & What People Want to Share, *CHI '05: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, pp. 81–90.

Cresswell, J. C. (2003). *Research Design*, 2nd edn, Sage Publications.

Dabbish, L. and Kraut, R. (2004). Controlling Interruptions: Awareness Displays and Social Motivation for Coordination, *In Proc of CSCW 2004, ACM Press*, ACM Press, pp. 182–191.

Danezis, G., Lewis, S. and Anderson, R. (2005). How Much is Location Privacy Worth?, *Proceedings of the Fourth Workshop on the Economics of Information Security*.

Davenport, T. H. and Beck, J. C. (2000). Getting the Attention You Need, *Harvard Business Review* **78**(5): 118–126.

De Guzman, E. S., Sharmin, M. and Bailey, B. P. (2007). Should I Call Now? Understanding What Context is Considered When Deciding Whether to Initiate Remote Communication via Mobile Devices, *GI '07: Proceedings of Graphics Interface 2007*, ACM Press, pp. 143–150.

DeVito, J. A. (1992). *The Interpersonal Communication Book*, 6th edn, HarperCollins Publishers.

Dourish, P. (2004). What we talk about when we talk about context, *Personal Ubiquitous Computing* **8**(1): 19–30.

Dourish, P. and Anderson, K. (2006). Collective Information Practice: Exploring Privacy and Security as Social and Cultural Phenomena, *Human-Computer Interaction* **21**(3): 319–342.

Feenberg, A. (2003). Active and Passive Bodies: Comments on Don Ihde's Bodies in Technology, *Techn: Research in Philosophy and Technology* **7**(2).
URL: *http://scholar.lib.vt.edu/ejournals/SPT/v7n2/feenberg.html*

Fillery, P. F., Rusli, A. and James, H. L. (1996). Describing the Problem Situation in IS Studies Using SSM: a Practitioner's View, *Proceedings of the 1996 Information Systems Conference of New Zealand (ISCNZ '96)*, IEEE Computer Society, pp. 2–9.

Gray, D. E. (2004). *Doing Research in the Real World*, Sage Publications.

Handley, M. and Jaconson, V. (1998). *SDP: Session Description Protocol*, RFC 2327, Internet Engineering Task Force.

Hartley, P. (1993). *Interpersonal Communication*, Routledge.

Henry-Labordère, A. and Jonack, V. (2004). *SMS and MMS Interworking in Mobile Networks*, Artech House, Inc.

Hevner, A. R., March, S. T., Park, J. and Ram, S. (2004). Design Science in Information Systems Research, *MIS Quarterly* **28**(1): 75–105.

Horvitz, E. and Apacible, J. (2003). Learning and Reasoning about Interruption, *ICMI '03: Proceedings of the 5th International Conference on Multimodal Interfaces*, pp. 20–27.

Horvitz, E., Apacible, J., Subramani, M., Sarin, R., Koch, P., Cadiz, J., Narin, A. and Rui, Y. (2003). Experiences with the Design, Fielding, and Evaluation of a Real-Time Communications Agent, *Technical Report MSR-TR-2003-98*, Microsoft Research.

Horvitz, E., Koch, P., Kadie, C. M. and Jacobs, A. (2002). Coordinate: Probabilistic Forecasting of Presence and Availability, *Proceedings of the 18th Conference in Uncertainty in Artificial Intelligence (UAI'02)*, pp. 224–233.

Independentsoft (2010). SIP.NET. Cited: 21 November 21 2010.
URL: *http://www.independentsoft.de/sip/*

International Telecommunication Union (2009). Measuring the Information Society - The ICT Development Index, *Technical report*, International Telecommunication Union.
URL: *http://www.itu.int/*

Johnson, G., Kolsky, E., Basso, M., Elliot, B., Sumic, Z., Maoz, M., Moore, C., Steenstrup, K., Cramoysan, S., Fenn, J., Shen, S., Hafner, B., Drakos, N., Smith, D., Girard, J., Mann, J. and Cain, M. W. (2007). Hype Cycle for Enterprise Communication Applications, *Technical report*, Gartner. Gartner Research Report, G00148589.
URL: *http://www.gartner.com/*

Jones, J. (2008). Opinions on Cell Phone Use on Airplanes, Congestion, and Telecommuting – from the 2006 and 2007 Omnibus Household Survey, *Technical report*, U.S. Department of Transportation, Bureau of Transportation Statistics (BTS).
URL: *http://www.bts.gov/*

Kaba, B., N'Da, K. and Mbarika, V. (2008). Understanding the Factors Influencing the Attitude Toward and the Use of Mobile Technology in Developing Countries: A Model of Cellular Phone Use in Guinea, *HICSS '08: Proceedings of the Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, IEEE Computer Society, Washington, DC, USA, p. 127.

Karat, C., Brodie, C. and Karat, J. (2006). Usable Privacy and Security for Personal Information Management, *Communications of the ACM* **49**(1): 56–57.

Karatzouni, S., Furnell, S., Clarke, N. and Botha, R. (2007). Perceptions of User Authentication on Mobile Devices, *Proceedings of the ISOneWorld Conference*. CD Proceedings.

Khalil, A. and Connelly, K. (2006). Context-aware Telephony: Privacy Preferences and Sharing Patterns, *CSCW '06: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, ACM Press, pp. 469–478.

Kononen, V., Mantyjarvi, J., Simila, H., Parkka, J. and Ermes, M. (2010). Automatic feature selection for context recognition in mobile devices, *Pervasive and Mobile Computing* **6**(2): 181–197.

Lamble, D., Kauranen, T., Laakso, M. and Summala, H. (1999). Cognitive load and detection thresholds in car following situations: safety implications for using mobile (cellular) telephones while driving, *Accident Analysis and Prevention* **31**(6): 617–623.

Light, A. (2008). Transports of delight? What the experience of receiving (mobile) phone calls can tell us about design, *Personal and Ubiquitous Computing* **12**(5): 391–400.

Ling, R. (2004). *The Mobile Connection: The Cell Phone's Impact on Society*, Morgan Kaufmann.

Ljungstrand, P. (2001). Context awareness and mobile phones, *Personal and Ubiquitous Computing* **5**(1): 58–61.

Loke, S. (2007). *Context-Aware Pervasive Systems*, Auerbach.

Love, S. and Perry, M. (2004). Dealing with Mobile Conversations in Public Places: some implications for the design of socially intrusive technologies, *CHI '04: CHI '04 Extended Abstracts on Human Factors in Computing Systems*, pp. 1195–1198.

March, S. T. and Smith, G. F. (1995). Design and natural science research on information technology, *Decision Support Systems* **15**(4): 251–266.

Marczyk, G. R., DeMatteo, D. and Festinger, D. (2005). *Essentials of Research Design and Methodology*, John Wiley & Sons, Inc.

Marmasse, N., Schmandt, C. and Spectre, D. (2004). WatchMe: Communication and Awareness Between Members of a Closely-Knit Group, *Ubiquitous Computing*, Springer, pp. 214–231.

Microsoft (2005). Microsoft Releases Windows Mobile 5.0. Cited: 21 November 2010.
URL: *http://www.microsoft.com/presspass/press/2005/may05/05-10windowsmobile5pr.mspx*

Microsoft TechNet (2010). What is Presence? Cited: 14 March 2010.
URL: *http://technet.microsoft.com/en-us/library/bb894395.aspx*

Miles, R. and Hamilton, K. (2006). *Learning UML 2.0*, O'Reilly Media.

Milewski, A. E. and Smith, T. M. (2000). Providing Presence Cues to Telephone Users, *CSCW '00: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, ACM Press, pp. 89–96.

Monk, A. and Howard, S. (1998). The Rich Picture: A Tool for Reasoning About Work Context, *Interactions* **5**(2): 21–30.

Moor, J. H. (1997). Towards a Theory of Privacy in the Information Age, *ACM SIGCAS Computers and Society* **27**(3): 27–32.

Mouton, J. (2008). *How to succeed in your Master's and Doctoral Studies*, 11th edn, Van Schaik Publishers, Pretoria.

Nelson, L., Bly, S. and Sokoler, T. (2001). Quiet Calls: Talking Silently on Mobile Phones, *CHI '01: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, pp. 174–181.

Nickerson, R. C. and Isaac, H. (2006). "Am I Disturbing You?": A Research Model for Investigating the Use of Mobile Phones in Social Settings, *ICMB '06: Proceedings of the International Conference on Mobile Business*, IEEE Computer Society, Washington, DC, USA, pp. 35–41.

Nickerson, R. C., Isaac, H. and Mak, B. (2008). A multi-national study of attitudes about mobile phone use in social settings, *International Journal of Mobile Communications* **6**(5): 541–563.

Nielsen, J. (1993). *Usability Engineering*, Morgan Kaufmann.

Nissenbaum, H. (1997). Toward an Approach to Privacy in Public: Challenges of Information Technology, *Ethics & Behavior* **7**(3): 207–219.

Nissenbaum, H. (1998). Protecting Privacy in an Information Age: The Problem of Privacy in Public, *Law and Philosophy* **17**(5–6): 559–596.

Nissenbaum, H. (2004). Privacy as Contextual Integrity, *Washington Law Review* **79**(1): 119–158.

Open Mobile Alliance (2005). WV-041 Features and Functions, *Technical report*, Open Mobile Alliance.
URL: *http://www.openmobilealliance.org/technical/*

Open Mobile Alliance (2007a). IMPS Architecture, *Technical report*, Open Mobile Alliance.
URL: *http://www.openmobilealliance.org/technical/*

Open Mobile Alliance (2007b). OMA Presence SIMPLE 2.0 Requirements, *Technical report*, Open Mobile Alliance.
URL: *http://www.openmobilealliance.org/technical/*

Open Mobile Alliance (2008). Presence SIMPLE Requirements, *Technical report*, Open Mobile Alliance.
URL: *http://www.openmobilealliance.org/technical/*

Oulasvirta, A. (2008). Designing Mobile Awareness Cues, *MobileHCI '08: Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services*, ACM Press, pp. 43–52.

Oulasvirta, A., Raento, M. and Tiitta, S. (2005). ContextContacts: Re-Designing SmartPhone's Contact Book to Support Mobile Awareness and

Collaboration, *MobileHCI '05: Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices and Services*, ACM Press, pp. 167–174.

Palen, L. (2002). Mobile Telephony in a Connected Life, *Communications of the ACM* **45**(3): 78–82.

Palen, L. and Dourish, P. (2003). Unpacking "Privacy" for a Networked World, *CHI '03: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 129–136.

Palen, L., Salzman, M. and Youngs, E. (2000). Going Wireless: Behavior & Practice of New Mobile Phone Users, *CSCW '00: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, pp. 201–210.

Pedersen, E. R. (2001). Calls.calm: Enabling Caller and Callee to Collaborate, *CHI '01: CHI '01 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, pp. 235–236.

Pering, C. (2002). Taming of the Ring: Context Specific Social Mediation for Communication Devices, *CHI '02: CHI '02 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, pp. 712–713.

Petersen, J. K. (2002). *The Telecommunications Illustrated Dictionary*, 2nd edn, CRC Press.

Poikselkä, M., Mayer, G., Khartabil, H. and Niemi, A. (2004). *The IMS IP Multimedia Concepts and Services in the Mobile Domain*, John Wiley & Sons, Ltd.

Raento, M. and Oulasvirta, A. (2008). Designing for privacy and self-presentation in social awareness, *Personal and Ubiquitous Computing* **12**(7): 527–542.

Raento, M., Oulasvirta, A., Petit, R. and Toivonen, H. (2005). ContextPhone: A Prototyping Platform for Context-Aware Mobile Applications, *IEEE Pervasive Computing* **4**(2): 51–59.

Rannenberg, K. (2000a). How Much Negotiation and Detail Can Users Handle? Experiences with Security Negotiation and the Granularity of Access

Control in Communications, *ESORICS '00: Proceedings of the 6th European Symposium on Research in Computer Security*, Springer-Verlag, pp. 37–54.

Rannenberg, K. (2000b). Multilateral Security A Concept and Examples for Balanced Security, *NSPW '00: Proceedings of the 2000 workshop on New security paradigms*, ACM Press, pp. 151–162.

Redelmeier, D. A. and Tibshirani, R. J. (1997). Association between Cellular-Telephone Calls and Motor Vehicle Collisions, *The New England Journal of Medicine* **336**(7): 453–458.

Reichenbach, M., Damker, H., Federrath, H. and Rannenberg, K. (1997). Individual Management of Personal Reachability in Mobile Communication, *Information Security in Research and Business, Proceedings of the IFIP TC11 13th International Conference on Information Security (SEC '97)*, Chapman & Hall, pp. 164–174.

Rennecker, J. and Godwin, L. (2005). Delays and interruptions: A self-perpetuating paradox of communication technology use, *Information and Organization* **15**(3): 247–266.

Roloff, M. E. (1981). *Interpersonal Communication: The Social Exchange Approach*, Sage Publications.

Rosenberg, J. (2000). Presence: The Best Thing That Ever Happened To Voice. Cited: 05 Oct 2005.
URL: *http://www.cconvergence.com*

Rosenberg, J. (2006). A Data Model for Presence, *Technical Report RFC 4479*, Internet Engineering Task Force.
URL: *http://www.ietf.org/rfc/rfc4479.txt*

Rosenberg, J. (2007a). Extensible Markup Language (XML) Formats for Representing Resource Lists, *Technical Report RFC 4826*, Internet Engineering Task Force.
URL: *http://www.ietf.org/rfc/rfc4826.txt*

Rosenberg, J. (2007b). Presence Authorization Rules, *Technical Report RFC 5025*, Internet Engineering Task Force.
URL: *http://www.ietf.org/rfc/rfc5025.txt*

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M. and Schooler, E. (2002). SIP: Session Initiation Protocol, *Technical Report RFC 3261*, Internet Engineering Task Force.
URL: *http://www.ietf.org/rfc/rfc3261.txt*

Saint-Andre, P. (2004). Extensible Messaging and Presence Protocol (XMPP): Instant Messaging and Presence, *Technical Report RFC 3921*, Internet Engineering Task Force.
URL: *http://www.ietf.org/rfc/rfc3921.txt*

Saint-Andre, P., Smith, K. and Troncon, R. (2009). *XMPP: The Definitive Guide*, O'Reilly Media.

Salin, P. (2004). *Mobile instant messaging systems - a comparative study and implementation*, Master's thesis, Department of Computer Science and Engineering, Helsinki University of Technology.

Salkintzis, A. K. and Passas, N. (eds) (2005). *Emerging Wireless Multimedia Services and Technologies*, John Wiley & Sons.

Santos, A. C., Cardoso, J. M., Ferreira, D. R., Diniz, P. C. and Chanho, P. (2010). Providing user context for mobile and social networking applications, *Pervasive and Mobile Computing* **6**(3): 324–341.

Schmidt, A., Takaluoma, A. and Mäntyjärvi, J. (2000). Context-Aware Telephony Over WAP, *Personal and Ubiquitous Computing* **4**(4): 225–229.

Schulzrinne, H. (2004). The tel URI for Telephone Numbers, *Technical Report RFC 3966*, Internet Engineering Task Force.
URL: *http://www.ietf.org/rfc/rfc3966.txt*

Schulzrinne, H., Tschofenig, H., Morris, J., Cuellar, J., Polk, J. and Rosenberg, J. (2007). Common Policy: A Document Format for Expressing Privacy Preferences, *Technical Report RFC 4745*, Internet Engineering Task Force.
URL: *http://www.ietf.org/rfc/rfc4745.txt*

Shacham, R., Kellerer, W., Schulzrinne, H. and Thakolsri, S. (2007). Composition for Enhanced SIP Presence, *12th IEEE Symposium on Computers and Communications (ISCC 2007)*, pp. 203–210.

Simon, H. A. (1996). *The Sciences of the Artificial*, 3rd edn, The MIT Press.

Sutherland, E. (2009). Counting customers, subscribers and mobile phone numbers, *info* **11**(2): 6–23.

Tang, J. C. and Begole, J. (2003). Beyond Instant Messaging, *ACM Queue* **1**(8): 28–37.

Tang, J. C., Yankelovich, N., Begole, J., Van Kleek, M., Li, F. and Bhalodia, J. (2001). ConNexus to Awarenex: Extending awareness to mobile users, *CHI '01: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, pp. 221–228.

Tavani, H. T. (2008). Informational Privacy: Concepts, Theories, and Controversies, *in* K. E. Himma and H. T. Tavani (eds), *The Handbook of Information and Computer Ethics*, John Wiley & Sons.

Tavani, H. T. and Moor, J. H. (2001). Privacy Protection, Control of Information, and Privacy-Enhancing Technologies, *ACM SIGCAS Computers and Society* **31**(1): 6–11.

The Wireless Village (2001). The Wireless Village initiative: White Paper, *Technical report*, The Wireless Village.
URL: *http://www.openmobilealliance.org/tech/affiliates/wv/wvindex.html*

Thompson, K. (2006). Are we keeping up with cellular tech? Cited: 23 June 2006.
URL: *http://www.itweb.co.za/sections/telecoms/2006/0606231425.asp*

Trenholm, S. and Jensen, A. (1996). *Interpersonal Communication*, 3rd edn, Wadsworth Publishing Company.

Vaishnavi, V. K. and Kuechler Jr., W. (2008). *Design Science Research Methods and Patterns: Innovating Information and Communication Technology*, Auerbach Publications.

Vaughan-Nichols, S. (2003). Presence Technology: More than Just Instant Messaging, *IEEE Computer* **36**(10): 11–13.

Weaver II, R. L. (1993). *Understanding Interpersonal Communication*, 6th edn, HarperCollins College Publishers.

Webster, J. and Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review, *MIS Quarterly* **26**(2): 13–23.

Wiberg, M. and Whittaker, S. (2005). Managing availability: Supporting Lightweight Negotiations to Handle Interruptions, *ACM Transactions on Computer-Human Interaction* **12**(4): 356–387.