

Gesellschaft für Informatik (GI)

publishes this series in order to make available to a broad public recent findings in informatics (i.e. computer science and information systems), to document conferences that are organized in cooperation with GI and to publish the annual GI Award dissertation.

Broken down into the fields of

- Seminar
- Proceedings
- Dissertations
- Thematics

current topics are dealt with from the fields of research and development, teaching and further training in theory and practice. The Editorial Committee uses an intensive review process in order to ensure the high level of the contributions.

The volumes are published in German or English.

Information: <http://www.gi-ev.de/service/publikationen/lni/>

ISSN 1617-5468

ISBN 978-3-88579-240-6

This volume contains the proceedings of the fourth conference organized by the GI special interest group on Mobility and Mobile Information Systems (GI-Fachgruppe MMS). The aim of both, the interest group and the conference, is to offer a platform for discussion on all subjects relevant to mobile and ubiquitous technologies and applications. Topics of interest range from areas of classical computer science like mobile databases or security aspects of mobile applications to subjects from business informatics like mobile business processes and mobile commerce.



M. Bick, M. Breunig, H. Höpfner (Hrsg.): MMS 2009

GI-Edition

Lecture Notes in Informatics

Markus Bick, Martin Breunig,
Hagen Höpfner (Hrsg.)

Mobile und Ubiquitäre Informationssysteme – Entwicklung, Implementierung und Anwendung

Proceedings zur
4. Konferenz Mobile und Ubiquitäre
Informationssysteme (MMS 2009)
Münster, 3. März 2009



Markus Bick, Martin Breunig, Hagen Höpfner (Hrsg.)

**Mobile und Ubiquitäre Informationssysteme –
Entwicklung, Implementierung und Anwendung**

**Proceedings zur 4. Konferenz
Mobile und Ubiquitäre Informationssysteme (MMS 2009)**

**3. März 2009
in Münster, Germany**

Gesellschaft für Informatik e.V. (GI)

Lecture Notes in Informatics (LNI) - Proceedings

Series of the Gesellschaft für Informatik (GI)

Volume P-146

ISBN 978-3-88579-240-6

ISSN 1617-5468

Volume Editors

Prof. Dr. Markus Bick

ESCP-EAP Europäische Wirtschaftshochschule Berlin

Email: markus.bick@escp-eap.de

Prof. Dr. Martin Breunig

Universität Osnabrück

Email: mbreunig@uni-osnabrück.de

Dr. Hagen Höpfner

International University in Germany, Bruchsal

Email: hagen.hopfner@i-u.de

Series Editorial Board

Heinrich C. Mayr, Universität Klagenfurt, Austria (Chairman, mayr@ifit.uni-klu.ac.at)

Hinrich Bonin, Leuphana-Universität Lüneburg, Germany

Dieter Fellner, Technische Universität Darmstadt, Germany

Ulrich Flegel, SAP Research, Germany

Ulrich Frank, Universität Duisburg-Essen, Germany

Johann-Christoph Freytag, Humboldt-Universität Berlin, Germany

Thomas Roth-Berghofer, DFKI

Michael Goedicke, Universität Duisburg-Essen

Ralf Hofestädt, Universität Bielefeld

Michael Koch, Universität der Bundeswehr, München, Germany

Axel Lehmann, Universität der Bundeswehr München, Germany

Ernst W. Mayr, Technische Universität München, Germany

Sigrid Schubert, Universität Siegen, Germany

Martin Warnke, Leuphana-Universität Lüneburg, Germany

Dissertations

Dorothea Wagner, Universität Karlsruhe, Germany

Seminars

Reinhard Wilhelm, Universität des Saarlandes, Germany

Thematics

Andreas Oberweis, Universität Karlsruhe (TH)

© Gesellschaft für Informatik, Bonn 2009

printed by Köllen Druck+Verlag GmbH, Bonn

Vorwort

Die von der GI-Fachgruppe Mobilität und mobile Informationssysteme (GI-MMS) ausgerichtete 4. Konferenz Mobile und Ubiquitäre Informationssysteme findet am 3. März 2009 – wieder eingebettet in die BTW 2009 – an der Westfälischen Wilhelms-Universität Münster statt.

Ziel der Konferenz ist es, ein Forum zur Präsentation und Diskussion der umfassenden Aspekte und Fragestellungen mobiler und ubiquitärer Technologien zu bieten. Dabei werden, der Zielsetzung der Fachgruppe MMS folgend, durch die Konferenz die verschiedenen beteiligten Wissenschaftsgebiete zusammengebracht. Dementsprechend wird im Anschluss an die MMS 2008 die breite Gliederung entlang von sechs Schwerpunkten beibehalten:

- Ambient Intelligence in Arbeits-, Lern- und Freizeitprozessen
- Mobile and Ubiquitous Commerce
- Anwendungs- und Datenbank-Entwicklungen für Mobile Informationssysteme
- Das Internet der Dinge
- Technologieentwicklungen und Integrationslösungen für Mobile Mehrwertdienste
- M-Enterprise – mobile und ubiquitäre Geschäftsprozesse

Für die MMS 2008 wurden 18 Artikel eingereicht, von denen nach einem mehrstufigen strengen Begutachtungsprozess neun Beiträge angenommen werden konnten (Annahmequote 50 %). Diese Beiträge sind im vorliegenden Band abgedruckt und werden durch zwei Kurzbeiträge zu laufenden Arbeiten ergänzt.

An dieser Stelle möchten wir uns bei den Autoren und allen Mitgliedern des Programmkomitees der MMS 2009 für ihr Engagement und ihre Unterstützung bedanken.

Berlin, Osnabrück und Bruchsal, im Februar 2009

Markus Bick, Martin Breunig, Hagen Höpfner

Organisation

- Prof. Dr. Markus Bick (ESCP-EAP Europäische Wirtschaftshochschule Berlin)
- Prof. Dr. Martin Breunig (Universität Osnabrück)
- Dr. Hagen Höpfner (International University in Germany, Bruchsal)

Programmkomitee

- Prof. Dr. Michael H. Breitner (Universität Hannover)
- Prof. Dr. Christian Bunse (International University in Germany, Bruchsal)
- Prof. Dr. Rony G. Flatscher (Wirtschaftsuniversität Wien)
- Dr. Christoph Gollmick (SAP AG)
- Prof. Dr. Thomas Hess (LMU München)
- Prof. Dr. Thomas Kirste (Universität Rostock)
- Prof. Dr. Birgitta König-Ries (Universität Jena)
- Dipl.-Kfm. Tyge-F. Kummer (ESCP-EAP Europäische Wirtschaftshochschule Berlin)
- Prof. Dr. Franz Lehner (Universität Passau)
- Prof. Dr. Rainer Malaka (Universität Bremen)
- Dr. Essam Mansour (International University in Germany, Bruchsal)
- Dr.-Ing. Holger Meyer (Universität Rostock)
- Juniorprof. Dr. Daniela Nicklas (Universität Oldenburg)
- Prof. Dr. Alexis Papathanassis (Hochschule Bremerhaven)
- Prof. Dr. Jan M. Pawlowski (University of Jyväskylä)
- Prof. Dr. Kai Rannenberg (Goethe-Universität Frankfurt)
- Prof. Dr. Thomas Ritz (Fachhochschule Aachen)
- Prof. Dr. Jörg Roth (Fachhochschule Nürnberg)
- Prof. Dr. Günther Specht (Universität Innsbruck)
- Dr. Frederic Thiesse (Universität St. Gallen)
- Dr.-Ing. Can Türker (ETH Zürich)
- Dr. Arno Wilfert (PricewaterhouseCoopers)
- Dr. Karsten Weroneck (Fraport AG Frankfurt)

Inhalt

Konferenzbeiträge

Mohamed Hamdy, Birgitta König-Ries

Leader Election Modes of the Service Distribution Protocol for Mobile Ad Hoc Networks 11

Yann Michel, Annika Hinze

Traditional Pre-fetching and caching of limited use for mobile applications 25

Katharina Hahn, Heinz F. Schweppe

Exploring Transactional Service Properties for Mobile Service Composition 39

Michael Decker

Ein Überblick über Ansätze zur Vermeidung der Manipulation von Ortungsverfahren 53

Klaus Kornmeier

Endkundenakzeptanz mobilkommunikationsbasierter Zahlungssysteme 67

Tyge-F. Kummer, Markus Bick

Kausalanalytische Untersuchung von Akzeptanzproblemen ambienter Technologien zur Vermeidung von Behandlungsfehlern in deutschen Krankenhäusern 82

Henrik Mühe, Wolfgang Wörndl, Georg Groh

Realisierung eines dezentralen Recommender Systems für PDAs 95

Britta Fuchs Thomas Ritz

Blended Shopping 109

Andreas Zauner, Holger Hoffmann, Jan Marco Leimeister, Helmut Krcmar

Automotive Software und Service Engineering (ASSE) – Eine Exploration von Herausforderungen und Trends aus Sicht von Branchenexperten 123

Kurzbeiträge

Thomas Wiechert, Andreas Schaller, Frederic Thiesse

Near Field Communication Use in Retail Stores: Effects on the Customer Shopping Process 137

Ulrich Bretschneider, Jan Marco Leimeister, Helmut Krcmar

RFID-Systeme für Messen und Konferenzen: Untersuchung der Einflussgrößen auf die Besucherakzeptanz 142

Leader Election Modes of the Service Distribution Protocol for Ad Hoc Networks

Mohamed Hamdy and Birgitta König-Ries
Institute of Computer Science, Friedrich-Schiller-University Jena,
07743 Jena, Germany
{hamdy|koenig}@informatik.uni-jena.de

Abstract: Ad hoc networks are characterized by high dynamics in particular with respect to the formation of network partitions. This behavior makes it difficult to (s)elect a suitable leader for a given network partition at any given time. However, such leaders or representatives are needed by many protocols for ad hoc networks. In this paper, we will illustrate the leader election process and its influence on protocol performance using the example of the service distribution protocol. Effects of different modes of leader (s)election on the replica placement (allocation) process are investigated as well as the general service availability and success ratio of the protocol.

1 Introduction

Mobile ad hoc networks (MANETs) are a kind of mobile network in which mobile nodes (hosts) move freely anywhere, join/disjoin the network, and communicate with each other without any centralized administration or help. If two mobile nodes are close enough, a wireless link between them is formed. Due to nodes' movements, the topology of the network is ever-changing. Thus, the network usually consists of a number of distinct network partitions. This partitioning behavior makes some/all of the network resources inaccessible for the network participants. In order for the decentralized network to function, the mobile nodes are supposed to share their resources in a collaborative manner in order to achieve the basic required network functionalities like routing and data transportation. However, not only computing power, memory buffers, link utilization, and data should be considered as resources, but also functionality offered by the nodes. Recently, service orientation has become a popular paradigm for sharing functionalities among connected peers. There, the functionalities to be shared are represented as services. These services are provided by providers to all interested clients. Considering the ad hoc network partitioning, however, access to any given service on the provider sites could not be guaranteed. This, of course, affects deeply the service execution between the mobile nodes and needs to be ameliorated, if service orientation is to be used as a basis for functionality sharing in ad hoc networks. Replication of resources is a classical approach to increase the availability and is successfully being used, e.g., in distributed database management systems (DDBMS), or domain name servers (DNS). Unfortunately, none of the approaches developed for these domains can be directly applied for service replication on ad hoc networks.

What we basically want to achieve is the following behavior: Functionality is offered in the network in the form of replicable services. Examples for such services are, e.g., gateway services to other networks (i.e. Internet and GPS), and ad hoc key distribution and certificate issuing [HKR08b]. For simplicity of presentation, we regard just one service being deployed on the network. We assume that all nodes are basically able to run and cache the service albeit at different costs. Initially, the service is running on just one node. Each time when a new partition is about to be formed, a replication mechanism should ensure that there will be a service replica inside this newly formed partition. The created replica should be placed on a selected node, which should be the optimal place to receive and operate the replica according to some criterion. In distributed systems [SSG02, Tel00, Kau96, Toh01] the question of where to place the new replica is answered in the context of the leader election issue.

The leader election problem in MANETs is difficult to address as the topological and resource properties are always changing. Most of the service replication protocols employ proactive leader selection approaches based on partition prediction approaches. When two or more network partitions are being merged, a leader is elected from the concurrent services in the newly formed partition to transform the multiple independent copies of the service into one synchronized service.

Most hierarchical routing protocols for MANETs like CEDAR [SSB99] and ZBR [DHY03] address the problem of leader election; mostly rich-resourced nodes are selected to host information used to produce the required routes by the nodes inside the clusters of these leaders. The TORA [PC97] routing protocol represented the milestone for the work of [MWV00] to introduce a self stabilizing leader election approach in two modes of single and multiple topological changes. In [DB08], another self-stabilizing leader election algorithm based on TORA is presented for which it has been proven that it ensures that a network partition can converge to a legitimate state within a finite time even if topological changes occur during the convergence time. [DE08] proposes a hierarchical leader election protocol which clusters the network into balanced clusters, builds a spanning tree of the locally selected cluster leaders, and ends with the election of a super-leader. [VDI⁺03] considers the problem of trust among the participants in many applications and introduces a set of secure leader election algorithms for the two synchronous and asynchronous modes for MANETs.

Even more directly related to our work are service replication protocols for MANETs. [DJ07, Jus05] introduce mechanisms for mobile service replication by the initial service. A node moderates the replication process per predicted partition, replicas in the new formed partitions are supposed to be hosted by a powerful elected node. For consistency, a master node is kept in order to keep services synchronized among the partitions. The replication process is based on the assumption of availability of a global view for the current network status by all of the participants. [MHKS08] proposes a replication approach for mobile service which is modeled as a validated Petri net model for the dependability. A use case is illustrated to show how the approach configuration parameters could be determined and optimized. In [DBB05, DB05], by estimating the link quality with a partition prediction mechanism based on TORA, the authors supply two mechanisms for pull based replication; (a) replication (pre-partition formation) and (b) merging (after two partitions

merged) mechanism. [WL02] presents a group of algorithms that enable service replication and replica placement selection based on partition prediction and server selection mechanisms by both clients and servers using a distributed grouping mechanism at regular service discovery intervals. [HSC01], based on the concepts of link quality (robustness), evaluates the replication decision, without being involved with the velocity of the underlying network or location information. Based on the available information from the routing component, the network partitioning is not considered as the only trigger of the replication process but also required quality requirements like avoiding long routes in the high density network. As an important result, this could involve the process of replica placement and the related leader election involved in the replication decision which means that finding a good place (node) to be elected could be a trigger for the replication process. A survey on the required quality parameters which affect efficiency of the service execution over a MANET is presented in [RT06]. These parameters are defined and evaluated to a typical efficiency matrix.

In most of the approaches described above, both partition prediction and leader election require a global view of the network and/or querying the lower network layers. The first is rather unrealistic to achieve in a MANET, the latter has several drawbacks: (a) coupling to the lower layers of the network makes these approaches network architecture dependent, (b) querying specific lower layers implies a coupling to specific network components and their architecture and (c) enquiring and processing of the lower network layers is expensive. Additionally, most of these approaches assume the same importance for the service over all of the network operation time and do not take into consideration that the importance of the service is variant over time and may be different for different clients.

In order to avoid the previously mentioned deficiencies of these approaches, in our prior work, we have introduced the service distribution protocol (SDP) [HKR08c, HKR08a]. This protocol operates on the application layer, only. It utilizes the popularity of a specified service, without any partitioning behavior prediction as an indicator for the need for replication. When common interest (*Gross Interest*) for a specified service by a certain group of clients reaches a threshold, the interesting service is replicated. On the other hand those services having lower interest will be shut down (hibernated). The tightly coupled two mechanism of replication and hibernation guarantee higher service availability and better performance for service execution over MANETs, see [HKR08b] for details. Based only on the amount of interest in a service, new providers will be chosen to host replicas of this service. This choosing of providers amounts to a sort of leader election. Since the network topology is ever-changing, these new providers are likely to travel to other partitions and prevail their replicas. In a given network partition, the hibernation mechanism dominates the leader election process. In case of concurrently running replicas, those replicas that are cheapest (according to some measure) from the clients' point of view, will get more requests (interest by the clients in this partition). After a while, the hibernation mechanism will shut down all of the not interesting replicas and only one replica will continue. Again, this is implicitly a leader election process.

In this paper, we show different modes for leader election of the SDP and investigate how the general performance of this protocol is affected by the election process. Especially, we are interested in the replica allocation (placement) process. For that purpose, two

use scenarios are drawn which differ in the service requesting behavior. An extensive simulation approach was employed to investigate the effects for these different use cases. The rest of the paper is organized as follows: An overview of SDP is presented in Section 2. Concepts and consideration about the process of leader election by the protocol are presented in Section 3. The used service model and the relating calling model and how they shape the service gross interest is discussed. Two very different usage scenarios are drawn in Section 5. The used network model is proposed in Section 6. Finally, the simulation and results are discussed in Section 7.

2 The Service Distribution Protocol for MANETs

The main measurement behind the replication decision in the service distribution protocol is the service interest. Interest of a group of clients regarding a specified service could be measured based on many criteria. For example, the number of requests gained by a service provider during a certain time interval, the size of transmitted data from/to a service provider, and/or the call-length could be indicators for the interest in a service at a given time. The protocol employs the frequency of calls (requests) in a certain interval of time as the measure of interest in the service. If a client achieves a certain number of calls per given time interval, it will ask for its own replica of the service. On the other hand, if a service provider does not receive at least a given number of calls (requests) during a certain interval it will shut down (hibernate) its service. The hibernated service may be cached for later activation (restoring). Of course, both the replication and caching decisions are built on the availability of resources on the host side. In this work, we make the assumption that sufficient resources are available on the mobile node to host the service. Performing these processes of replication/hibernation/caching/restoring ensures service prevalence over all network partitions. The protocol consists of three main mechanisms: *Replication*, *Hibernation* and *Caching*:

Replication Mechanism: Based on an optimistic replication model [MPV06], the functionalities of this mechanism are shared between both the service provider and the service clients as follows:

On the provider side:

- **Pass a replica:** Allows the provider to pass a replica to a new (proposed) provider and keep track of information about this replication.
- **Check Correctness:** Enables the service provider to check the correctness of a new replica placement as discussed in the following sections.
- **Publish:** Enables a provider to publish the new status of a service (active/hibernated).

On the client side:

- **Find least cost service:** The service cost is an index labeling each of the service/replicas, clients are supposed to find the least cost service and call it. This index represents a service request-cost which should come out of a combination of criteria.

- **Count and ask:** For simplicity, the frequency of calls (request) has been selected to represent the interest of a client in a specific service. This functionality enables a client to count the number of calls regarding a specific service and if this number achieves a certain value (replication threshold), it allows the client to ask for its own replica of that service.
- **Switch into a provider:** After receiving a replica, this mechanism enables the client to activate and publish the status of its new replica.

Hibernation Mechanism: This is triggered by a service provider which detected a loss of interest regarding his active service. It includes the following functionalities:

- **Count and decide:** For a given time interval, the number of requests to the provider by its group of clients is counted. If this number is not sufficient to let this service be hibernated.
- **Cache it:** Enable the service provider to cache its hibernated service based on its available resources.
- **Publish:** Publishes the new hibernated status of the deactivated service/replica.

Caching Mechanism: This includes the required functionalities to manage the cached service/replicas by an ex-provider, it includes the following functionalities:

- **Restore a service:** Restores a hibernated service/replica from the cache when the current node (as a client) achieves the replication threshold (instead of activating the replication mechanism) or based on the unavailability of the requested service (which is also cached) in the current partition. In cases where the current node restored a replica while there exist another active replica on another node (provider) site, after a while and based on the type of the used hibernation mode (mentioned in Section 3, a selection process takes place and one of these replicas is continued and the rest (if there is more than one active replicas in the same partition) will be hibernated.
- **Publish:** Publishes the status of the restored service/replica.

3 Leader Election Modes of the Service Distribution Protocol

At a certain time the service may be replicated over more than one hosting mobile node within the same network partition. This produces a set of equivalent concurrently running replicas. Based on the hibernation mechanism, after a given time interval (hibernation threshold), at each of the concurrent provider sites, a check for the gained interest will take place. Since all of the replicas' clients are oriented to select the least-cost per request replica in this network partition, just one replica will gain enough interest and will continue. The rest will be hibernated and cached. The main parameter in this selection process besides the request cost (which steers the common interest to a specific service) is the length of the given time interval to perform the hibernation checks and decisions. This interval is supplied by the hibernation threshold. The hibernation threshold is formed

as (number-of-calls:hibernation-time-interval). The length of the hibernation time interval determines the leader election modes of the service distribution protocol: Longer lengths of the hibernation check interval leads to more simultaneously running replicas inside a given partition and vice versa. In the following, we will investigate different settings for this time interval and the implications they have on protocol performance in more detail.

3.1 Long-election Mode of Hibernation

In this mode, the hibernation time interval of the hibernation threshold is fixed determined to be either one or the multiple of the calling (requesting) time unit. In this work, the calling time unit is a minute. This mode enables replicas to remain activated for a longer time to be evaluated. We assume the long-election to be helpful for very high mobility networks, in which the frequency of network partition formation is very high. In the work of [HKR08b], effects of different (multiplies of this long mode hibernation check time intervals) are applied and analyzed.

3.2 Short-election Mode of Hibernation

The basic idea in this mode is to evaluate the initial interest regarding a given (just published) service/replica as soon as possible. This means evaluation should take place after the minimum possible time required by the service provider to receive, activate, publish, and wait for a slice time for calls. Once this very tight time interval ends, the hibernation test takes place. If the service received at least a given number of calls, this means the service will continue, any other active service will be hibernated. This election mode is very conservative regarding the number of concurrent replicas inside a network partition.

4 Service Model and Gross Interest

In this research, the proposed deployed service is an abstract service. We assume that the pre-required resources for this service are available by all of the mobile nodes, this means, all of the network nodes can participate in the replication process. A "Requirement index" is describing each of the replicas. The requirement index represents both how much resources are supposed to be available by a client node to be eligible to ask for its own replica from a provider node and the cost of the request of this service. In reality, although all services are equivalent, not all available concurrent services should be provided with the same cost. All candidate client nodes are supposed to search for a service with the minimum available requirement index in the network partition.

4.1 Calling Model

The calling model describes the behavior of the service requesting by the clients. As in [HKR08c, HKR08a], The frequency of calls (requests) and the time of call length are the main parameters indicating the proposed calling model. For simplicity, the proposed calling model is based on only the frequency of calls. Initially, all of the created nodes seek for the initial (original) service provider node, only those nodes with at least one feasible path to the provider node are assumed to start evaluating the service calling and be involved in the related replication-hibernation-caching processes. After a while, service/replicas prevalence through the network is supposed to cover as much as possible of the ad hoc formed partitions [HKR08a]. A variant calling rates are maintained by each node, the calling rate is generated between $[0..Max-Rate]$ calls per minute, The "Max-Rate" indicates the number of client's groups maintained by the network, for example if Max-Rate = 3 calls per minutes then there are four client's groups (0 calls/minutes, 1 calls/minutes, 2 calls/minutes, and 3 calls/minutes) are supposed to be presented in the network. The calling rate is supposed to be constant during a calling period, the calling period is supposed to be a "Calling-Length" and its multiplications between $[1..Max-Calling-Units]$ in minutes, and after a pause time equals to a "Pause-Length" or its multiplications $[0..Max-Pause-Units]$ in minutes the node is supposed to select another calling rate and so on. All of calling rate, calling period, and pause period are uniformly randomly selected. The values of the $\{Calling-Length, Max-Calling-Units\}$ and $\{Pause-Length, Max-Pause-Units\}$ are given and evaluated as in the following sections.

4.2 Gross interest

Based on the calling behavior of a specific set of the service clients, the service will be replicated and prevail over the different network partitions. The different settings to the proposed calling model will generate different types of calling patterns and leads to a resultant gross interest regarding a specific service at certain time. Richer Gross interest indicated by higher calling frequencies with shorter pause intervals will require shorter time to produce better service prevalence over the network, on the other hand, poorer gross interest will generate the opposite situation.

5 Scenarios

In this section, two extremely different, rich and poor gross interest, scenarios are developed to reflect the concept of the gross interest, as mentioned in Section 4.2, moreover, a discussion about quantifying the difference between these two scenarios is presented in the following subsections.

Scenario 1: Rich Interest Scenario: An active calling behavior with short periods of pausing is modeled in this scenario. The mobile node has a *calling-length* equal to 10

minutes and the *max-calling-units* equal to 5 units. This means that the calling period could be either 10 minutes, 20 minutes ..., or 50 minutes. On the other hand, each of the mobile nodes is assumed not to call (wait) for a pause period. The pause period settings are as follows; the *pause-length* equals to 3 minutes and the *max-pause-units* equals to 5 units. So, the pause periods will be 0 minutes, 3 minutes ..., or 12 minutes [HKR08b].

Scenario 2: Poor Interest Scenario: The second scenario presents a lazy calling behavior with long periods of pausing. The *calling-length* is set to 2 minutes, the *max-calling-units* is set to 3 which means that the clients have 2 minutes, 4 minutes, or 6 minutes of calling periods. The *pause-length* equals to 10 minutes and the *max-pause-units* equals to 5 units, which means that the pause periods should be 0 minutes, 10 minutes, ..., or 40 minutes [HKR08b].

Counting Active Clients: Active clients are those clients which are being in a calling period as in Section 4.1. The expected number of the active clients in at certain time in a given network partition is related to the calling model. At certain time (t), the expected number of active clients (Ex) [HKR08b] ,

$$Ex = Pz_i \cdot \frac{Ex_n}{Ex_n + R \cdot Ex_m} \dots (1)$$

where (Ex_n) is the expected value of the Max-Calling-Units, (Ex_m) is the expected value of Max-Pause-Unites, and R is the ratio between a single "Pause-Length" to a single "Calling-Length" (see Section 4.1), and (Pz_i) is the " i^{th} " network partition size. Based on computing the (Ex) value, two extremely different scenarios are proposed to be used in this work evaluation.

Quantify the Gross Interest Difference Between Scenarios: Assume a partition with size equals to 25 mobile host, this leads to, for scenario 1, about 20 mobile host could be active client a time, which is a really high number of clients, on the other hand, for scenario 2, just about 4.1 clients will be active. As one of our experimental results, we computed the weighted sum (weight: size of a partition to the network size) of our resultant network partitions during our evaluations using the proposed specifications of the introduced mobility model, we found that for a network size equals to 70 node, the expected partition size will be about 17 node. Again, For scenario 1, there will be about 14 active client, and for scenario 2 there will be just about 2.8 active client nodes [HKR08b].

6 Network Model

The used network in this work is modeled at certain time t as an undirected unweighted graph $G(N, E)$, where N are the mobile nodes and E the groups of formed links between the participant nodes. The network is formed of a number of disjoint network partitions. $G_x(N_x, E_x)$ presents a specified network partition x , where $G(N, E) = G_1(N_1, E_1) \dots \cup G_x(N_x, E_x) \dots \cup G_k(N_k, E_k)$, $N = N_1 \dots \cup N_x \dots \cup N_k$, and $E = E_1 \dots \cup E_x \dots \cup E_k$. The wireless transmission range (R) of the mobile nodes is fixed during the whole operation time (75 meter). The network is placed in a square area

(600by600m²) [HKR08a]. A wireless like if formed if and only if two neighbor mobile nodes are located inside the transmission range of each other. The random way point mobility model [LNR04] is the movement model of the mobile nodes. Each mobile node picks a constant speed between $[0, V_{max}]m/s$ (applied:[0..6]m/s), then chooses a random location to go, after reaching its desire location, it is supposed to wait for a pause time between $[0, P_{max}]$ (applied:[0..15]minutes) and so on.

7 Simulation

A detailed simulation has been elaborated by this work, a rather sophisticated simulator has been built for that using C++¹. The general performance parameters namely service availability and success ratio have been discussed. The main parameters describing the quality of the replica placement process is presented and evaluated in this section.

7.1 Performance Parameters

Service Availability indicates the ratio between the summations of the whole time intervals that the service was active some where in the network, but it does not indicate if the service was accessible or not for the whole of the network participants. the importance of this parameters is that it reflects the presence of the service anywhere to be replicated and prevailed over the whole network partitions.

Success Ratio indicates the ratio between the number of succeeded request to the service/replicas overall the network. It presents the satisfaction of the clients requests by the current service distribution in the network.

Correctness of a Replica placement: One of the main issues in this research, as in [HKR08b], is to measure the correctness of the replica placement ratio which indicates the available number of the concurrent active replicas inside a given partition at certain time, moreover, this research aims to reflect also the effects of the proposed leader election modes on the service distribution protocol replica placement process. [HKR08a] suggested a linear replica allocation correctness. [HKR08b] introduced a detailed discussion on the linear computation method of the allocation correctness and concluded to another computation method (WRCR: Weighted Rational Correctness Ratio) for the ratio based on both of the weight of a given network partition to the whole network size and the number of the concurrent active replica inside it on a rational basis. The following equation shows the $RCCR_t$ (Rational Correctness Ratio) relation at given time (t) for the i^{th} network partition (P_i):

$$RCCR_t(P_i) = \begin{cases} 0 & R_i = 0 \\ 1 & R_i \in \{1, 2\} \quad \dots (2) \\ \frac{3}{R_i} \cdot \frac{Pz_i - R_i}{Pz - 2} & R_i > 2 \end{cases}$$

¹Microsoft Visual C++, version 8.50727

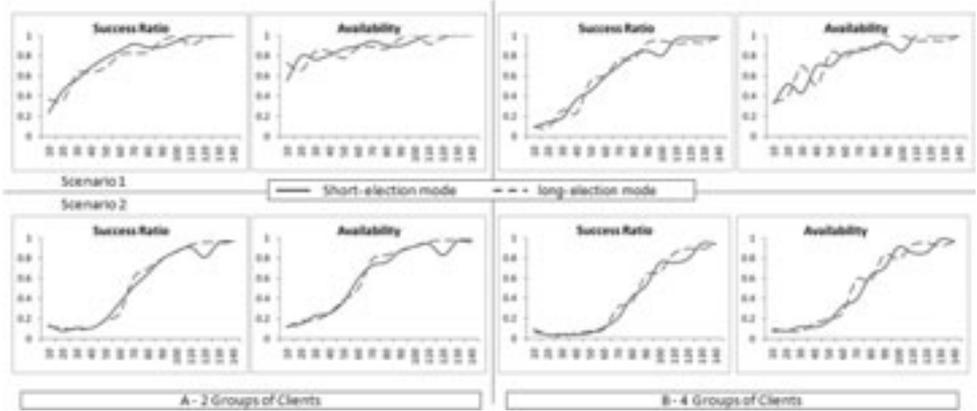


Figure 1: Success ratio and service availability (0-1) against the network size (10-140 node) for both of scenario 1 and 2 for (A)2 client groups and (B)4 client groups

where R_i is the number of the concurrent active replica inside (P_i), and (P_{z_i}) is the partition size of (P_i) and WRCR is as follows:

$$WRCR_t = \sum_{i=0}^n \left(\frac{P_{z_i}}{\text{size}(N)} \cdot RCR_t(P_i) \right) \dots (3)$$

where n is the number of the partitions at given time (t), and N is the network as defined in Section 6. In this research the WRCR is used as a measurement the correctness of the replica allocation process.

Typical Partition of the Network: The typical partition represent the most weighted partition regarding the size in the network, which can provide a good network partition sample. The Typical partition is a feature of a given network mobility model settings. The typical partition is used as in this work in order to analyze the number of replicas inside it. In fact it is a kind of computed (virtual) partition rather a physical one, it is computed as $Typ.Part_N$:

$$Typ.Part_N = \sum_{i=1}^n \left(P_{z_i} \cdot \frac{P_{z_i}}{\text{size}(N)} \right) \dots (4)$$

7.2 Configurations and Settings

The simulation run time for each experiment was set to be 2 hours, the network size is varying from 10 to 140 mobile hosts. Results come out of the average of 20 times runs. The number of client groups is either 2 groups(maximum call rate 1 call / minute) or 4 groups (maximum call rate 3 call / minute). The replication threshold for a client is achieving number of calls (requests) equals to the maximum call rate in a given minute (Max-call-rate:1), the hibernation threshold for a provider is gaining less than one call(request) in a

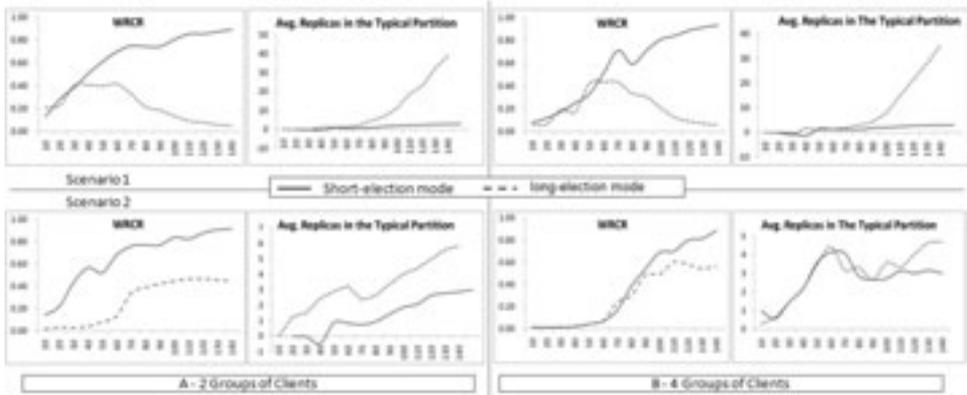


Figure 2: WRCR ratio (0-1) and Avg. number of replicas in the typical partition against the network size (10-140 node) for both of scenario 1 and 2 for (A)2 client groups and (B)4 client groups

given length of minutes, in this research, regarding the previously introduced two election modes, the hibernation threshold will be either at least one call (request) per a minute (1call:1minute, indicates the long election mode) or one call per the minimum waiting time interval shown in subsections (3.1,3.2), in our simulation we consider this required time interval to be 1 second [HKR08b].

7.3 Results

Figure 1 shows the general performance of the service distribution protocol, for both rich and poor interest scenarios, the service availability increases as the network becomes more dense, this leads to the same behavior on the success ratio since a higher success ratio (indicates the number of succeeded requests in the network) is tightly coupled to a higher service availability. No difference in the results come from the different election modes, in fact this can be concluded as a positive point for the short election mode: Shorter hibernation check intervals do not affect the high service availability and success ratio. As the number of the client groups increases both the service availability and accordingly the success ratio decreases, this is due to the smaller set of the interesting (in hosting replicas) clients in the higher number of the client groups. Table 1 shows the values of the average service availability and success ratio with the relative observed standard deviation for both scenarios, hibernation modes and number of client groups.

Figure 2 shows both the observed WRCR ratio and the average number of concurrent replicas in the typical partition in the network regarding varying the network sizes. For Scenario 1 and for 2 client groups, as the network increases, the WRCR ratio increases too, starting from 70 node network size WRCR is about/more than (0.80), as the client groups increase, the WRCR ratio growing rate becomes slower. This is due to the lack of service prevalence in the different network partitions. The number of partitions with

Scenario1	2 Groups	SM	SuccessRatio		Availability	
			Avg.	Avg.StD	Avg.	Avg.StD
			0.78	0.20	0.88	0.19
Scenario1	4Groups	LM	0.81	0.15	0.90	0.16
		SM	0.64	0.17	0.79	0.23
Scenario2	2 Groups	LM	0.64	0.17	0.80	0.24
		SM	0.53	0.14	0.62	0.16
Scenario2	4Groups	LM	0.55	0.11	0.64	0.13
		SM	0.41	0.13	0.52	0.17
		LM	0.42	0.12	0.53	0.17

Table 1: Average (Avg.) success ratio and service availability with the average value of the observed standard deviation (Avg.Std.) for both Short and long election modes (SM, LM) for both scenarios and client groups

no service coverages is increased because of the lower number of clients interesting to host a replica. The same behavior is presented by the curves of the short election mode for Scenario 1. On the other hand, in the long election mode, for both numbers of client groups, the observed WRCR ratio starts increasing with the increasing network sizes til an interval of network sizes (about [30,60], and [50,70] nodes for (2,4) client groups respectively) then the WRCR ratio starts decreasing as the network size increases. This behavior could be interpreted by the curves of the average replicas in the typical partition. For 2 client groups, the average number of concurrent running replicas is slightly growing til about network size of 70 nodes (less than 5 replicas in the typical partition). Then it grows quickly to which directly affects the observed WRCR ratio. In the 4 clients groups the WRCR collapses slower, because the average number of concurrent running replicas is less then in the 2 groups case. In fact, in the high density network (large network sizes), the typical partition size is mostly about the network size [HKR08b], and in cases of high gross interest and lower replication threshold, as in Scenario 1 with 2 client groups, the number of the interested clients to host replicas is very high and produces a higher number of concurrent replicas. In Scenario 2, in cases of poor gross interest, the protocol conserves higher WRCR ratio as the network size increases for both of the hibernation modes, increasing the number of the client groups affects negatively the growth of the WRCR ratio. Generally, applying the short hibernation election mode decreases the number of the concurrently running replicas in all scenarios and client groups (about 3 replicas at most at a time). This result is pretty good because, by applying such an election mode, we can ensure a very limited number of replicas inside the network (typical) partition at a time.

8 Conclusions

The main contribution of this work is to highlight how the service distribution protocol can converge to the optimum replica distribution situation during the life time of a mobile ad hoc network based on the proposed leader election modes. As mentioned, the protocol success and unique key is using just the available information about the service use of both service provider and clients in a service oriented environment. By utilizing these information and the service prevalence by mobility and replication actions, the protocol ensures higher service availability and querying success ratio although the partitioning behavior

of the ad hoc network. This paper provides a solution to how the leader election process could be met by the service distribution protocol during a continuous election process based on variant service interest generated by variant gross interest regarding a specific service. Two time based leader election modes, namely short and long election modes, are introduced for the service distribution protocol for MANETs. The two modes have been evaluated on two extremely different use scenarios which have been shown and quantified in Section 5. Then they have been evaluated using a detailed simulation. Results show that the short election mode can ensure a very limited number of concurrently running replicas from both the whole network and a network partition point of view. Although the long mode election can achieve high performance metrics regarding service availability and success ratio, it produces a relatively high number of concurrent replicas in the network. On the other hand, the short election mode can also achieve a similar service availability and success ratio with relatively constant number of concurrent replicas inside the network and its partitions. While these results were obtained for the SDP, we believe that they also provide some guidance for similar decisions in other protocols - at the very least they show the need for a careful examination of the choice of time interval for decision making in leader election.

References

- [DB05] Abdelouahid Derhab and Nadjib Badache. A pull-based service replication protocol in mobile ad hoc networks. *European Transactions on Telecommunications*, 18:1–11, 2005.
- [DB08] Abdelouahid Derhab and Nadjib Badache. A Self-Stabilizing Leader Election Algorithm in Highly Dynamic Ad Hoc Mobile Networks. *IEEE Transactions on Parallel Distributed Systems*, 19(7):926–939, 2008.
- [DBB05] Abdelouahid Derhab, Nadjib Badache, and Abdelmadjid Bouabdallah. A Partition Prediction Algorithm for Service Replication in Mobile Ad Hoc Networks. In *Second Annual Conference on Wireless On-demand Network Systems and Services (WONS 2005)*, Swi, 2005.
- [DE08] Orhan Dagdeviren and Kayhan Erciyes. A Hierarchical Leader Election Protocol for Mobile Ad Hoc Networks. In *8th International Conference on Computational Science (ICCS 2008)*, pages 509–518, Kraków, Poland, 2008.
- [DHY03] H. Du, H. Hassanein, and C. Yeh. Zone-based routing protocol for high-mobility MANET. In *Canadian Conference on Electrical and Computer Engineering*, 2003.
- [DJ07] Schahram Dustdar and Lukasz Juszczyk. Dynamic replication and synchronization of web services for high availability in mobile ad-hoc networks. *Service Oriented Computing and Applications*, 1:19–33, March 2007.
- [HKR08a] Mohamed Hamdy and Birgitta König-Ries. An Extended Analysis of an Interest-Based Service Distribution Protocol for Mobile Ad-hoc Networks. In *Proc. of the International Conference on Wireless Information Networks and Systems (WINSYS 2008)*, Porto, Portugal, 2008.
- [HKR08b] Mohamed Hamdy and Birgitta König-Ries. Service Availability, Success Ratio, Prevalence, Replica Allocation Correctness, Replication Degree, and Effects of Different Replication/Hibernation Behavior Effects of the Service Distribution Protocol for Mobile Ad Hoc Networks -A Detailed Study-. Technical report, Institute of Computer Science, Friedrich-Schiller-University Jena, Germany, December 2008.
- [HKR08c] Mohamed Hamdy and Birgitta König-Ries. A Service Distribution Protocol For Mobile Ad Hoc Networks. In *the proceedings of the International Conference on Pervasive Services (ICPS 08)*, Sorrento, Italy, 2008.
- [HSC01] Michael Hauspie, David Simplot, and Jean Carle. Replication decision algorithm based on link evaluation for services in MANET. Technical report, University of Lille, France, 2001.
- [Jus05] Lukasz Juszczyk. Replication and Synchronization of Web Services in Ad-hoc Networks. Master’s thesis, Technischen Universität Wien, 2005.
- [Kau96] Morgan Kaufmann. *Distributed Algorithms*. Nancy A. Lynch, 1996.

- [LNR04] Guolong Lin, Guevara Noubir, and Rajmohan Rajaraman. Mobility Models for Ad hoc Network Simulation. In *The 23rd Conference of the IEEE Communications Society (INFOCOM 2004)*, 2004. College of Computer & Information Science Northeastern University Boston.
- [MHKS08] Erling Matthesen, Ossama Hamouda, Mohamed Kaaniche, and Hans-Peter Schwefel. Dependability Evaluation of a Replication Service for Mobile Applications in Dynamic Ad-Hoc Networks. *Lecture Notes in Computer Science*, 5017/2008:171–186, 2008.
- [MPV06] Vidal Martins, Esther Pacitti, and Patrik Valduriez. Survey on data replication systems in P2P systems. Technical report, INRIA, Rennes, France, 2006.
- [MWV00] N. Malpani, J. L. Welch, and N. Vaidya. Leader Election Algorithms for Mobile Ad Hoc Network. In *Proceedings of the 4th international workshop on Discrete algorithms and methods for mobile computing and communications*, pages 96–103, 2000.
- [PC97] Vincent D. Park and M. Scott Corson. A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks. In *INFOCOM (3)*, pages 1405–1413, 1997.
- [RT06] Pablo Rossi and Zahir Tari. Software Metrics for the Efficient Execution of Mobile Services. In *The Proceedings of a workshop on Emerging Web Services Technology (WEWST'06), in conjunction with 4th European Conference on Web Services (ECOWS'06)*, 2006.
- [SSB99] Raghupathy Sivakumar, Prasun Sinha, and Vaduvur Bharghavan. CEDAR: A Core-Extraction Distributed Ad Hoc Routing Algorithm. *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, 17:1454–1465, 1999.
- [SSG02] S. Panchapakesan Shanti S. Gupta. *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations (Classics in Applied Mathematics)*. Society for Industrial Mathematics, 2002.
- [Tel00] Gerard Tel. *Introduction to Distributed Systems*. Cambridge University Press, 2000.
- [Toh01] C-K. Toh. *Ad Hoc Mobile Wireless Networks: Protocols and Systems*. Prentice Hall Publishers, 2001.
- [VDI⁺03] Sudarshan Vasudevan, Brian Declene, Neil Immerman, Jim Kurose, and Don Towsley. Leader Election Algorithms for Wireless Ad Hoc Networks. In *in Proceedings DARPA Information Survivability Conference and Exposition*, pages 22–24, 2003.
- [WL02] K. Wang and B. Li. Efficient and guaranteed service coverage in partitionable mobile ad hoc networks. In *IEEE INFOCOM02*, New York, USA, 2002.

Traditional Pre-fetching and caching of limited use for mobile applications

Yann Michel¹, Annika Hinze²

¹ Freie Universitaet Berlin, ymichel@inf.fu-berlin.de

² University of Waikato, New Zealand, a.hinze@cs.waikato.ac.nz

Abstract: This paper reports our experiences introducing location and context aware caching and pre-fetching for a mobile tourist information system. An extensive evaluation and comparison of selected caching and pre-fetching strategies has been carried out in this mobile environment. We discuss why the consideration of location-awareness is less influential and thus less successful than expected and hoped for.

1 Introduction

The Tourist Information Provider (TIP) is a context-aware mobile tourist information system providing sight information to its users. The available information is filtered according to a user's location and interest profile. Our goal for the research reported in this paper was to lower the response times for mobile clients and to restrict wireless network usage to times of good network quality. We wished to alleviate or even eliminate the effects of disconnectedness so that the user can continue working with the TIP client even though network connection is (temporarily) unavailable. We wanted to retain TIP's characteristics of location-awareness and personalization.

As Kirchner et al. [6] pointed out, pre-fetching in combination with caching may lead to lower response time, based on more locally available data in the cache and less "burst" load on the network due to most of the information being downloaded at times with sufficient bandwidth. On the other hand one needs to consider additional cost at the mobile device: additional CPU load to predict the user's movements and possibly wasted bandwidth in case there was a falsely predicted movement.

This paper reports on our experiences introducing location and context aware caching and pre-fetching for the mobile TIP. We provide details of our extensive evaluation and comparison of selected caching and pre-fetching strategies in a mobile environment. The findings of our analysis are surprising; we will show that the consideration of location-awareness is less influential and thus less successful than expected and hoped for. Our conclusion is that additional information, similar to the one used in recommender systems, will need to be used to support more effective location-aware caching and pre-fetching.

The remainder of the paper is structured as follows: Section 2 discusses strategies for caching and pre-fetching applicable in a location aware mobile context; Section 3 introduces implementation details. The main focus of the paper is on the evaluation reported in Section 4. The details about the TIP system are kept to a minimum to allow for extensive discussion of the evaluation results. TIP has been extensively described, e.g., in [4].

2 Caching and Pre-fetching

Data in a *cache* is stored ready to be used and is replaced when the available space is used.

Replacement strategies suitable for location-based mobile systems Temporal replacement strategies consider timing information of the stored items, such as in *first-in first-out (FIFO)* and *least-recently-used (LRU)*. Semantic replacement in mobile applications could refer to location, such as in the *Manhattan Distance* and the use of *Voronoi Diagrams*. For the Manhattan Distance [3] all cached elements are assigned to a region; this replacement strategy removes all elements within a region and not just a single element (as in LRU). The region with the greatest Manhattan distance between its centre of gravity to the current point is replaced. Voronoi Diagrams [9] aim at accelerating nearest neighbor queries. Direction and current speed of the users is taken into account. Three replacement strategies are based on cached area, on distance, and on common circle pane. Spatial replacement strategies are optimized for location-aware item handling. Two typical strategies are *Furthest Away Replacement (FAR)* and *Probability Area/Probability Area Inverse Distance*. Furthest away replacement [8] uses direction prediction for the next estimated location. Elements to replace are those that are furthest away from the predicted position and that are not in the direction of movement. PA/PAID methods [10] additionally consider *data distance* and *valid scope area*. Data in the larger valid scope area have a higher request probability. The items with the lowest probabilities are replaced.

Pre-fetching Pre-fetching in Linux uses the LRU-One Block Lookahead (LRU-OBL) [5]. It pre-fetches the logical block next to the currently referenced block. The cache is partitioned into two rooms, the *Weighing Room* and the *Waiting Room*, where weights of the stored items are determined and the not yet referenced items are kept, respectively. A critical parameter is the optimal ratio for the sizing of the two rooms. Predictive pre-fetching of data in a mobile environment [2] uses a velocity model based on current location, speed, and direction of the user. For pre-fetching data for a user at a certain location, speed and the direction of the user have to be determined (resulting in a delay). Average speed of the mobile users may be used [6] to smooth the changes of the pre-fetching zone. The higher the user's speed, the larger the pre-fetching zone. In a mobile environment, the next movements of the mobile user must be predicted. Two typical approaches for estimating a user's future position are linear interpolation and Lagrange interpolation.

Combined Location-aware Caching and Pre-fetching: Cao et al. [1] have investigated how to efficiently use caching and pre-fetching without letting both interfere with each other. They propose four rules for optimizing the combination of caching and pre-fetching strategies and two combined usage strategies: A *conservative* strategy simply tries to minimize the elapsed time while performing a minimum number of pre-fetches; and an *aggressive* strategy that always pre-fetches the next block that is not in the cache at the earliest possible moment by replacing the block whose request is furthest away in the future. In our current implementation, we consider these rules by applying the conservative strategy.

3 Implementation

The TIP server is the central information provider for the available data; it consists of several services and a central broker that routes the information. The TIP client resides on the mobile devices; it is the end user software component for travellers. The client consists of a broker and several services; it runs a cache and a pre-fetching service. A *Message Handler* maintains the cache and the stored cache items. If all replies to the incoming request are already cached, they are simply enqueued into the reception queue. If not all replies are locally available, the request is forwarded to the remote broker. Requests are handled as events. An incoming event is entered into the cache and then enqueued into the reception queue. This cache implementation provides an abstract framework for all cache methods used in TIP. As internal container for the cached objects we use a HSQL database¹ that runs in memory or by using a local storage such as a compact-flash or hard-disk. We keep two tables EVENTS and REPLIES. To support location-aware filtering of the stored events, several geo-spatial functions were added to the database. We analysed the following cache replacement policies for the TIP client: LRUcache (LRU replacement strategy) MHDCache (manhattan distance), FARCache (distance and direction).

The system pre-fetches data for locations the user is expected to move to, using the *PrefetchService*. The system distinguishes between prefetched and cached events by different markings: *cached* events were previously requested by the user, e.g., by a GPS-triggered Location event; *pre-fetched* events were requested by the submission of a predicted location. This results in a two-room scheme similar to the one introduced earlier. For the movement prediction of the TIP user, we use linear and Lagrange based prediction. The *PrefetchService* advertises *Location* events and subscribes to all events the *DisplayService* subscribes to. When a user arrives at a certain location and the cache was appropriately sized, the desired information should be locally available already and does not need to be specifically requested from the server.

4 Performance Analysis Setup

We start this central section by describing the evaluation parameter and measurements selection, and the setup of the tests.

4.1 Evaluation Parameters

Cache size: The cache size defines the number of items that can be stored in the cache.

Here we do not specify the overall storage size used but the maximum number of elements that can be stored. This simplification of the cache sizing was chosen because the implemented sample services submit events of similar size. We test caches with 250, 500, 1000 and 2000 elements.

Cache replacement strategy: Three cache replacement strategies can be chosen for TIP: least recently used (LRU), manhattan distance (MHD) and furthest away replacement (FAR) as described in Section 2

User location: This parameter specifies the location of the TIP user. We test three scenarios: triangle, circle and random (see Section 4.3).

¹<http://www.hsqldb.org/>

Pre-fetching strategy: Different future user locations are predicted using “linear pre-fetching” and “Lagrange pre-fetching” with 3, 4, or 5 historic location points, and the disabling of pre-fetching. This predicted location is submitted to the pre-fetching service by event message.

Cache retrieve distance ϵ : The implemented cache allows the retrieval of similar events that are within a distance of ϵ degrees. Here this value is set to zero to test only the underlying cache replacement strategies and not the retrieval of similar events.

4.2 Evaluation Measures

The evaluation focussed on four measures. The first three measures were probed inside the message handler. If an event is a request, the handler checks if the requested information is still cached (if yes, this is a cache hit, otherwise a cache miss).

Cache Hit-ratio The cache hit-ratio is a quality measure for the cache replacement strategy. A high hit-ratio indicates a good re-usage of cached items and therefore a good quality of the underlying cache replacement strategy. A low hit-ratio shows that less stored items could be re-used. A high cache hit-ratio indicates less network requests and therefore less energy consumption; a low cache hit-ratio indicates a larger amount of data that needs to be fetched via the network.

$$\text{cache hit-ratio} = \text{cache hits} / (\text{cache hits} + \text{cache misses})$$

Pre-fetch Hit-ratio If the event that is returned from the cache has the flag *pre-fetched* set to true, this is counted as a pre-fetched event. The flag is now changed to false and the event is returned. If the event was not pre-fetched, it is only counted as a returned event. A high pre-fetch hit-ratio implies that the pre-fetching service already downloaded the desired data; a low pre-fetch hit-ratio does not necessarily mean that the algorithm performs poorly but could also be the result of a high pre-fetch delete-ratio.

$$\text{pre-fetch hit-ratio} = \text{number of pre-fetched events} / \text{number of all events}$$

Pre-fetch Delete-ratio When the cache is full and a new item should be inserted, an already stored item has to be removed. If the determined victim still has the pre-fetched flag set, it was never accessed by the user. We count the deletion of all items, plus the items that were pre-fetched but never accessed by the user. A low pre-fetch delete-ratio indicates that much of the pre-fetched information was really accessed and therefore used. A high pre-fetch delete-ratio indicates that unnecessary downloads were generated and results in a waste of energy. The pre-fetch hit-ratio together with the pre-fetch delete-ratio show how effective the evaluated pre-fetching strategy works with the used cache replacement.

$$\text{pre-fetch delete-ratio} = \text{number of deleted pre-fetched events} / \text{number of all deleted events}$$

Prediction Distance The distance of the predicted location to the actual location measures the quality of the movement prediction. A low prediction distance is a good quality measure for the underlying algorithm. If the prediction distance is high it is likely that the desired information has not been cached on arrival at a future location. The quality measure for the predicted locations is probed inside of the pre-fetch service, which stores the last predicted location.

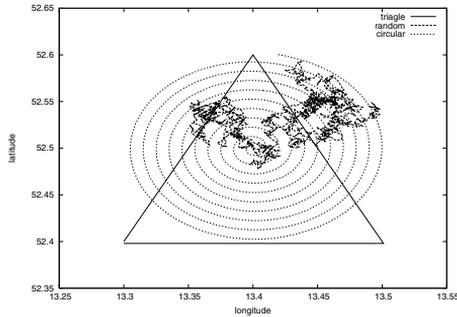


Figure 1: Example tracks for the test scenarios.

4.3 Evaluation Setup

We give an overview of the sample data used and give details about the test scenarios that were used to evaluate the system.

Sample Data The evaluation data are artificially generated but influenced by observed real data patterns. First we generated sample data for sights that can be visited and contains at least a name and corresponding coordinates. The sample-sights are uniformly distributed in the geo-spatial surrounding of Berlin: 13.3 to 13.5 degrees east and 52.4 to 52.6 degrees north. Distance between sights in longitudinal and latitudinal direction is 0.001 degrees (approximately 111m) each. 27,647 sample sight entries were generated.

Discussion of suitability of data and tracks We use artificial test data that are uniformly distributed (as described above) and a number of test scenarios: a forces scenario, a circular track, a triangle track, and a random track (described below). These tracks are somewhat restricted and, taken in isolation, unrealistic mobility models to simulate a user's movement. However, we believe even using these tracks will allow us to make interesting observations. After all, a user's movements will consist of a sequence of movements akin to the patterns evaluated here. We focus on situations where a user follows a seemingly random sequence of these movements. That is, no information about the terrain or other users' walking patterns is available. We will discuss later which improvements could be used if more information is available. We believe that our systematic approach of evaluating walking patterns that may in combination describe a user's track will allow for a general analysis of the evaluated algorithms.

Test Scenarios For the evaluation of the described parameters, three test scenarios were defined. Each of them describes a track (see Figure 1) that a virtual user traverses:

- S1. Triangular Track:** This track describes a linear movement of a user with two turning points at the corners of the triangle. Lines one to five generate the rising edge; lines six to ten generate the falling edge; lines eleven to fourteen generate the edge that returns to the start. Offset parameters specify the step size in longitude and latitude direction.

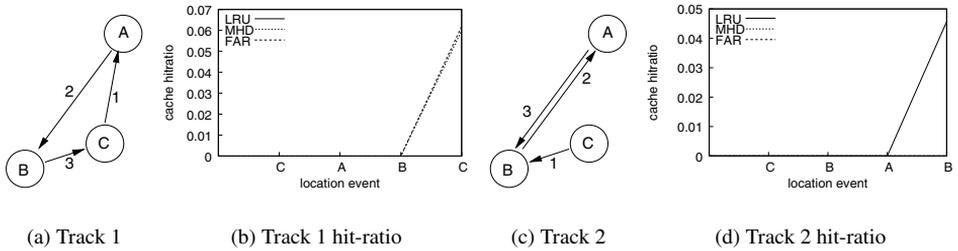


Figure 2: Example Tracks 1 and 2 for demonstrating the different behavior of location-aware and -unaware cache-replacement strategies shown in the hit-ratio graphs.

- S2. Circular Track:** This track describes a spiral movement of the user. The parameters δt and *magnitude* specify the sampling resolution and the distance between the drawn spiral-line respectively. Here, the offset parameters for longitude and latitude move the generated spiral to the center of the pane that was filled with the test data.
- S3. Random Track:** This track describes a randomly chosen track. Generated locations on the defined pane are kept, otherwise discarded. A multiplier is randomly generated between -0.5 and 0.5 . The offset parameters specify the factors in longitude and latitude direction that are used for the computation of the next location.

For the evaluations tracks were generated once and used for every test setup.

5 Performance Evaluation Analysis

For each test we first describe our hypothesis and then compare these to the results and discuss the significance of the results.

5.1 Caching

We first demonstrate the advantages and disadvantages of location-aware caching in a forced scenario. We then test the cache replacement strategies (LRU, MHD and FAR) for each test scenario and different sizes.

5.1.1 Forced Scenario

This test compares the behavior of the location-aware and location-unaware cache replacement strategies. We used two tracks created to be favorable or unfavorable, respectively, for each of the two categories. This scenario will highlight how different caches behave differently under different circumstances depending on the replacement strategy and the track. Each track consists of four location visits (see Figures 2(a) and 2(c)). To force the replacement of stored items, the cache size is restricted to three locations.

The results for Track 1 are shown in Figure 2(b). The location-aware algorithms both

have replaced items for region “A” because the distance to the stored items was furthest away from the current location. The items for the region “C” were retained resulting in a cache-hit when returning to that region. Figure 2(d) shows the results for Track 2. The location-aware algorithms replaced region “B” because it was furthest away from the current location. Consequently, when returning to “C” the items were no longer locally available and had to be re-fetched. The LRU-algorithm replaced the items for the region “C” resulting in a cache-hit when returning to “B”. These results give rise to the expectation that the time-aware algorithms perform better for “returning” users whereas the location-aware algorithms performs better for users moving in “circles”.

5.1.2 Test Scenario 1 (Triangle Track)

We expected the LRU- and the MHD-based caches to perform similarly because, except for the two turning points, the track describes a linear movement. The time-based replacement of LRU algorithm replaces the same items as MHD based on distance. The FAR-based algorithm is expected to perform similar until the first turning point has passed. Then we expect a temporary decrease in performance since it does not only consider the distance of the items but also the movement direction.

Only one example result is shown in Figure 3(a); as the results were almost identical for all four cache sizes tested (250, 500, 1000, 2000). Only cache size 1000 had a slight divergence in the hit-ratio for the FAR-based cache: The cache hit-ratio decreases after the first turning point of the triangle track but then shows a slightly increasing trend afterwards. The reason is that the victim chosen at the first turning point is disadvantageous to the cache hit-ratio.

Overall, the algorithms performed as expected. However, it was surprising to find that the cache size does not have any significant effect on the hit-ratio of the evaluated algorithms.

5.1.3 Test Scenario 2 (Circular Track)

For all four cache sizes and all algorithms, we expected the cache hit-ratio to decrease over the test track because the user walks from the center of the spiral outwards.

Two results are shown in Figure 3(b) and 3(c); different cache sizes resulted in almost identical results for the LRU-based cache. The MHD-based cache performs identically to the LRU-based algorithm for a cache size of 250/500 elements, with only slight changes for 1000 and 2000 elements. The FAR algorithm has the lowest cache hit-ratio for all sizes. The results for 250 elements are similar to the other caches; the hit-ratio decreases with increasing cache size.

The algorithms performed as expected. The low hit-ratio of the FAR-cache is caused by the direction determination. The larger the radius of the circular walk, the more items are cached and need replacement while moving along the circle. After a threshold, recently visited items have to be replaced first, leading to less items being available when returning to the same angle of the circle.

5.1.4 Test Scenario 3 (Random Track)

We expect the location-aware caches (MHD and FAR) to perform better than the location-unaware cache (LRU). We also expected larger cache sizes to improve the cache hit-ratios. Selected results are shown in Figures 3(d) to 3(f). The MHD-based cache performs slightly better than the LRU- and FAR-based caches for 250 elements. When increasing the

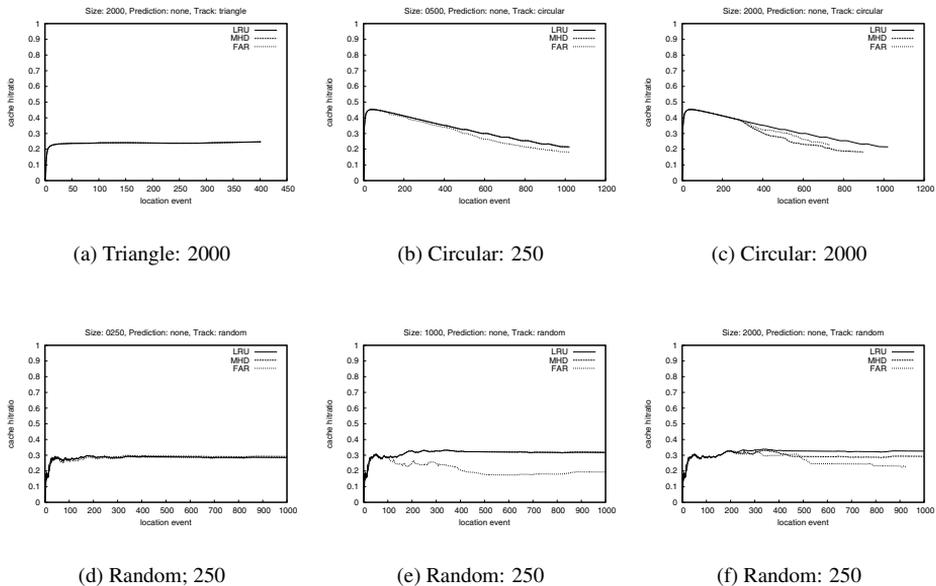


Figure 3: Cache hit ratio for Scenario S1, S2, S3 (track type: cache size)

cache size, the LRU cache hit-ratio increases (LRU-based cache has the best hit-ratio for 1000/20000). Figure 3(e) shows decreasing ratio for the FAR-based cache. For 2000 elements (Figure 3(f)), both location-aware caches perform worse than the LRU-based cache. Not all our expectations were met in this scenario. The bigger cache size only improved the LRU- and the MHD-based caches whereas it worsened the FAR-cache's results. The reason lies in the different replacement in FAR compared to MHD: FAR takes distance and movement direction into account. In addition, increasing the cache size seemed to improve only the LRU-based cache. Based on the similar graphs shapes, one can see that the basic character of the distance-based replacement (MHD) is similar to the time-based replacement (LRU). The location-aware algorithms performed best for the small sizes – for larger cache sizes, the LRU-based algorithm has more space to store its cached items.

5.1.5 Summary: Caching

We have shown that the system benefits from caching: 15 to 45 percent of the requests could be answered locally by re-using cached information. In these cases the mobile client did not need to re-download the information via the network, resulting in energy savings for network access. The effect of location-aware caching was not as significant as expected: For small cache sizes, the results of the location-aware (MHD and FAR) caches were largely comparable to those of the location-unaware cache (LRU). For larger cache sizes, the LRU-based cache replacement results in better hit-ratios than the other algorithms.

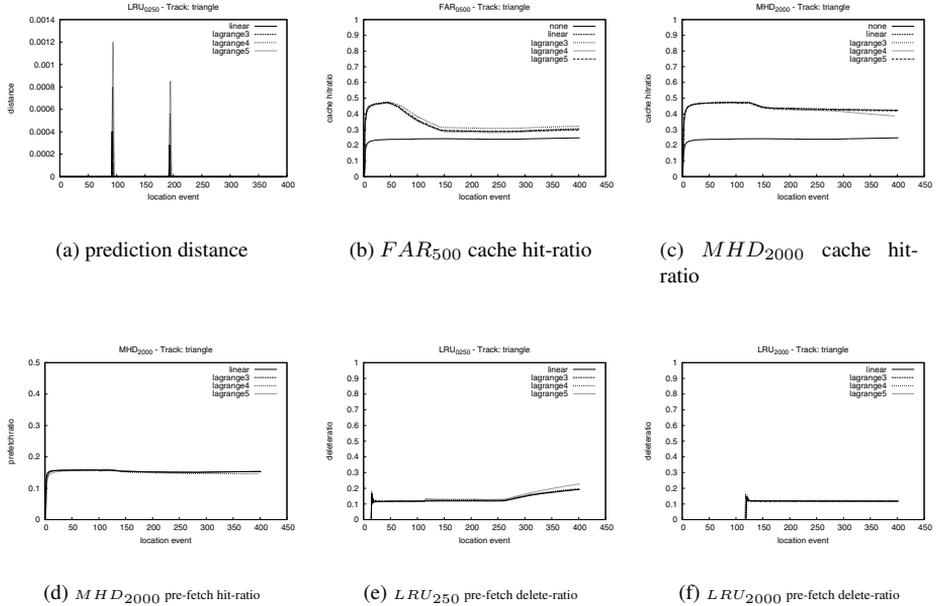


Figure 4: Selected pre-fetching results for Scenario S1 (triangle track).

5.2 Pre-fetching

In this subsection, we discuss the results of the pre-fetching algorithms, using the the same tracks and cache sizes as before.

5.2.1 Test Scenario 1 (Triangle Track)

The prediction distance for all implemented methods should be zero for linear movements. We expected all algorithms to perform similarly independent of the replacement strategy. We predicted slightly better cache hit-ratio for larger cache sizes. We show selected results that are indicative of the observed pre-fetching behavior.

1. Prediction distance: The prediction distance is the same for all replacement strategies and not affected by the cache size. Figure 4(a) shows the results for the LRU cache with 250 elements. Except for the mis-prediction at the two turning points, the prediction distance is always zero.
2. Cache hit-ratio: Introducing pre-fetching leads to a higher hit-ratio; all pre-fetching strategies lead to identical hit-ratio graphs for each selected cache replacement strategy and cache size (see Fig 4(b) and 4(c)). After a drop-off point, the hit-ratio decreases. At this point the cache is full and items have to be replaced. The decrease continues until a plateau is reached. Drop-off and plateau point are different for different caching strategies and cache sizes.

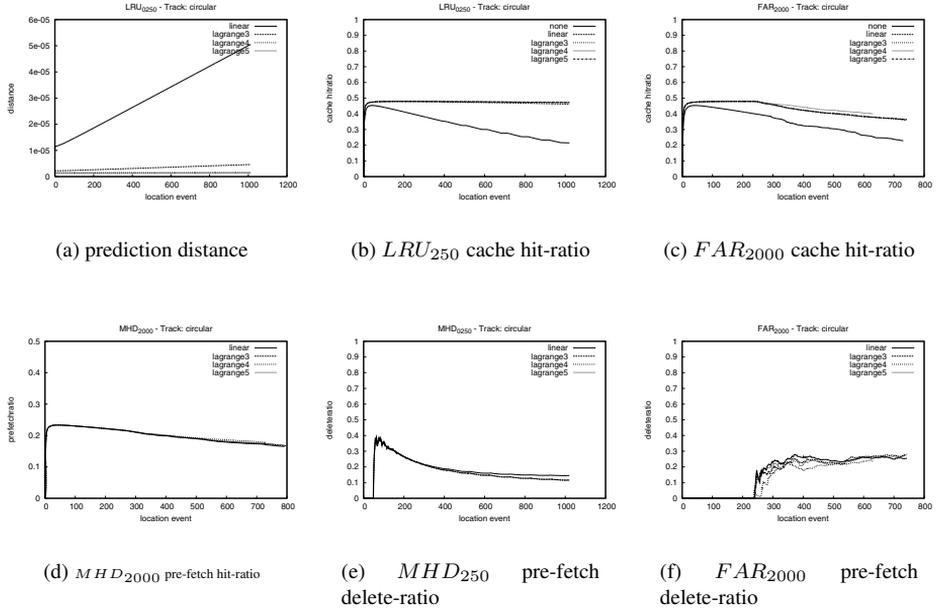


Figure 5: Selected pre-fetching results for Scenario S2 (circular track).

3. Pre-fetch hit-ratio: The different pre-fetching algorithms resulted in similar pre-fetch hit-ratios for the same cache size. An example is shown in Figure 4(d). A difference could only be measured for the FAR-based experiments of the size 500 and 1000 where we observed a slight decrease in the pre-fetch hit-ratio to 12 percent after the caches were filled.
4. Pre-fetch delete-ratio: Similar to the cache hit-ratio, pre-fetch delete-ratio is influenced by the cache replacement strategy and the cache size. Examples are shown in Figures 4(e) and 4(f). The delete ratio starts low and sharply increases at some point to stay relatively stable but higher afterwards. The results vary slightly for different location-prediction algorithms.

The evaluated pre-fetching strategies did not show any significant differences. The increase of the cache size did not increase the cache hit-ratio as expected but delayed the replacement of stored items. Therefore the decrease of the cache hit-ratio of the MHD and FAR caches was delayed. In addition, the increased size of the cache decreased the pre-fetch delete-ratio for the underlying caches. The pre-fetch delete ratio indicates that a stable percentage of items were pre-fetched but never used.

5.2.2 Test Scenario 2 (Circular Track)

We expected the prediction distance for the Lagrange-based algorithms to be better than the linear prediction because of better adaptation of the polynomial to the circular track.

The prediction distance is expected to improve especially if the prediction algorithm uses more historic locations (higher Lagrange index). An increase of the cache size should not increase the cache hit-ratio because the track never returns to the same location.

1. Prediction distance: The prediction distance is not affected by replacement strategies nor the cache size. Figure 5(a) shows the results for the LRU 250 cache. The Lagrange-based prediction methods perform better than the linear prediction. All prediction methods show only a very small difference from the predicted location to the actual one (at most $5^{-5} \text{degrees} = 5m$).
2. Cache hit-ratio: Different pre-fetching methods do not result in significantly different cache hit-ratios. A result for smaller caches is shown in Figure 5(b); the hit-ratio remains stable. For larger caches, the cache hit-ratio decreases. The actual shape of the graph is influenced by the replacement strategy; an example is shown in Figure 5(c)).
3. Pre-fetch hit-ratio: The different pre-fetching algorithms resulted in similar pre-fetch hit-ratios for the same cache size; an example is shown in Figure 4(d).
4. Pre-fetch delete-ratio: The pre-fetch delete-ratios for smaller caches are similar for all replacement strategies: the delete-ratio decreases after the cache is filled. Slight variations can be found for the FAR cache and for different prediction algorithms (see Figure 5(e)).

With larger cache sizes, the delete-ratio graphs change: it is initially zero and sharply rises at a point to about 22 percent and from then on it continues to increase slowly with variations. The exact shape depends on the pre-fetching algorithm. The pre-fetching algorithm shows a wider variation for the FAR cache, especially at the start of the deletions, where we could measure a difference of up to 10 percent.

The prediction distances were very small. The Lagrange-based algorithms performed better than the linear algorithm especially when using more locations. These comparatively small differences between the prediction algorithm are reflected in the minimal difference of the cache hit-ratios. Surprisingly, an increased cache size decreased the hit-ratio for the FAR and MHD caches and the used pre-fetching algorithms. The decreasing cache hit-ratio is caused by an increased pre-fetch delete-ratio, i.e., the victims are disadvantageously chosen. This effect is amplified while increasing the cache size. The LRU cache, however, was not affected by the increased cache size.

5.2.3 Test Scenario 3 (Random Track)

We expected this test to perform similarly to the one without pre-fetching: for the location-aware caches (MHD and FAR) to perform better than the time-aware cache (LRU). We also expected a better prediction accuracy and therefore a higher cache hit-ratio whenever a test setup uses a Lagrange-based prediction rather than the linear prediction. Generally, we expected a larger cache size to improve the cache hit-ratio measures.

1. Prediction distance: The measurements for the prediction distance are the same for all experiments and did not change for any used cache replacement strategy or cache size (see example in Figure 6(a)). The prediction distance of the linear prediction method is approximately 400 m. The prediction distance of the Lagrange methods using three, four and five points are 600 m, 1.1 km and 2 km, respectively.

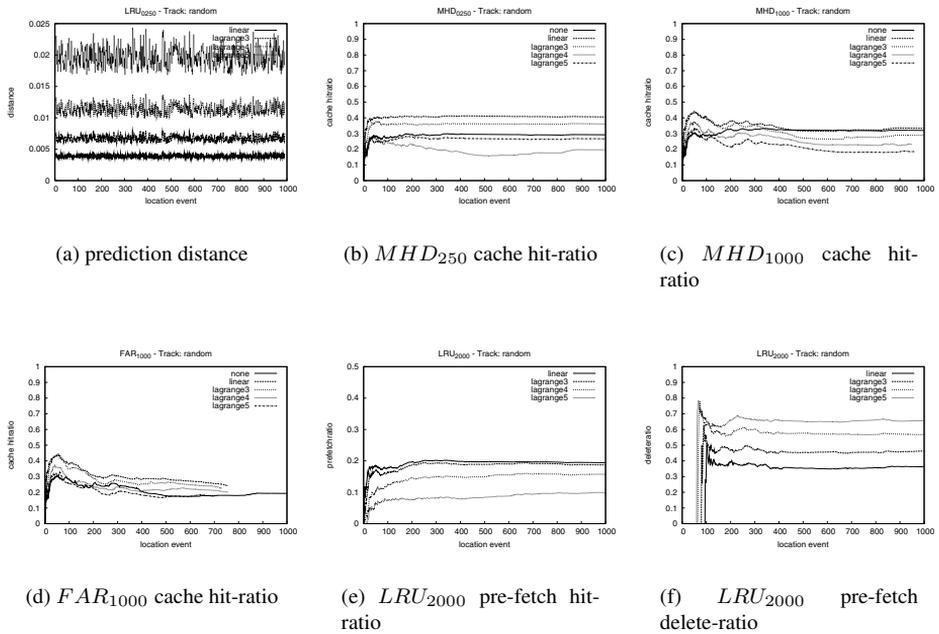


Figure 6: Selected pre-fetching results for Scenario S3 (random track).

2. Cache hit-ratio: For the smaller caches, all algorithms show the best hit-ratios for the linear and Lagrange3 pre-fetching (see Figure 6(b)). An increased cache size of 500 elements improved the cache hit-ratios of the LRU and MHD caches. For the FAR cache, all pre-fetching algorithms resulted in a lower cache hit-ratio than without pre-fetching. Results for Lagrange with 4 and 5 points improve with increased cache size. However, the MHD cache results are worse than the non pre-fetching cache hit-ratio (see Figure 6(c)). The FAR cache shows the best results for the linear pre-fetching (see Figure 6(d)). The largest experiment size of 2000 storable elements did not improve the results for the LRU cache. The results of the linear and lagrange3 pre-fetching using the MHD or FAR cache were only slightly better than the results of the non pre-fetching experiment.
3. Pre-fetch hit-ratio: The pre-fetch hit-ratio can be improved for all caches by increasing the size. The best pre-fetch hit-ratios can be achieved for the linear pre-fetching (see Figure 6(e)) followed by the Lagrange-based algorithms.
4. Pre-fetch delete-ratio: The pre-fetch delete-ratio using LRU or MHD can be reduced by increasing the cache size. For the FAR cache, this only applies for the linear and lagrange3 pre-fetching algorithms. Lagrange with 4 and 5 points using an FAR cache do not show clear results. An example for the general pre-fetch delete-ratio of the used pre-fetch algorithms is shown in Figure 6(f).

Prediction distance for the linear algorithm was the smallest for all experiments. Against our expectations, the Lagrange-based algorithms did not show the smallest distance for the prediction, which may have been caused by the noise in the Lagrange polynomials generated. Only the three point Lagrange-interpolation came close to the prediction distance of the linear algorithm. Although the cache hit-ratio and pre-fetch hit-ratio for the LRU_{2000} have similar results for the linear or lagrange3 pre-fetching (see Figure 6(e)), the pre-fetch delete-ratio for the lagrange3 algorithm is significantly higher (see Figure 6(f)). The cache size did not improve the cache hit-ratio, however, it decreased the pre-fetch delete-ratio for the LRU and MHD caches.

5.2.4 Summary: Pre-fetching

We have shown that the overall cache hit-ratio can be improved by pre-fetching data. We highlighted that this is not always the case, e.g., for the FAR_{500} experiment using the random track, the resulting cache hit-ratio was lower with pre-fetching than without. The linear prediction performed best for the random track; the Lagrange-based algorithms had the lowest distance on the circular track; all prediction algorithms performed the same on the triangle track. The pre-fetch hit-ratio did not show any significant differences for the different pre-fetch algorithm on the triangle and circular tracks. For the random track, the pre-fetching algorithms generated different results: The hit-ratio of the linear pre-fetch did perform best for all cases, with lagrange3 algorithm closest in quality. The pre-fetch delete-ratio for the triangle and circular scenarios did not show any significant difference for the used pre-fetching algorithms. Again, for the random track, the best results were achieved by using the linear prediction. The size of the underlying cache did not significantly influence the cache hit-ratio of the linear or circular tracks. For the random track the cache hit-ratio was slightly improved by an increased cache. In addition, the pre-fetch delete-ratio could be lowered by increasing the cache size.

6 Concluding Discussion

This paper reports about an extensive evaluation of an location-aware cache and pre-fetching service implemented for a mobile tourist information system. Selected results were discussed, more details are available in [7].

We showed that the cache performance not only depend on the replacement strategies used but also on the user movements. Using a pre-fetching service, we further improved the results of the cache hit-ratio. However, we note that pre-fetching information can decrease the performance when the wrong replacement items are chosen. We also determined that not all pre-fetched information was used and was therefore downloaded unnecessarily.

The overall results of the performance evaluation can be summarized as follows: Caching is generally beneficial for the mobile system resulting in less downloads and therefore less energy consumption. Pre-fetching of information can additionally improve the cache hit-ratio but false-positives may occur, i.e., unnecessary download of items that are never used. Surprisingly, the simple cache replacement algorithm (LRU) and the simple pre-fetching algorithm (linear) performed best or at least competitively in comparison to the other algorithms. As a next step we plan to pre-fetch information depending on the available network quality and test with tracks that reflect this feature. We also observed how the cache performance depends on the scenarios: For a future evaluation, one could adapt the replacement strategy during runtime.

What are the benefits of this evaluation? The analysis was done in a systematic way, not depending on particular walks in a given city but by evaluating the building blocks of a user's track: linear, circular, and random movements. As outlined before, we assumed here that no information about the area, or the (typical) movements of other users is available. We found that caching and pre-fetching strategies that are traditionally employed do not work well in these settings. Some of the issues were anticipated but the extend to which these traditional algorithms are unsuitable is much larger than hoped for.

We conclude that effective caching and pre-fetching in a mobile location-aware environment has to use additional information to improve the quality. Candidates are (1) maps or other information about a city layout, that is, information tracks available to the user at any given location, (2) information about typical user behaviour – usage statistics for the tracks known from the map, and (3) information about *this* user's behaviour - that is, usage statistics that reveal preferences defined, e.g., by work and home location. Techniques similar to recommender strategies may be used to create a specific location profile for a user that allows movement prediction in locations the user has not visited before.

Security and Privacy issues were not targeted so far. Also, currently the client cannot verify the validity of the received information, which is important for usage in P2P networks.

References

- [1] P. Cao, E. W. Felten, A. R. Karlin, and K. Li. A Study of Integrated Prefetching and Caching Strategies. In *Proc. ACM SIGMETRICS Performance Evaluation Review*, May 1995.
- [2] G. Cho. Using Predictive Prefetching to Improve Location Awareness of Mobile Information Service. In *Proceedings of the ICCS'02*, London, UK, 2002.
- [3] S. Dar, M. J. Franklin, B. T. Jónsson, D. Srivastava, and M. Tan. Semantic Data Caching and Replacement. In *Proc. of the 22th VLDB*, Sept. 1996.
- [4] A. Hinze and G. Buchanan. The Challenge of Creating Cooperating Mobile Services: Experiences and Lessons Learned. In *Twenty-Ninth Australasian Computer Science Conference (ACSC 2006)*, Hobart, Australia, Jan. 2006.
- [5] H. S. Jeon and S. H. Noh. Dynamic Buffer Cache Management Scheme Based on Simple and Aggressive Prefetching. In *Proceedings of the 4th Annual Showcase And Conference (LINUX-00)*, Berkeley, CA, Oct. 2000. The USENIX Association.
- [6] H. Kirchner, R. Krummenacher, T. Risse, and D. Edwards-May. A Location-aware Prefetching Mechanism. In *Proc. of 4th International Network Conference (INC 2004)*, University of Plymouth, Plymouth, 2004.
- [7] Y. Michel. Location-aware caching in mobile environments. Master's thesis, Freie Universität Berlin, Department of Computer Science, June 2006.
- [8] Q. Ren and M. H. Dunham. Using Semantic Caching to manage Location-dependent Data in Mobile Computing. In *Proceedings of MobiCom'00*, New York, USA, 2000.
- [9] B. Zheng and D. L. Lee. Semantic Caching in Location-Dependent Query Processing. In *Proc. of the 7th International Symposium on Advances in Spatial and Temporal Databases*, Redondo Beach, CA, USA, July 2001.
- [10] B. Zheng, J. Xu, and D. L. Lee. Cache Invalidation and Replacement Strategies for Location-Dependent Data in Mobile Environments. *IEEE Trans. Comput.*, 51(10):1141–1153, Oct. 2002.

Exploring Transactional Service Properties for Mobile Service Composition

Katharina Hahn, Heinz F. Schweppe
Institute for Computer Science, Freie Universität Berlin
Takustr. 9, 14195 Berlin
{khahn,schweppe}@mi.fu-berlin.de

Abstract: Service oriented computing provides suitable means to technically support distributed collaboration of heterogeneous devices such as in mobile environments. However, wireless communication links are unstable. When supporting collaboration in such environments, failures have to be optimistically coped with in order to still provide suitable correctness guarantees and avoid inconsistent system states. In this paper, we explore transactional properties of services in order to reasonably integrate transactional coordination with composition of services in mobile networks and provide suitable correctness guarantees.

1 Introduction

When deploying applications in mobile networks, one has to provide suitable means to cope with the characteristics of such environments. Due to the mobility of participants and the wireless networking technologies, mobile networks are *more dynamic* than fixed networks. This leads to *less stability* of communication links. On account of the inherent network dynamics, the *execution environment of an application is not known* at designtime and might differ from execution to execution. This also holds for the *heterogeneity* of portable devices which can be integrated in collaborative wireless computing.

Service oriented architectures are a powerful approach to support collaboration in such heterogeneous environments as they allow for loosely coupling of components thus respecting to their autonomy. Services can be dynamically discovered and composed into new value-added, so called *composed services*. Service discovery and binding at runtime provides suitable means to deal with composition of services in mobile environments in which the execution context, i.e., available services at runtime, is not previously known.

Especially due to the possibly physical distribution of several services and less stability of network links, it is indispensable to be able to cope with failures of different kinds to guarantee *correct execution*. This might come at the cost of relaxing correctness criteria such as the strict atomicity and isolation of all components which are guaranteed in databases. As strict guarantees rely on blocking of resources, services might be blocked for an arbitrary long time in case of disconnections of participants. In order to avoid blocking, “correctness” is relaxed and e.g. assured by avoiding any inconsistent and non-recoverable system states. This is achieved either through *forward-recovery* by still allowing the workflow

to successfully complete or *backward-recovery* by resetting the system to a previously consistent state.

In this paper, we develop a relaxed atomicity criterion for composed services which explores the non-functional (i.e. in this context transactional) properties of services. While avoiding blocking if possible to respect the autonomy of mobile participants, transactional coordination of services is reasonably integrated in workflow management. We adapt the composition of services at runtime in order to support correct execution, which we define by the notion of a relaxed atomicity criterion (see Section 5).

Consider for example a Tourist Ticketing System which supports users in finding points of interest and booking activities according to their previously defined preferences. An example of a mobile ticket vendor, whose business model is to sell tickets for touristic events, such as an exhibition, is shown in Figure 1. To make his offer more attracting to the user, he additionally offers to organize and book according transportation facilities to the exhibition venue and reserve a table at a nearby restaurant.

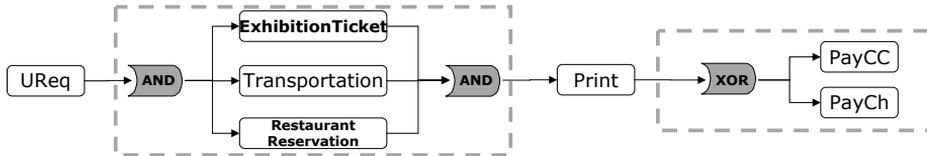


Figure 1: Combined ticketing for tourists.

After specifying the user’s request (UReq), including e.g. number of persons, the workflow is split to parallelly buy tickets for the *exhibition*, book *transportation* facilities and *reserve* a restaurant table. If all three components successfully finish, the tickets are *printed*, and the user is then asked to either pay by *credit card* or by *cash*. In this workflow, no failure handling has been specified yet. We argue, that it can be automated by exploring the workflow and participating services.

The rest of the paper is structured as follows: In Section 2 related work is presented. Section 3 introduces the formal model of transactional composition of services. In Section 4, transactional properties of components are explored, while Section 5 defines the atomicity guarantee. In Section 6, we outline alterations of the composition in order to support transactional execution. We conclude in Section 7.

2 Related Work

Many advanced transaction models (ATM), e.g. [GMS87, WS92] have been proposed which support transactional processing in distributed and heterogeneous databases [JK97]. These use less strict notions of atomicity and isolation in order to avoid blocking situations. Although they are very powerful, they are not capable of integrating structural requirements of complex applications in one transaction. A variety of mobile transaction models (such as [GGGG04, PA00, PA02]) have been proposed, which are able to cope with

failures due to frequent disconnections. However, they are not able to integrate different structural patterns as well.

Workflow execution and transactional coordination standards for Web-Services (WS) are two separate concerns. The execution of workflows is usually defined in BPEL (Business Process Execution Language), and controlled by workflow engines (e.g. [Apa]). Those provide support for the design, execution, visualization and analysis of workflows but do not integrate transactional guarantees. Transactional coordination has been specified by the WS-Transaction Framework (WS-Tx) [IBM05] which offers means for coordination of different services. It specifies different coordination types for short- and long-running activities and employs convenient backward recovery mechanisms to guarantee correct execution. As with most advanced transaction models proposed for asynchronous and decentralized transaction processing, WS-Tx lacks the flexibility to map different structural patterns to different transaction semantics.

Forward-recovery for composite services as e.g., proposed by [SDN07], is a promising approach to deal with the unstable availability of single participants in service oriented computing. In [SDN07], the authors propose the use of an abstract service provider. Its responsibility is to dynamically replace a failing service at runtime with a semantically equivalent service. Thus, specific failure situations are covered. But transactional execution of the whole workflow or subparts of it, is not considered.

Fauvet et. al [FDDB05] propose a high level operator for composing Web-Services according to transactional properties. Transactional execution relies on the tentative hold protocol (THP). Services are distinguished according to their additional capabilities: Support of 2PC, compensatability or neither. While this approach is interesting and powerful it uses a proprietary operator. We rather try to integrate transactional composition in existing standards by exploring non-functional properties.

In order to verify the execution of Web-Service workflows, several formalisms have been used, such as petri nets [HB03] or finite-state-machines [BFHS03]. These introduce powerful means to formally verify the execution of composite Web-Services but do not focus on transactional verification. Gaaloul et al. [GRGH07] use an event based-approach to model transactional composite services. As it provides suitable means to specify transactional behavior, we base our work on this formalism. It captures static verification however adaptation of the workflow in order to guarantee correct execution is left to the designer.

Binding services at runtime is integrated in existing workflow languages, e.g. in BPEL using *dynamic bindings using lookups* as partner links and can be performed by workflow engines. However, in mobile environments, discovery and matching of services at runtime is a challenging task. Many powerful approaches to service discovery in mobile networks exist, such as [RFH+01, CJFY06, RCJF02], as well as semantic description languages, e.g., OWL-S, WSDL-S or DSD [MDS02, WSD05, Kle04] to support matching of service offers and requests.

We also want to relate to the work done in the area of workflow scheduling, which identifies the problem of finding correct execution sequences for workflow activities, obeying inherent constraints, e.g. temporal or causality constraints [ASSR93, DKRR96]. Other approaches focus on minimizing communication costs or ensuring prearranged QoS obli-

gations defined in service level agreements [DD07]. As opposed to those, our work focuses on the analysis in order to reschedule activities to guarantee transactional correctness.

3 Formally Modeling Transactional Composition of Services

In the following, we introduce the model used to focus on the transactional behavior of services and composite services. The model is based on the event-algebra presented by [GRGH07]. It is also used to specify the relaxed transactional guarantees to be given for composite services, which we define in Section 5.

3.1 Service Model

The behavior of a service is modeled using a state-machine. Regardless its properties, a service has at least the following states: *Initial* indicates that it has not been activated yet, after activation its status is *active*. *Failed* and *cancelled* indicate failed execution (i.e., no changes are made persistent), either due to an internal error or externally triggered. If the service completed successfully, its state is *completed*. Additionally, if the service is compensatable, such as the example in Figure 2, it also has a *compensated* state.

The transitions between these states are either *internally* triggered (indicated by solid lines), i.e., by the service itself, or *externally* triggered (indicated by dashed lines), i.e., by another entity, such as the workflow engine, another service or a person.

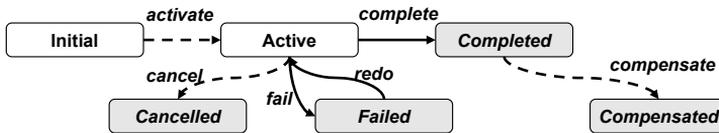


Figure 2: State machine of service.

3.2 Transactional Composition of Services

A composite service consists of a set of components and a set of axioms which define the correlation between the components. The normal execution defines the execution flow in case no failure occurs. *Transactional* composition of services additionally defines the relation of the components in case of failure in order to avoid inconsistent states, e.g. by compensation. So far, failure handling is left to the designer. By defining a workflow, the designer defines the regular execution; using compensation handler, he is able to specify behavior in case of *specific* failure situations. Thus, transactional guarantees for the whole workflow are not generally given.

The regular execution of services is modeled using *normal execution dependencies*. A

normal execution dependency between services S_i and S_j states that the completion of S_i triggers the activation of S_j :

Normal execution dependency: $depNrm(S_i, S_j) : S_i.complete() \Rightarrow S_j.activate()$

The standard failure handling mechanisms are cancellation of active services, compensation for completed services or activation of alternatives. According to these mechanisms and the possible failure situations, the following dependencies are defined:¹

- Alternative dependency: $depAlt(S_i, S_j) := S_i.fail() \Rightarrow S_j.activate()$
- Fail-Cancel dependency: $depFlCln(S_i, S_j) := S_i.fail() \Rightarrow S_j.cancel()$
- Fail-Compensate dependency: $depFlCps(S_i, S_j) := S_i.fail() \Rightarrow S_j.compensate()$
- Cancel-Cancel dependency: $depClnCln(S_i, S_j) := S_i.cancel() \Rightarrow S_j.cancel()$
- Cancel-Compensate dependency:
 $depClnCps(S_i, S_j) := S_i.cancel() \Rightarrow S_j.compensate()$
- Compensate-Compensate dependency:
 $depCpsCps(S_i, S_j) := S_i.compensate() \Rightarrow S_j.compensate()$

In order to guarantee transactional execution, we propose to automatically add these transactional dependencies and thus provide transactional guarantees. In the stated example (recall Figure 1), an *alternative* dependency has to be added between *PayCC* and *PayCh*, since the later is an alternative for the former one. Additionally *fail-cancel* dependencies are to be added between the *ticketing*, *transportation* and *reservation* services, as those should either all complete or none. Which dependencies are to be added, depends on the transactional properties of a service as well as its structural context. This is examined using workflow patterns.

3.3 Workflow Patterns

The structure of a composite service is represented by workflow patterns. Formally, a *workflow pattern* is a function that given a set of services returns the control flow [ABEW00]. Components of patterns are either services or contained patterns. By $WP(S)$, we denote a workflow pattern with a set of elements S . In the following, we exemplarily present three common workflow patterns and their characteristics for transactional workflow management.

SEQUENCE-Pattern This is a basic pattern which is used by the designer to specify, that one service is activated after the completion of the previous one. It is also known as *sequential* or *serial routing*. The invocation of the services is done in the same control thread. Arranging services S_i and S_j in sequence always infers a normal execution dependency (see Section 3.2) between S_i and S_j .

¹Note, that besides the alternative dependency, their denotation encodes cause and effect.

AND-Pattern Services, which are arranged in an AND-pattern are executed in parallel. Thus, the control flow is split up in parallel threads which are executed independently of each other. Therefore, we assume that not data dependencies between components of one AND-pattern exist. The control flow is synchronized at the join point and the subsequent workflow is activated as soon as *all* elements of the AND-pattern are completed. E.g., in our example *print* is invoked if *ticketing*, *transportation* and *reservation* are completed. The AND-pattern is also known as *parallel split* or *parallel routing*.

XOR-Pattern Based on any control data, one branch out of many is chosen. As these branches are never executed in parallel, the workflow is continued as soon as *one* branch completes. The XOR-pattern is also known as *exclusive choice*, *switch* or *conditional routing*. We focus on situations in which the choice relies on non-functional properties of services. Considering the formal model, alternative execution dependencies between elements of an XOR-pattern exist, as the failure of one can be recovered by the execution of another one. Thus, the XOR-pattern can be used to integrate forward-recovery, as the failure of one participant might be compensated by the execution of another one (e.g. *PayCC* and *PayCh*).

These patterns specify the control flow of the composite service and additionally the execution semantics. In an AND-pattern all included elements have to be completed while an XOR-pattern specifies, that one and only one element is to be completed. This already indicates the transactional dependencies to be added to avoid inconsistent states: E.g., the semantics of the AND-pattern infer, that in case that one element fails, all of the others have to be canceled. Thus *fail-cancel* dependencies between all elements are added. By contrast, if an element within an XOR-pattern fails, an alternative can be executed, thus *alternative* dependencies are added between elements of the XOR-pattern. However, the transactional dependencies are also influenced by the transactional properties of services, which are studied in the next section.

4 Exploring Non-Functional Service Properties

Within this section, we introduce non-functional properties of services which are relevant to define appropriate failure handling mechanisms. Additionally, we derive properties for workflow patterns according to the contained elements.

4.1 Transactional Service Properties

In order to ensure correct execution of composite services, we examine the transactional behavior of services: I.e., whether services can fail, whether completed services can be compensated for and whether they need to be recovered in case of failure. The first two properties (redoability and compensatability) have already been considered in the context

of flexible transactions [ZNBB94]. We additionally classify services according to their need for compensation in case of failure.

Compensatability

The compensatability of a service indicates whether its effects can be undone after completion. Thus a service S is compensatable (denoted as $S.comp = 1$ and $S.comp = 0$ accordingly for non-compensatable services) if there exists a service C which semantically undoes the effects of S .² This is formally specified by the *compensate*-transition and a *compensated*-state as shown in Figure 2.

Redoability

The redoability of services specifies whether the service's execution can fail, i.e. a redoable service S (denoted as $S.redo = 1$) will definitely complete after being activated. This is important for example for compensating services, as it is assumed that their execution will not fail. Redoability of a service is modeled by an internal *redo*-transition (see Figure 2). Thus, once invoked the completion of the service can be guaranteed.

Consistent Closure

By including a service in a workflow, the designer states whether its completion is inevitable for the success of the workflow (e.g., if no alternatives exist). So far, it is vice versa assumed, that a service is only allowed to be completed if the workflow is completed. However, some services may allow for *inconsistent closure*, i.e. they complete although the workflow may be canceled. This is usually given by semantics of the service or the consistency demands of the data it operates on. Examples are a log-in service which authenticates the user to the provider or a printing service which prints a confirmation. We therefore additionally consider the following transactional property of services:

A service S demanding consistent closure³ (denoted as $S.consComp = 1$) needs recovery in case the workflow is rolled back. Thus, its completion infers the completion of the workflow. A service, that is allowed to complete *inconsistently* ($S.consComp = 0$) does not need recovery, in case the workflow execution fails. Thus it states, whether the effects of a service *have to be* undone in case of recovery. Within our model, $\{fail, cancel, compensate\}$ -compensate dependencies (*-compensate dependencies for short) to a service demanding consistent completion have to be added.

Derived Property: Recoverability

For validation and defining the appropriate backward recovery mechanisms, i.e. adding the transactional dependencies, it is of interest to know, whether a service (or a pattern) can be backward-recovered in case of failure. We denote this by *recoverability* of an element. A service S is *recoverable* ($S.recover = 1$), if and only if it is compensatable or does not need consistent completion. This property is especially interesting for patterns: A pattern is regarded to be *recoverable* (denoted as $WP(S).recover = 1$), if all of the executed services which demand consistent completion are compensatable and all contained patterns are recoverable.

²Non-compensatable services are sometimes also referred to as *pivot* services.

³We use the term *consistent completion* interchangeably.

4.2 Inferring Pattern Properties

As stated before, we want to analyze the workflow with the bound services at runtime. We thus analyze the composite service bottom-up and derive the transactional properties for patterns according to the contained elements, as indicated by the dashed lines in Figure 1. The transactional properties of a pattern $WP(S)$ are denoted just as the transactional service properties: $WP(S).comp$, $WP(S).consCompl$, $WP(S).redo$ and $WP(S).recover$. They are determined according the properties of included elements as well as the pattern (i.e., its execution semantics). As the SEQUENCE and the AND-pattern both state that *all* included elements have to be completed, their transactional properties are determined in the same way.

Transactional Properties of the SEQUENCE and AND-Pattern

A pattern $WP(S)$ which is a SEQUENCE-pattern or an AND-pattern

- is compensatable, if and only if all included services are compensatable:
 $WP(S).comp = 1 \Leftrightarrow \forall S_i \in S : S_i.comp = 1$
- needs consistent completion, as soon as one service needs consistent completion:
 $WP(S).consCompl = 1 \Leftrightarrow \exists S_i \in S : S_i.consCompl = 1$
- is redoable, if all included services are redoable:
 $WP(S).redo = 1 \Leftrightarrow \forall S_i \in S : S_i.redo = 1$
- is recoverable, if all included services are either compensatable or allow for inconsistent completion:
 $WP(S_i).recover = 1 \Leftrightarrow \forall S_i \in S : S_i.comp = 1 \vee S_i.consCompl = 0$

Transactional Properties of the XOR-pattern

Due to the different execution semantics (one and only one included element is to be executed), the transactional pattern properties for an XOR-pattern are determined differently than for the SEQUENCE- and the AND-pattern. As it cannot be previously to execution stated which service will complete, the pattern properties can only be previously determined for the following (not all) cases. Otherwise they cannot be determined until execution. The XOR-pattern

- is compensatable, if all included services are compensatable. It is non-compensatable, if all included services are non-compensatable. Otherwise, it is not known previous to execution.
 $WP_{XOR}(S).comp = 1 \Leftarrow \forall S_i \in S : S_i.comp = 1$
 $WP_{XOR}(S).comp = 0 \Leftarrow \forall S_i \in S : S_i.comp = 0$
- needs consistent completion, if all included services demand consistent completion. If none of the included services demands consistent closure, the pattern allows for inconsistent closure. Otherwise, it is not known until execution.
 $WP_{XOR}(S).consCompl = 1 \Leftarrow \forall S_i \in S : S_i.consCompl = 1$
 $WP_{XOR}(S).consCompl = 0 \Leftarrow \forall S_i \in S : S_i.consCompl = 0$

- is redoable, if at least one service is redoable. This service can be invoked in case of failure of any other included service. Else, the pattern is non-redoable.

$$WP_{XOR}(S).redo = 1 \Leftrightarrow \exists S_i \in S : S_i.redo = 1$$

- is recoverable, if all included services are either compensatable or allow for inconsistent completion.

$$WP_{XOR}(S).recover = 1 \Leftarrow \forall S_i \in S : S_i.Comp = 1 \vee S_i.consCompl = 0$$

$$WP_{XOR}(S).recover = 0 \Leftarrow \forall S_i \in S : S_i.comp = 0 \wedge S_i.consCompl = 1$$

An arbitrary combination of the properties *compensatability*, *consistent closure* and *redoability* is possible. The *recoverability* is derived by the former properties for services and by the properties of the included components for patterns. Note, that for verification purposes, the compensatability of a service is disregarded in case of inconsistent completion. However, it is important when adding transactional dependencies to a composite service. We denote the transactional properties of a service or a pattern S as P_S defined as follows:

$$P_S = (S.comp, S.consCompl, S.redo, S.recover)$$

Accordingly, a service S with $P_S = (0, 1, 1, 0)$ is a service which is not compensatable, demands consistent completion, is redoable and thus not recoverable. Based on the this model, we will now introduce the transactional guarantees that we support when dealing with transactional workflows.

5 Relaxing Atomicity for Transactional Composite Services

Blocking of resources is counterproductive in the environment of loosely coupled and especially mobile services. Due to the autonomy of mobile services and the fluctuant availability of those, relaxed atomicity guarantees have to be defined. These specify the criteria for *correct execution*. In order to avoid blocking situations, different notions of relaxed atomicity e.g., *semantic atomicity* [GM83] and *semi-atomicity* [ZNBB94], which allow the commitment of subtransactions at different times have been proposed for database transactions. Convenient backward-recovery mechanisms ensure that already committed subtransactions are recovered in case of failure. In the model of flexible transactions [ZNBB94], semi-atomicity is validated by reviewing the order of subtransactions according to their non-functional properties: The commitment of compensatable subtransactions *precedes* the commitment of pivot subtransactions. As their commitment infers the commitment of the whole transaction, it is only *followed* by redoable subtransactions.

We adapt the model of semi-atomicity defined for flexible transactions and extend it to comprise transactional workflow management in mobile environments. By defining accepted termination states (ATS), the designer defines representational sets of services whose completion reflect the successful execution of the composite service. Multiple sets exists, if alternatives in the workflow exist (or even multiple ATS exist).

We define *semi-atomic* execution of the workflow with respect to the transactional properties of components (as presented in Section 4) as follows:

Semi-Atomicity of Composite Services: Semi-atomic execution of a composite services with defined ATS is ensured if

- either all services belonging to one valid execution path to an ATS are completed and all other services which demand consistent completion are not completed
- or no service demanding consistent completion is completed.

As recovery for services which do not demand consistent completion is disregarded, this relaxes the semi-atomicity as defined for flexible transactions.

6 Ensuring Semi-Atomicity for Composite Services

In this section, we present how semi-atomicity of a composite service as defined in Section 5 is supported. Due the increased autonomy of mobile devices as opposed to fixed wired networks, we try to avoid the use of blocking protocols such as the 2PC. By exploring the transactional service properties in the context of workflow patterns at runtime, the structure of the composite service is altered and choice on alternatives is influenced to support semi-atomic execution.

6.1 Alter Execution Type and Order

At first, we consider the execution order of components. It is initially given by the workflow pattern, they are arranged in (either parallel or sequential). In case of failure, it has to be ensured, that all already completed services are recoverable or forward-recovery for the failing services exists. This is identified by considering the transactional properties of components. Recall, that the transactional properties of a component S are denoted as:

$$P_S = (S.comp, S.consCompl, S.redo, S.recover)$$

Assuming no data dependency between two services S_i and S_j , then their execution is aligned (i.e., S_i precedes S_j) in the following cases.

1. $P_{S_i} = (*, *, 0, 1)$ and $P_{S_j} = (0, 1, *, 0)$, or
2. $P_{S_i} = (0, 1, 0, 0)$ and $P_{S_j} = (0, 1, 1, 0)$.

Otherwise, in case S_j completes but S_i fails, the semi-atomicity of the workflow is harmed as in all stated cases, S_j cannot be recovered. If at least two services S_i and S_j both are not recoverable and not redoable, i.e. $P_{S_i} = P_{S_j} = (0, 1, 0, 0)$, then semi-atomic execution is only assured by *coordinating* them within a subtransaction, e.g. using WS-Tx. Otherwise, in case of failure of one of them and the completion of the other, the composite service is in an inconsistent state which cannot be recovered.

Any other combination of elements can be executed *independently* of each other without harming the semi-atomicity of the execution.

Recall the example of the combined ticketing from Figure 1. The designer indicated to execute the ticketing T_i , transportation T_a and reservation R services in parallel. The T_i be a local service with $P_{T_i} = (1, 1, 0, 1)$. Providers for T_a and R are discovered at runtime. Assume that services with the properties $P_{T_a} = (0, 1, 1, 0)$ and $P_R = (0, 0, 1, 1)$ are discovered. The workflow is then altered as shown in Figure 3a. T_i precedes the execution of T_a , R is executed parallelly.

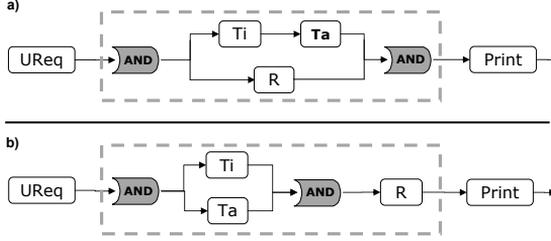


Figure 3: Altering execution order of components.

Assume that executed again different services with the following properties $P_{T_a} = (1, 1, 0, 1)$ (i.e., compensatable) and $P_R = (0, 1, 0, 0)$ (i.e., non-compensatable and non-redoable), then the composition is altered as shown in Figure 3b. The execution of T_a and T_i is parallelized, while R is executed after them. Otherwise, if the workflow was executed as originally intended, then failure of either T_a or T_i with concurrent completion of R would lead to a non-recoverable system state. In both cases the non-functional properties of the pattern WP included in the dashed lines are $P_{WP} = (0, 1, 0, 0)$.

6.2 Service Selection According to Transactional Properties

According to their context, transactional properties of services are used to determine a preference relation on which service to include in an XOR-pattern. Consider for example Figure 4: At runtime, either branch S_i or S_j is to be taken within the XOR-pattern, before the subsequent workflow S_{subseq} is invoked. Let the transactional properties be $P_{S_i} = (1, 1, 0, 1)$ and $P_{S_j} = (0, 1, 1, 0)$. According to Section 4.2, the XOR-pattern is thus redoable, as S_j is redoable. If $P_{S_{prev}} = (1, 1, *, 1)$ and $S_{subseq}.redo = 0$ then, S_i must be chosen in order to guarantee semi-atomicity. Otherwise, in case of failure of S_{subseq} , the XOR-pattern cannot be recovered. If in contrast $P_{S_{prev}} = (0, 1, *, 0)$ and $S_{subseq}.redo = 1$, then the XOR-pattern has to complete to ensure semi-atomicity. This is given through the redoability of pattern which is already assured through the presence of S_j (as $S_j.redo = 1$). Thus, in this case, the choice between those two services can be done according to other non-functional properties.

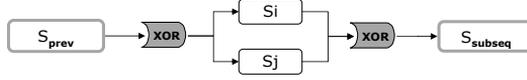


Figure 4: Choosing S_i or S_j according to transactional properties.

6.3 Automation of Recovery: Adding Transactional Dependencies

After dynamically adjusting the workflow, transactional dependencies have to be added at runtime in order to define the appropriate forward- and backward-recovery mechanisms. We will add dependencies according to the workflow patterns and alter those according to the non-functional properties of services.

1. All components S_i and S_j for which a normal execution dependency in the form $depNrm(S_i, S_j)$ exists, the appropriate backward-recovery in case of failure, cancellation or compensation are added. Thus dependencies of the form $depFlCps(S_j, S_i)$, $depClnCps(S_j, S_i)$ and $depCpsCps(S_j, S_i)$ are added ($dep * Cps(S_j, S_i)$ for short).
2. All components which do not need consistent completion, are omitted in backward recovery. Thus if S_j does not need consistent closure and any transactional dependencies in the form $depXCps(S_k, S_j)$ and $depXCps(S_j, S_i)$ exist, for $X \in \{Fl, Cln, Cps\}$, then those are consolidated to an appropriate dependency of the form $depXCps(S_k, S_i)$. Thereby, S_j is skipped in case of backward recovery.
3. All compensatable components in an AND-pattern $WP_{AND}(S)$ are to be compensated in case of failure. Thus, the following dependencies are added to the compensatable components $S_c \in S$ from the subsequent workflow S_{sub} : $dep * Cps(S_{sub}, S_c)$. Additionally, all parallely arranged components are to be cancelled, if failure or cancellation of one of them occurs. Thus, for all $S_j \in S$, $depFlCln(S_j, S)$ and $depClnCln(S_j, S)$ are added.
4. Components aligned in an XOR-pattern $WP_{XOR}(S)$ are considered to be alternatives. After analyzing their preference relation, alternative dependencies of the form $depAlt(S_i, S_j)$, regarding the ranking order are added (i.e., S_j is an alternative for S_i). If S_j is redoable, the alternative dependency in the form of $depAlt(S_j, S_k)$ is deleted, as failure of S_j will not occur.
5. Finally, for any S_j in the composite service, which is not compensatable, any dependency of the form $dep * Cps(S_k, S_j)$ is deleted (as it is not compensatable). On the other hand, if S_j is redoable, then it cannot fail, thus any failure dependency in the form $depFl * (S_j, S_k)$ is deleted.

The approaches introduced in Section 6.1 to 6.3 enable adaptations of the composite service at runtime according to transactional properties of all components. These adaptations and the appropriate recovery mechanisms support semi-atomic execution of the workflow.

7 Conclusion

In this paper, we have introduced adapted notion of semi-atomicity which explores transactional properties of services to define correct execution for composition of mobile services. This is especially interesting for mobile networks, as those are more dynamic and error-prone. We considered the properties whether services can fail, can be compensated and need to be compensated in the context of workflow patterns. As the execution context is not previously known in mobile environments, we outlined dynamical adaptations of the composition at runtime in order to support semi-atomic execution in the current execution context. Appropriate backward- and forward recovery mechanisms are integrated by transactional dependencies.

By adapting the workflow, i.e. changing order and type of invocation, and influencing the preference relation for alternatives, it is possible to support semi-atomic execution. Ordering services avoids coordination through blocking and still provide transactional execution guarantees. As opposed to existing approaches, we explore *existing* standards for service composition to *automate* transactional execution of composite services. As part of future work, we want to design an adaptive, mobile workflow engine which exploits transactional service properties to automate transactional execution at runtime.

References

- [ABEW00] W. v. d. Aalst, P. Barthelmess, C.A. Ellis, and J. Wainer. Workflow Modeling Using Proclerts. In *Proceedings of the 7th International Conference on Cooperative Information Systems (COOPIS'2000)*, pages 198–209, 2000.
- [Apa] Apache Orchestration Director Engine. <http://ode.apache.org/>.
- [ASSR93] P. C. Attie, M. P. Singh, A. Sheth, and M. Rusinkiewicz. Specifying and Enforcing Intertask Dependencies. In *Proceedings of the 19th VLDB Conference*, 1993.
- [BFHS03] Tevfik Bultan, Xiang Fu, Richard Hull, and Jianwen Su. Conversation Specification: A New Approach to Design and Analysis of E-service Composition. In *WWW '03: Proceedings of the 12th International Conference on World Wide Web*, pages 403–410, New York, NY, USA, 2003. ACM.
- [CJFY06] Dipanjan Chakraborty, Anupam Joshi, Tim Finin, and Yelena Yesha. Toward Distributed Service Discovery in Pervasive Computing Environments. *IEEE Transactions on Mobile Computing*, February 2006.
- [DD07] Dmytro Dyachuk and Ralph Deters. Service Level Agreement Aware Workflow Scheduling. In *Proceedings of International Conference on Services Computing (SCC)*, volume 0, pages 715–716. IEEE Computer Society, 2007.
- [DKRR96] H. Davulcu, M. Kifer, C. R. Ramakrishnan, and I. V. Ramakrishnan. Logic Based Modeling and Analysis of Workflows. In *Proceedings of the 17th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pages 1–3. ACM Press, 1996.
- [FDDB05] M.-C. Fauvet, H. Duarte, M. Dumas, and B. Benatallah. Handling Transactional Properties in Web Service Composition. In *WISE*, pages 273–289, 2005.

- [GGGG04] Ankur Gupta, Nitin Gupta, R. K. Ghosh, and M. M. Gore. Team Transaction: A New Transaction Model for Mobile Ad Hoc Networks. In *ICDCIT*, 2004.
- [GM83] Hector Garcia-Molina. Using semantic knowledge for transaction processing in a distributed database. *ACM Trans. Database Syst.*, 8(2):186–213, 1983.
- [GMS87] H. Garcia-Molina and K. Salem. Sagas. *SIGMOD Rec.*, 16(3):249–259, 1987.
- [GRGH07] Walid Gaaloul, Mohsen Rouached, Claude Godart, and Manfred Hauswirth. Verifying Composite Service Transactional Behavior Using Event Calculus. In *OTM Conferences (1)*, pages 353–370, 2007.
- [HB03] Hamadi and Benatallah. A Petri Net-based Model for Web Service Composition. In *Proceedings of the 14th Australasian Database Conference (ADC'03)*, 2003.
- [IBM05] WebServices AtomicTransaction, 2005. <http://www.ibm.com/developerworks/library/specification/ws-tx/>.
- [JK97] Sushil Jajodia and Larry Kerschberg, editors. *Advanced Transaction Models and Architectures*. Kluwer, 1997.
- [Kle04] Michael Klein. Handbuch zur DIANE Service Description. Technical Report 2004-17, Universität Karlsruhe, Faculty of Informatics, December 2004.
- [MDS02] F.V. Harmelen J.Hendler I.Horrocks D.L. McGuinness P.F. Patel-Schneider M. Dean, D. Connolly and L.A. Stein. Web Ontology Language (OWL) Reference Version 1.0, 2002. <http://www.w3.org/TR/2002/WD-owl-ref-20021112>.
- [PA00] G. Pardon and G. Alonso. CheeTah: a Lightweight Transaction Server for Plug-and-Play Internet Data Management. In *VLDB '00: Proceedings of the 26th International Conference on Very Large Data Bases*, pages 210–219, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.
- [PA02] A. Popovici and G. Alonso. Ad-Hoc Transactions for Mobile Services. In *Proceedings of the 3rd VLDB Workshop on Transactions and Electronic Services (TES '02)*, 2002.
- [RCJF02] O. Ratsimor, D. Chakraborty, A. Joshi, and T. Finin. Allia: Alliance-based Service Discovery for ad-hoc Environments. In *Proceedings of the 2nd international workshop on Mobile commerce*, pages 1–9, New York, NY, USA, 2002. ACM.
- [RFH+01] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker. A scalable content addressable network. In *Proceedings of ACM SIGCOMM*, 2001.
- [SDN07] Michael Schäfer, Peter Dolog, and Wolfgang Nejdl. Engineering Compensations in Web Service Environment. In *Proceedings of 7th Intl. Conference on Web Engineering (ICWE)*, pages 32–46, Como, Italy, 2007.
- [WS92] Gerhard Weikum and Hans-Jorg Schek. Concepts and Applications of Multilevel Transactions and Open Nested Transactions. In *Database Transaction Models for Advanced Applications*, pages 515–553. 1992.
- [WSD05] Web Service Semantics WSDL-S, 2005. <http://www.w3.org/Submission/WSDL-S/>.
- [ZNBB94] A. Zhang, M. Nodine, B. Bhargava, and O. Bukhres. Ensuring Relaxed Atomicity for Flexible Transactions in Multidatabase Systems. *SIGMOD Rec.*, 23(2), 1994.

Ein Überblick über Ansätze zur Vermeidung der Manipulation von Ortungsverfahren

Michael Decker

Institut AIFB, Universität Karlsruhe (TH)

76128 Karlsruhe

decker@aifb.uni-karlsruhe.de

Abstract: Die Verfügbarkeit von Ortungsverfahren zur Positionsbestimmung von mobilen Endgeräten wie Mobilfunktelefonen oder PDAs ermöglicht eine eigene Klasse von neuartigen Anwendungen, die sog. Location Based Services (LBS). In den einschlägigen Arbeiten wird jedoch fast nie betrachtet, inwiefern die Ortungsverfahren robust gegenüber Manipulationsversuchen („Location-Spoofing“) durch Dritte oder die (möglicherweise illegitimen) Besitzer des Endgerätes selbst sind. Es werden deshalb zur Motivation zunächst einige Anwendungsszenarien genannt, für die spezielle manipulationsresistente Ortungsverfahren benötigt werden. Darauf aufbauend werden dann verschiedene grundlegende Ansätze zur Vermeidung von Location-Spoofing klassifiziert und beschrieben. Es wird auch darauf eingegangen, welche besonderen Anforderungen die einzelnen Verfahren an die Endgeräte stellen und für welche Anwendungsszenarien sie geeignet sind.

1 Einleitung und Motivation

Unter Ortung versteht man die Ermittlung des Aufenthaltsortes eines mobilen Gerätes. Meistens handelt es sich bei diesem Gerät um ein von einem Nutzer ständig mitgeführtes persönliches Kommunikationsgerät wie z.B. Mobilfunktelefon, Notebooks oder PDA mit der Möglichkeit zur drahtlosen Datenkommunikation über Technologien wie GPRS, UMTS oder WiFi. Mittlerweile stehen zahlreiche technische Verfahren zur Realisierung von Ortung zur Verfügung (z.B. [TuPo04, 74ff] oder [Küpp05, 155ff]):

- Das wohl bekannteste Ortungssystem ist das satellitengestützte „Global Positioning System“ (NAVSTAR GPS) der USA, mit dem eine Genauigkeit bis auf ca. 15 m Meter erzielt werden kann. Mit zusätzlichen Korrektursignalen von terrestrischen Referenzstationen (sog. Differential GPS, D-GPS) kann eine Genauigkeit bis auf wenige cm erreicht werden.
- In zellular aufgebauten Funknetzen (z.B. Mobilfunknetze) kann anhand der Kenntnis der verwendeten Basisstation eine Positionsbestimmung vorgenommen werden (Zellortung). Je nach Größe der von einer Basisstation bedienten Funkzelle beträgt die Ortungsgenauigkeit zwischen 100 Metern (in Ballungszentren) oder 35 Kilometern (im ländlichen Raum). Durch Hinzunahme weiterer Basisstationen (Triangulation) oder Berücksichtigung der Laufzeiten kann die so erzielte Genauigkeit weiter verbessert werden.

- Die genannten Verfahren sind für Anwendungen in Gebäuden leider wenig geeignet, so dass auch spezielle Indoor-Verfahren entwickelt wurden, z.B. WLAN-Fingerprinting oder RFID-basierte Systeme. Mit diesen Systemen sind Genauigkeiten von 1-2 m erreichbar.

Mobile Anwendungen, die den über ein oder mehrere Ortungsverfahren ermittelten Aufenthaltsort eines mobilen Endgerätes auswerten, werden als „Location-based Services“ (LBS) bezeichnet [TuPo04, 73ff]. Hierbei muss das geortete Endgerät nicht notwendigerweise das Gerät des Dienstaufrufers sein: bei sog. Tracking-Diensten (z.B. <http://www.track-your-kid.de>) initiiert der Nutzer mit seinem Endgerät die Ortung eines anderen Endgerätes. Ein weiteres klassisches Beispiel für LBS sind sog. POI-Finder-Dienste: ein POI (Point-of-Interest) ist hierbei eine bestimmte stationäre Einrichtung, die typischerweise für einen Reisenden oder Ortsunkundigen von Interesse ist, z.B. eine Sehenswürdigkeit, ein Restaurant, Hotel oder Tankstelle. Mit einem POI-Finder-Dienst kann sich der Nutzer ohne mühsame Eingabe seines aktuellen Aufenthaltsortes — den er vielleicht auch gar nicht genau genug kennt — den nächstgelegenen POI einer bestimmten Kategorie (z.B. Apotheke) anzeigen lassen, ggf. ist auch eine Navigation an diesen Ort hin möglich.

Bei der Entwicklung von Ortungsverfahren steht meist die räumliche Präzision der erzielten Ortung im Vordergrund; die Absicherung eines Ortungsverfahrens gegen bewusste Manipulationsversuche durch den mobilen Nutzer selbst (interner Angriff) oder einen Dritten (externer Angriff) wird meist nicht behandelt, wohl weil sich aus den üblicherweise herangezogenen Anwendungsszenarien keine Motivation für einen Beteiligten oder potenzielle Angreifer ergibt, erheblichen technischen Aufwand zu betreiben, um die Ortung zu manipulieren. Es lassen sich aber durchaus Anwendungen finden, bei denen dies nicht der Fall ist, so dass bei der Entwicklung die Absicherung gegen interne und/oder externe Manipulationen von Bedeutung ist:

- Unter Zugriffskontrolle (Access Control) werden Maßnahmen und Techniken verstanden, um zu entscheiden, ob einem bestimmten Nutzer eines Informationssystems gestattet werden soll, eine bestimmte Operation (z.B. lesen, schreiben, löschen, ausführen) auf einer bestimmten Ressource (z.B. Datei, Datenbankobjekt, Dienst) auszuführen. Die hierfür zuständige Komponente (Soft- und/oder Hardware) wird als Referenz-Monitor bezeichnet. Bei der sog. ortsabhängigen Zugriffs wird für diese Entscheidung nicht nur die Identität oder Gruppen-/Rollenzugehörigkeit des jeweiligen Nutzers ausgewertet, sondern auch dessen aktueller Aufenthaltsort (z.B. [DaBP07]). So lassen sich Sicherheitspolitiken umsetzen, die etwa den Zugriff auf vertrauliche Dateien außerhalb des Betriebsgeländes oder aus Ländern mit Spionagegefahr verbieten. Dieser Anwendungsfall stellt übrigens die ursprüngliche Motivation für den vorliegenden Artikel dar. Bemerkenswerterweise wird in den einschlägigen Veröffentlichungen mit ortsbewussten Zugriffskontrollmodellen die Manipulationsresistenz von Ortungsverfahren nicht thematisiert.
- Anstelle die Ortung als zusätzliche Information neben der Identität eines Nutzers für eine Zugriffsentscheidung auszuwerten, kann es auch sinnvoll sein, *nur* die Ortsinformation auszuwerten, wenn nämlich der Aufenthalt an einem bestimmten Ort

schon als Berechtigungsnachweis gewertet werden kann. Dies wäre etwa der Fall für eine Unternehmensabteilung oder ein Gelände, deren/dessen Zugang mit herkömmlichen Mitteln (etwa Wachpersonal, Zäune, verschließbare Türen, Mauern) gesichert ist. Eine rein ortsbasierte Zugriffskontrolle hat den Vorteil, dass nicht automatisch mit dem Zugriff auf gesicherte Ressourcen ein Nutzungsprofil der jeweiligen Nutzer erstellt wird (z.B. Nutzer X erscheint immer erst um 10 Uhr am Arbeitsplatz).

- Bestimmte mobile Dienste sollen nur dann verfügbar sein, wenn sich der Nutzer in unmittelbarer Nähe zu einem bestimmten physischen Objekt befindet, z.B. bei Diensten zur Fernsteuerung von (beweglichen) Maschinen/Fahrzeugen, Entriegelung von Türen oder Zugriff auf Drucker.
- Werden Navigationsdienste für militärische Fahrzeuge oder Werttransporte genutzt, so könnte eine externe Manipulation der Ortung genutzt werden, das Fahrzeug in einen Hinterhalt zu locken¹.
- Ortungstechnologien werden auch zur Nachverfolgung des Aufenthaltsortes von verurteilten Straftätern („elektronische Fußfessel“²) und Werttransporten verwendet, so dass hier interne Manipulationen zu befürchten sind. Weiter gibt es die Anwendung, dass ein Alarm ausgelöst werden soll, wenn hochwertige bewegliche Güter (z.B. Baumaschinen) einen bestimmten definierten Bereich (z.B. Baustelle, Firmengelände) verlassen („Geofencing“).
- Im Rahmen von Digital Rights Management (DRM) sind Fälle denkbar, in denen digitaler Content mit mobilen Endgeräten nur an bestimmten Orten (z.B. Ländern) wiedergegeben werden kann [GaWo98, Mund05], da z.B. ein Spielfilm vom entsprechenden Anbieter nur für ein bestimmtes Land lizenziert wurde (vgl. auch DVD-Region-Codes). Denkbar wäre auch eine Beschränkung der Lizenz auf ein Firmengelände (z.B. Campus- oder Site-Lizenz). Unter der Berücksichtigung der Ortszeit könnte so auch den einschlägigen jugendschutzrechtlichen Bestimmungen bei der Ausstrahlung digitaler Inhalte ohne zusätzliche Authentifizierungsmaßnahmen entsprochen werden, z.B. Streaming von Filmen mit Altersfreigabe ab 16 Jahren erst ab 22 Uhr Ortszeit gem. Jugendmedienschutz-Staatsvertrag (JMStV) §5.

Die bewusste Manipulation von Ortungsverfahren wird in der Literatur mit „Location Spoofing“ (engl. to spoof: täuschen, reinlegen) bezeichnet, wobei dieser Begriff sowohl für externe (z.B. [WaJo03]) als auch interne Manipulationen (z.B. [Mund05]) Verwendung findet. Der Begriff wurde wohl in Anlehnung an Angriffe gewählt, bei denen in drahtgebundenen Netzwerken eine Adressangabe gefälscht wird (z.B. DNS-, IP- oder ARP-Spoofing). Eine reine Verhinderung der Ortung (Denial-of-Service, DoS) etwa durch Störsender („Jamming“) ist meist nicht so gefährlich wie ein Spoofing-Angriff, da dieser vom Angegriffenen erkannt werden kann. Ein interner Angriff liegt auch vor, wenn ein mobiles Endgerät durch Verlust oder Diebstahl in den Besitz einer unbefugten

¹ Z.B. <http://www.jamesbondfilms.co.uk/tomorrow-never-dies.htm> (letzter Abruf: 17.11.2008)

² <http://www.justiz.baden-wuerttemberg.de/servlet/PB/menu/1229914/index.html> (letzter Abruf: 20.11.2008)

Partei gelangt ist, die dann z.B. unter Laborbedingungen Ortungssignale erzeugen oder Manipulationen an Soft- und/oder Hardware durchführen kann.

Ziel des vorliegenden Artikels ist es, einen Überblick über verschiedene in der Literatur beschriebene grundlegende Anti-Spoofing-Verfahren zugeben. Hierzu wurde auch ein Klassifikationsschema entworfen.

Der verbleibende Teil des vorliegenden Artikels ist wie folgt gegliedert: in Kapitel 2 werden zunächst eine Klassifikation verschiedener Ansätze zur Vermeidung von Spoofing sowie einige für das weitere Verständnis notwendige Konzepte eingeführt. Diese Ansätze werden in den darauf folgenden Kapiteln 3 bis 7 im Einzelnen behandelt. In Kapitel 8 wird kurz auf GPS und Galileo eingegangen, bevor im letzten Kapitel eine Zusammenfassung gegeben wird.

2 Überblick und Grundlagen

Die in der Literatur gefundenen Verfahren zur Vermeidung von Spoofing lassen sich wie folgt klassifizieren (siehe Abbildung 1):

- Plausibilitätskontrollen
- Tamperproof Hardware: Spezielle gegen physische Manipulationen geschützte Hardware
- Location-Keys
- Request-Response-Protokolle
- Funktechnische Maßnahmen

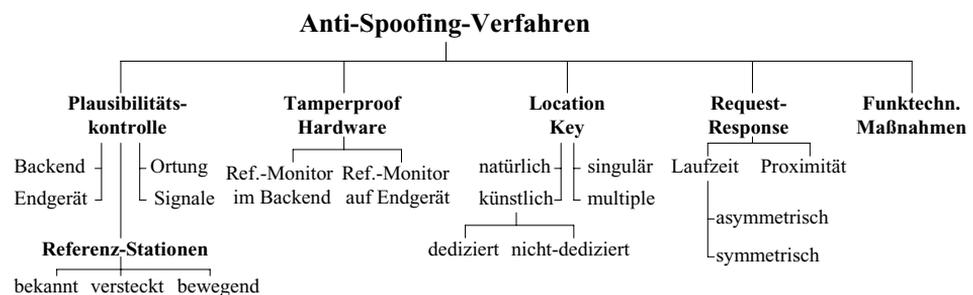


Abbildung 1: Überblick Anti-Spoofing-Verfahren („Anti-Spoofing-Baum“)

Die einzelnen Klassen von Verfahren werden in dieser Reihenfolge in je einem eigenen Kapitel behandelt, wobei dabei noch die einzelnen Unterklassen beschrieben werden. Falls eine entsprechende Aussage möglich ist, wird auch erörtert, ob ein Verfahren nur für Eigenortung (das mobile Endgerät kann selbst die Ortung errechnen) oder Fremdortung (die Ortung wird im stationären Backend des Ortungsnetzwerkes errechnet) ein-

setzbar ist. Weiter wird auch erwähnt, ob die mobilen Endgeräte über bestimmte Ausstattungsmerkmale wie hochpräzise Uhren oder einen drahtlosen Kommunikationskanal mit geringer Latenzzeit verfügen müssen.

Eine Form des Angriffs, die für mehrere Verfahren relevant ist, sind sog. Rerouting-Angriffe (auch „Wormhole-Attack“ genannt): hierbei wird ein Funksignal an einer Stelle empfangen und über ein anderes Medium („out-of-Band“) an einen Sender weitergeleitet (getunnelt). Ein mobiles Endgerät könnte sich hiermit also Signale weiterleitet lassen, die nur an einem anderen Ort empfangbar sind. Rerouting-Angriffe sind generell nur durch die mit ihnen einhergehende Verzögerung bedingt durch die Out-of-Band-Weiterleitung erkennbar. Durch ein Aufspreizen des Signals auf ein möglich breites Spektrum (siehe Kapitel 7) können sie aber technisch erheblich erschwert werden. Für GPS stehen mit sog. GPS-Repeatern oder Reradiatoren schon kommerzielle Produkte zur Verfügung, die die von einer Antenne empfangenen GPS-Signale an einem über Kabel verbundenen Sender wiedergeben („rerouten“), um damit GPS-Empfang in Verkaufs- oder Laborräumen zu ermöglichen³. Beim Replay-Angriff wird ein Signal aufgezeichnet und bewusst zeitverzögert — ggf. an einem anderen Ort — wiedergegeben.

3 Plausibilitätskontrollen

Bei Plausibilitätskontrollen kann unterschieden werden, ob sie auf der errechneten Ortung oder schon den empfangenen „Rohsignale“ ansetzt. Ein weiteres Unterscheidungskriterium ist die Komponente des Ortungssystems, die die Kontrolle vornimmt: hier sind das mobile Endgerät selbst, das stationäre Backend des Ortungsnetzwerkes oder dedizierte Referenzstationen möglich.

3.1 Ebene der Plausibilitätskontrolle

Bei Plausibilitätskontrollen auf Signalebene kann die absolute oder relative Stärke der vom Ortungsnetz empfangenen Funksignale überprüft werden [WaJo03]. Dies ist etwa bei GPS eine gute Möglichkeit, von künstlichen Sendern („Pseudoliten“) ausgestrahlte Signale zu erkennen, da diese meist vielfach stärker als die aus dem All mit einer Stärke von nur noch -160 dBW empfangbaren Signale sind. Aber auch nur ein mäßiger Pegelanstieg der von mehreren Satelliten empfangenen Signale kann ein Hinweis sein, dass ein Spoofing-Angriff gerade eingesetzt hat. Speziell bei GPS sind die Almanachdaten (groben Laufbahnen der Satelliten für etwa drei Monate im Voraus) über das Internet verfügbar, so dass anhand dieser überprüft werden kann, ob die am vermeintlichen Ort sichtbare Satellitenkonstellation plausibel ist.

Eine Plausibilitätskontrolle kann auch erst dann durchgeführt werden, wenn die Ortung bereits errechnet wurde: hier kann etwa anhand mehrerer aufeinander folgender Ortungen betrachtet werden, ob sich der Nutzer mit einer realistischen Geschwindigkeit fortbewegt oder ob plötzlich „Sprünge“ des vermeintlichen Aufenthaltsortes oder Bewegung-

³ Siehe etwa <http://www.gps-world.biz/products/reradproducts.php> oder http://www.navilock.de/produkte/gruppen/22/GPS_Repeater (letzter Abruf jeweils am 12.11.2008)

gen „durch die Wand“ auftreten. Stehen zusätzliche Sensoren am mobilen Endgerät zur Verfügung (Entfernungsmesser, (Kreisel-)Kompass, Beschleunigungs- und Geschwindigkeitsmesser), so kann anhand der von diesen gelieferten Informationen eine Fortschreibung der Ortung vorgenommen werden (Koppelnavigation oder „Dead Reckoning“), so dass spätere Ortsberechnungen des zu überwachenden Ortungssystems auf Plausibilität hin untersucht werden können.

Die erste Form der Plausibilitätskontrolle auf Ebene der „Rohsignale“ eignet sich hauptsächlich für Szenarien mit Eigenortung, die zweite Form kann auch für Fremdortung eingesetzt werden.

3.2 Ort der Plausibilitätskontrolle

Die Plausibilitätskontrolle kann nicht nur vom mobilen Endgerät oder dem Ortungsnetzwerk vorgenommen werden, sondern auch von speziellen Referenzstationen: diese sind eine optionale Erweiterung des Ortungsnetzwerkes mit bekanntem Standort. Für das GPS-System werden weltweit mehrere solcher Stationen betrieben (z.B. in Colorado Springs oder auf den Ascension Islands), um Fehlfunktionen einzelner Satelliten rechtzeitig zu erkennen. Referenzstationen können aber auch eingesetzt werden, um externe Spoofing-Angriffe zu erkennen: hierbei errechnet die Station ihre Position und vergleicht sie mit der bekannten Position; gibt es Abweichungen über einen bestimmten Schwellwert hinaus, so werden die mobilen Endgeräte und/oder Backendsysteme gewarnt. Ein externer Angreifer könnte daher seine Signale zur Überlagerung des eigentlichen Ortungssystems so ausstrahlen, dass die Referenzstation hiervon nicht betroffen ist, sehr wohl aber das anvisierte mobile Endgerät. Es gibt deshalb den Ansatz, versteckte Referenzstationen zu verwenden [CaCS06]. Da solch ein Versteck mit der Zeit aufgedeckt werden kann oder je nach Situation Referenzstationen an unterschiedlichen Orten benötigt werden, können auch mobile Referenzstationen (z.B. in KFZ installiert) eingesetzt werden.

4 Tamperproof Hardware

In der Literatur werden Systeme für den Anwendungsfall beschrieben, dass ein Endgerät nur an bestimmten Orten verschlüsselte Inhalte wie Spielfilme wiedergeben darf [Mund05, GaWo98]. Für das Anwendungsgebiet von Set-Top-Boxen wird sogar gefordert, dass eine Entschlüsselung von ausgestrahlten Inhalten nicht in Gaststätten oder öffentlichen Plätzen möglich sein soll, sondern nur in Privatwohnungen. Da das Endgerät sich hier im Zugriffsbereich des potentiellen Angreifers befindet, ist dies nur unter Verwendung spezieller Hardwarebausteine möglich, die insbesondere auch gegen direkte physische Manipulation resistent sind („Tamperproof Hardware“). Diese Module müssen hierbei die Funktionalitäten für die (Eigen-)Ortung sowie die Entschlüsselung beinhalten.

Das von Mundt in [Mund05] beschriebene System sieht zusätzlich ein Uhrenmodul vor: er geht davon aus, dass die von einem Satellitennavigationssystem ausgestrahlten Systeme einen Zeitstempel beinhalten und digital signiert sind. So soll ein Empfänger Re-

routing-Angriffe erkennen können, da Rerouting zu einer Zeitverzögerung führt, die mit diesen Zeitstempeln erkannt werden können. Um aber die entsprechenden Zeitabweichungen erkennen zu können, ist eine hochpräzise Uhr nötig (wird ein Signal mit Lichtgeschwindigkeit weitergeleitet benötigt es für eine Strecke von z.B. 100 km weniger als 3,4 Millisekunden). Eine hinreichende zeitliche Genauigkeit kann nur mit Atomuhren erreicht werden, die aber für mobile Endgeräte im Konsumerbereich aufgrund der Kosten und des Gewichts nicht in Frage kommen. Im System von Mundt muss deshalb die Uhr im Abstand von ca. drei Stunden neu mit einer externen Zeitquelle unter Verwendung eines speziellen kryptografischen Protokolls synchronisiert werden. Es wird also nicht ständig eine Kommunikationsverbindung benötigt.

Aber auch wenn die Ortung nicht auf dem mobilen Endgerät lokal benötigt wird, sondern auf einem Backendsystem für eine Zugriffsentscheidung durch einen sog. Referenz-Monitor, kann es sinnvoll sein, ein gegen Manipulation geschütztes Hardwaremodul auf dem Endgerät für die Ortung zu haben: in diesem wird nicht nur die Funktionalität zur Errechnung der Eigenortung untergebracht, sondern auch noch ein privater Schlüssel, um die Ortung digital zu signieren, bevor sie an das Backend übertragen wird. Das Backend kann mit dem zugehörigen öffentlichen Schlüssel dann die Authentizität der Ortung überprüfen.

5 Location-Keys

Bei Location-Keys muss das mobile Endgerät eine Information, die nur an bestimmten Orten empfangbar ist, an das stationäre Backend weiterleiten. Dieses Prinzip eignet sich insbesondere, wenn Eigenortung für eine verteilte mobile Anwendung eingesetzt wird, bei der auf dem Backend eine vertrauenswürdige Ortung für Zugriffsentscheidungen benötigt wird.

Wir unterscheiden zunächst, ob es sich um Location-Keys natürlichen oder nicht-natürlichen Ursprungs handelt. Bei den künstlichen Location-Keys kann weiter unterschieden werden, ob sie speziell für die Vermeidung von Spoofing ausgestrahlt wurden (dedizierte Location-Keys); nicht-dedizierte Location-Keys sind also Signale, die primär für einen anderen Zweck ausgestrahlt werden, etwa die eigentliche Ortung oder zur drahtlosen Kommunikation.

Ein Anti-Spoofing-Verfahren, das auf nicht-dedizierten Location-Keys basiert, ist der CyberLocator [DeMa96], bei dem es sich um eine Erweiterung für GPS handelt. Die mobilen Endgeräte müssen hierbei an das stationäre Backend nicht nur die errechnete Ortung weiterleiten, sondern auch noch die von den Satelliten empfangenen „Rohsignale“ (Radio Fingerprint). Diese sind auch unter Kenntnis der aktuellen Satellitenlaufbahnen (Ephemeride) nicht vorhersagbar, da sie durch komplexe atmosphärische Effekte (z.B. Änderung Wetter, Einfluss Ionosphäre) ständig beeinflusst werden. Am Backend werden die vom mobilen Endgerät gemeldeten Signale mit Messungen von vertrauenswürdigen Referenzstationen verglichen. Diese Referenzstationen müssen sich in einer Entfernung von 2.000-3.000 km Entfernung zu dem mobilen Endgeräten befinden. Es sind aber Rerouting-Angriffe denkbar, bei denen das mobile Endgerät von einem am

vorgeblichen Ort befindlichen Empfänger die relevanten Signale weitergeleitet bekommt. Um dies zu verhindern wird gefordert, dass das Endgerät die Location-Keys innerhalb von 5 ms an das Backend liefert (zum Vergleich: für UMTS-HSDPA beträgt die Latenzzeit etwa 150 ms, für drahtgebundenes DSL etwa 20 bis 60 ms). Leider bieten die Autoren des CyberLocators keine Herleitung der max. möglichen Entfernung zwischen Endgerät und Referenzstation, das System ist anscheinend auch nie implementiert worden.

Ein weiterer Ansatz, mit dem Endgeräte ihren Aufenthaltsort in von WLAN abgedeckten Gebieten nachweisen können, wird „Location Aware Access Control“ (LAAC) genannt und ist in [CBGo06] beschrieben. Hierbei werden von Funkbaken spezielle Signale mit begrenzter Reichweite ausgestrahlt, es handelt sich also um dedizierte Location-Keys. Diese Keys bestehen einfach aus langen zufällig erzeugten Bit-Strings, die periodisch erneuert werden. Das Endgerät verknüpft alle empfangenen Schlüssel mittels der XOR-Funktion und wendet auf das Ergebnis eine Hash-Funktion an. Das Ergebnis dieser Berechnung wird an das Backend weitergeleitet. Da das Backend über einen gesicherten Kanal die Location-Keys der einzelnen Basis-Stationen erhält, kann es die gleiche Berechnung wie das Endgerät durchführen und somit die Location-Keys verifizieren. Zusätzlich kann über spezielle Sektor-Antennen der Abstrahlungswinkel dieser Signale eingeschränkt werden. Zwei Baken mit einem Abstrahlungswinkel von jeweils 90 Grad können so angeordnet werden, dass das von beiden überdeckte Gebiet die Form eines Rechtecks hat. Es können also gezielt Flächen wie Ladenlokale oder Firmengelände abgedeckt werden. Die Möglichkeit eines Rerouting-Angriffs wird von den Autoren von LAAC als zu aufwändig erachtet, als dass sie dafür Gegenmaßnahmen vorsehen würden. Lediglich ein Replay-Angriff wird durch das regelmäßige wechseln des Location-Keys verhindert (ein Location-Key sollte eine Lebensdauer in der Größenordnung von höchstens einigen wenigen Sekunden haben). Dies ist aber vor dem Hintergrund zu sehen, dass LAAC speziell dafür entwickelt wurde, eine Zugriffskontrolle für drahtlosen Internetzugriff über WLAN-Hotspots (die gleichzeitig auch die Baken sind) auf Ladenlokale zu beschränken, so dass ein Angreifer, der über eine schnelle drahtlose Kommunikationsverbindung verfügt, die für einen Rerouting-Angriff notwendig wäre, kaum einen Nutzen davon hat, wenn er das LAAC-System überlistet.

Ein ähnlicher Ansatz wird in [Mich02] mit „Pervasive Access Control“ (PAC) beschrieben: hierbei verwendet das mobile Endgerät den von einer Bake empfangenen Location-Key als Schlüssel für eine Hashfunktion, mit der eine Signatur für einen Dienstrequest an einen zentralen Server erzeugt wird. Der Server liefert dann ein sog. „Ticket“ an das Endgerät zurück, das damit gegenüber einem lokalen Gerät (z.B. Drucker) seine Zugriffsberechtigung nachweisen kann.

In [Mala07] ist ein weiterer Ansatz für WLANs beschrieben, mit dem Endgeräte ihren Aufenthaltsort gegenüber einem Backend nachweisen, wenn davon ausgegangen wird, dass nicht autorisierte Nutzer keinen Zutritt zu dem Gelände haben, das mit dem WLAN versorgt werden soll. Hierzu übermittelt das Endgerät seine Position (die z.B. mit GPS oder einem infrarotbasierten Ortungsverfahren berechnet wird) sowie die Signalstärken, mit denen es mehrere Accesspoints (AP) des Netzwerkes empfangen kann, an das Backend. Die gemessenen Signalstärken stellen einen künstlichen und nicht dedizierten

Location-Key dar: dem Backend sind an einzelnen Referenzpositionen innerhalb des Geländes die tatsächlichen Messwerte bekannt, der Angreifer kann diese aber nur abschätzen. Es gibt zwar spezielle Modelle, mit denen die Empfangsstärke eines APs an einem bestimmten Ort vorhergesagt werden kann, hierfür ist es aber notwendig, die Bebauung (Wände) und Möblierung sowie die spezifischen Dämpfungsfaktoren der verwendeten Materialien zu kennen.

Wird ein digitales Signal von einem geostationären Satelliten ausgestrahlt, so unterscheiden sich die in einem bestimmten Zeitintervall an verschiedenen Punkten der Erdoberfläche empfangenen Nachrichten aufgrund der Phasenverschiebung, die aus den unterschiedlichen Entfernungen zum Satelliten resultieren. In [GaWo98] wird ein Verfahren beschrieben, in dem dieser Effekt ausgenutzt wird, um einen Location-Key zu erhalten. Hierzu ist es aber notwendig, dass das Endgerät mit hoher zeitlicher Präzision die Aufzeichnung des Satellitensignals startet und beendet; es werden deshalb von den Autoren verschiedene Wege vorgeschlagen, wie dieser „Aufzeichnungsbefehl“ an die Endgeräte übermittelt werden kann, z.B. über ein Broadcastsignal, das von einem terrestrischen Netzwerk ausgestrahlt wird. Da es sich bei dem aufgezeichneten Signal um ein digitales TV-Programm handelt, liegt hier wieder ein nicht-dedizierter Location-Key vor.

In der Literatur wird noch kein Anti-Spoofing-Verfahren beschrieben, das mit natürlichen Location-Keys arbeitet, jedoch wäre hier prinzipiell die kosmische Hintergrundstrahlung als Key geeignet, da diese — bedingt u.a. durch atmosphärische Effekte — an jedem Ort der Erde zu einem bestimmten Zeitpunkt ein spezifisches Empfangsmuster ergibt.

Neben der Unterscheidung der Herkunft des Location-Keys kann noch weiter unterschieden werden, ob ein Location-Key immer für einen Ort steht (singulärer Location-Key, wie beim CyberLocator), oder ob ein Key durch Überlagerung mit anderen Keys für mehrere Orte stehen kann (multiple Location-Keys, wie bei LAAC).

Der prinzipielle Nachteil von Location-Key-Ansätzen ist, dass das Endgerät hierbei über einen Kommunikationskanal mit geringer Latenzzeit verfügen muss.

6 Request-Response-Protokolle

Request-Response-Protokolle basieren darauf, dass das mobile Endgerät („Prover“) von dem vertrauenswürdigen und stationären „Verifier“ eine Nachricht (Request) mit nicht vorhersehbarem Inhalt erhält, die er sofort beantworten muss (Response). Im einfachsten Fall handelt es sich bei dieser Nachricht um einen zufällig erzeugten Bitstring hinreichender Länge („Nonce“), der einfach reflektiert werden muss. Die Grundidee ist, dass der Prover hierzu nur in der Lage ist, wenn er sich tatsächlich an dem von ihm vorgegebenen Ort aufhält. Ist er nämlich weiter entfernt, so können die Funksignale diese Strecke nicht überbrücken oder die Signallaufzeit ist zu lange. Die Grundannahme bei Auswertung der Signallaufzeiten ist es, dass selbst bei einem Rerouting-Angriff die Nachricht nicht schneller als mit Lichtgeschwindigkeit übertragen werden kann. In der Klassi-

fikation der Anti-Spoofing-Verfahren wird deshalb für „Request-Response“ zwischen „Proximität“ und „Laufzeit“ unterschieden.

Ein schönes Beispiel für solch ein auf Laufzeitmessung beruhendes System wird von Sastry et al. beschrieben [SaSW03]: Der Prover sendet zunächst per Funk eine Nachricht mit seinem angeblichen Aufenthaltsort an den Verifier. Wenn der Verifier sich für diesen Aufenthaltsort für zuständig hält, schickt er ebenfalls per Funk den Request mit dem Nonce und startet die Zeitmessung. Sobald der Prover diesen Request erhält, sendet er ihn zurück an den Verifier, allerdings über Ultraschall. Der Verifier stoppt die Zeitmessung, sobald er die Response erhalten hat und überprüft anschließend, ob die Laufzeit unter Berücksichtigung der Signalgeschwindigkeiten von Funk und Ultraschall hinreichend klein ist.

Die Besonderheit dieses Verfahren ist, dass die drahtlose Datenkommunikation über zwei unterschiedliche Medien — nämlich Funk und Ultraschall — geschieht. Dies ist dadurch motiviert, dass bestimmte Messungenauigkeiten der Laufzeit berücksichtigt werden sollen, insbesondere die Verarbeitungszeit, die der Prover u.U. benötigt, um die Antwort zu erzeugen. Die sich aus der Lichtgeschwindigkeit ergebenden Entfernungsungenauigkeiten sind aber so groß, dass diese das Verfahren unbrauchbar machen würden. Es wird deshalb auf Ultraschallwellen zurückgegriffen, die sich mit einer um sechs Größenordnungen geringeren Geschwindigkeit ausbreiten (331 m/sec gegenüber etwa 3×10^8 m/sec). Eine zeitliche Messungenauigkeit von 0,1 Sekunden (z.B. Reaktionszeit auf Prover) resultiert damit nur in einer räumlichen Ungenauigkeit von ca. 33 m statt ca. 30.0000 km, womit das Verfahren sogar für die meisten Indoor-Anwendungen noch brauchbar wäre. Die langsame Ultraschallkommunikation ist aber prinzipiell mit Rerouting angreifbar: die Autoren wählen deshalb bewusst Ultraschall für den „Rückweg“, da hier ihrer Ansicht nach der technische Aufwand für Rerouting zu groß wäre.

In [WaFe03] findet sich die Beschreibung eines weiteren Verfahrens Request-Response-Verfahrens, es werden jedoch sowohl der Request als auch die Response über Funk übertragen, somit handelt es sich also um ein symmetrisches Verfahren.

Die Kommunikation eines Mobilfunktelefons mit einer Basisstation kann auch als Request-Response-Protokoll aufgefasst werden. Je nach Mobilfunkstandard bietet hierbei auch die Verschlüsselung der Daten auf der Luftschnittstelle gegen externe Angreifer eine gewisse Sicherheit, wobei aber etwa der für GSM verwendete Algorithmus A5/1 gebrochen wurde; der Verschlüsselungsalgorithmus A5/3 für UMTS gilt bisher als sicher. Mobilfunkzellen können aber bis zu 35 km Durchmesser haben, was für viele Anwendungsszenarien zu ungenau ist. Es kann aber zusätzlich die Laufzeit der Signale zwischen Endgerät und Basisstation berücksichtigt werden; diese muss lt. GSM-Protokoll ohnehin ständig bekannt sein, da ein Endgerät mit zunehmender Entfernung von der Basisstation die Übertragung seiner Datenpakete („Bursts“) vorziehen muss, um dem ihm gemäß dem TDMA-Verfahren (Zeitmultiplexing) zugewiesenen Zeitschlitz der Basisstation zu treffen. Ohne diesen „Time Advance“ (TA) genannten Mechanismus könnte es zu Kollisionen von Bursts zweier Endgerät kommen, wenn das Endgerät mit dem direkt folgenden Zeitschlitz wesentlich näher an der Basisstation ist, da der Burst des weiter entfernten Endgeräts aufgrund der großen Laufzeit verspätet eintrifft und

nicht beendet ist, wenn der Burst des zweiten Gerätes die Basisstation erreicht. Über die TA-Messungen kann die Distanz zwischen Endgerät und Basisstation mit einer Genauigkeit von 550 m bestimmt werden. Aus dem Maximalwert der zugehörigen TA-Variablen im GSM-Protokoll ergibt sich auch die Höchstentfernung von ca. 35 km zwischen Basisstation und Endgerät. In [WuLC03] wird dieses System daraufhin untersucht, inwieweit es als Mechanismus gegen Location-Spoofing verwendet werden kann. Sie kommen zu dem Ergebnis, dass kein Spoofing-Schutz vorliegt, wenn die Entfernung zwischen Endgerät und Basisstation geringer als 4.600 m ist. Hierbei wird davon ausgegangen, dass eine Manipulation von clientseitiger GSM-Hardware für die meisten Fälle nicht realistisch ist.

Eine Variation der Proximität-Erkennung (ohne Laufzeitberechnung) findet sich in [VoNe06]: im zu überwachenden Gebiet sind hierbei mehrere Verifier installiert. Gibt ein Endgerät vor, sich in diesem Gebiet zu befinden, muss ein Signal auch von den entsprechenden Verifiern zu empfangen sein. Die Autoren leiten dabei Aussagen her, wie viele Verifier für Gebiete bestimmte Größe und Form benötigt werden. Zusätzlich gibt es aber auch „Rejection Verifier“ außerhalb des Schutzgebietes: empfängt ein solcher Rejector das Signal des Provers, so wird der Zugriff verweigert.

7 Funktechnische Maßnahmen

Es gibt spezielle funktechnische Maßnahmen, die das Überlagern oder Auslöschen der Ortungssignale durch Angreifer erschweren. Die entsprechenden Arbeiten sind dem Bereich der Nachrichtentechnik zuzuschreiben und sollen deshalb nur kurz angerissen werden:

- Frequenzspreizung (Spread-Spectrum-Verfahren): Hierbei wird ein schmalbandiges Signal in ein Signal mit einer größeren Bandbreite umgewandelt. Durch die erhöhte Bandbreite ist es dann für einen Angreifer wesentlich aufwändiger, das Signal zu stören [PiSM82]. Konkrete Frequenzspreizungsverfahren sind etwa *Frequency Hopping Spread Spectrum* (FHSS) oder *Direct Sequence Spread Spectrum* (DSSS). Frequenzspreizung kann aber auch eingesetzt werden, um das Signal robuster gegenüber natürlichen Störungen (z.B. Interferenzen) zu machen.
- Manchester-Coding: Für einen Angreifer ist es schwieriger, ein gesendetes Bit (1) „auszulöschen“ als ein nicht gesetztes Bit (0) zu setzen. Beim sog. Manchester-Coding wird deshalb jedes Bit der Originalnachricht auf zwei Bits abgebildet (0→01, 1→10), so dass ein Angreifer, der keine gesendeten Bits auslöschung kann, nicht Nachrichten derart manipulieren kann, so dass dies der Empfänger nicht entdeckt. [CaRC07]
- Erkannte Störsender können unter Verwendung einer speziellen Antenne mit Nullsteuerung (sog. „Phased Array Antenna“) unterdrückt werden. [DoHä04, 172]
- Ähnlich der Frequenzspreizung ist der Ansatz, für die Signale verschiedener Sender in Ortungsnetzwerken unterschiedliche Frequenzen zu verwenden. Beim russischen GLONASS-System etwa sendet jeder Satellit auf einer eigenen Frequenz [DoHä04,

172]. Beim US-amerikanischen GPS hingegen senden alle Satelliten auf den gleichen Frequenzen, da die Navigationsnachrichten aber mit dem CDMA-Verfahren codiert sind ist es für den Empfänger trotzdem möglich, einzelne Satelliten „herauszuhören“.

8 Absicherung von GPS und Galileo

GPS ist das wohl bekannteste Ortungssystem, das inzwischen auch von vielen Privatanutzern tagtäglich für Navigation eingesetzt wird. Alle nominal 24 Satelliten des Systems strahlen auf denselben Frequenzen L1 und L2 Navigationsnachrichten aus, die u.a. die Nummer des jeweiligen Satellits oder die Bahnlaufdaten (Ephemeride) enthalten. Um diese Nachrichten mit CDMA zu modulieren, stehen zwei Codes zur Verfügung: der Coarse Acquisition Code (C/A) für den auf zivile Nutzung ausgelegten Standard Positioning Service (SPS) und der Precise Code (P-Code), der militärischen Nutzern für den sog. Precise Positioning Service (PPS) vorbehalten sein sollte und durch die Verwendung des geheimen Y-Schlüssels zum P(Y)-Code wird [DoHä04, 173f; Küpp05, 165ff]. Während der C/A-Code nur auf der L1-Frequenz ausgestrahlt wird, wird der P(Y)-Code sowohl auf der L1- als auch der L2-Frequenz ausgestrahlt. Der P(Y)-Code kann als symmetrische Chiffre aufgefasst werden, da zur Ver- als auch Entschlüsselung der gleiche Schlüssel Y verwendet wird. Weil er eine um den Faktor 10 höhere Frequenz als der C/A-Code hat, lässt sich mit ihm auch eine deutlich bessere Ortungsgenauigkeit erzielen. Dieser Schlüssel muss auf allen autorisierten GPS-Empfängern vorhanden sein; wird nur eines dieser Geräte kompromittiert, so könnte der Angreifer eigene mit dem P(Y)-Code verschlüsselte Navigationsnachrichten ausstrahlen, um eine Spoofing-Attacke durchzuführen. Da dieselbe Navigationsnachricht mit dem öffentlichen C/A-Code als auch dem P(Y)-Code verschlüsselt wird, kann auch ein Known-Plaintext-Angriff durchgeführt werden. Der Y-Schlüssel wird deshalb alle 24h Stunden ausgetauscht (Rekeying). Jenseits von Spoofing-Angriffen sind Angriffe auf den P(Y)-Code aber mittlerweile uninteressant, weil es mit sog. Codeless Receivern/Kinematic Receivern auch ohne dessen Kenntnis möglich ist, die höhere Ortungsgenauigkeit des PPS zu erreichen [DoHä04, 229f]. Weiter sind für GPS mit „RAIM“ (Receiver Autonomous Integrity Monitoring) verschiedene Formen der Plausibilitätskontrolle der empfangenen Signale auf dem Endgerät verfügbar, für die aber redundante Messungen (z.B. Sichtbarkeit von mehr als den vier systembedingt notwendigen Satelliten) verfügbar sein müssen [Lang99].

Das von der EU geplante Galileo-System soll vergleichbar den beiden GPS-Diensten verschiedene — teilweise kostenlose — Dienste anbieten, die sich in Genauigkeit oder Integritätsgarantien unterscheiden. Um Spoofing zu vermeiden ist geplant, die von einigen Diensten ausgestrahlten Navigationsnachrichten mit einem Public-Key-Verfahren digital zu signieren: hierbei wird die Nachricht vom Satelliten unter Verwendung des geheimen Schlüssels signiert, der Empfänger kann dann mit dem zugehörigen öffentlichen Schlüssel verifizieren, dass das Signal tatsächlich von einem Galileo-Satelliten ausgestrahlt wurde. Ein Angreifer müsste an den nur in der Kontrollstation oder auf den Satelliten gespeicherten geheimen Schlüssel gelangen, um selbst erzeugte Navigationsnachrichten digital zu signieren. Um ein Rerouting anhand des ebenfalls signierten Zeitstempels zu erkennen, müsste der Empfänger aber über eine hochpräzise Uhr verfügen.

9 Zusammenfassung und Ausblick

Anhand eines Klassifikationsschemas wurden verschiedene Prinzipien vorgestellt, mit denen bewusste Manipulationen von Ortungsverfahren verhindert werden können. Nicht behandelt wurden Ansätze, die speziell für infrastrukturlose Netzwerke (Sensornetzwerke, MANETs) entwickelt wurden, z.B. [CaCS06]. Neben verschiedenen Formen der Plausibilitätskontrolle, Location Keys und Request- und Response-Protokollen wurden auch manipulationssichere Hardware und spezielle funktechnische Maßnahmen betrachtet.

Plausibilitätskontrollen können sowohl bei Eigen- als auch Fremddortung durchgeführt werden, wobei eine Auswertung auf Signalebenen vor allem für ersteres geeignet ist. Manipulationssichere Hardware ist immer dann unabdingbar, wenn die Zugriffsentcheidung selbst lokal auf dem mobilen Endgerät durchgeführt werden soll, wie dies etwa bei DRM-Szenarien der Fall ist. Location-Key-Verfahren können durch Rerouting angegriffen werden; wenn durch die Ortung aber nur der Zugriff auf Internetzugriff (z.B. über WLAN-Hotspot) gesteuert werden soll ist dieses Verfahren ausreichend. Bei Request-Response-Protokollen mit Laufzeitmessungen werden hochpräzise Uhren benötigen, bei proximitätsbasierenden Ansätzen sind komplexe und teilweise auch dynamische Abschattungen zu berücksichtigen (z.B. durch Gebäude, Passanten oder Fahrzeuge). Auch eine Kombination mehrerer Ortungsverfahren ist denkbar, da es für den Angreifer dann aufwändiger wird, mehrere Verfahren gleichzeitig so zu manipulieren, dass keine Inkonsistenzen auftreten.

Eine generelle Empfehlung für ein Anti-Spoofing-Verfahren kann nicht gegeben werden, es müssen immer die für den jeweiligen Anwendungsfall gegebenen Anforderungen und Möglichkeiten berücksichtigt werden.

Literatur

- [CaRC07] Capkun, S.; Rasmussen, K.B.; Cagalj, M.: SecNav: Secure Broadcast Localization and Time Synchronization in Wireless Networks. International Conference on Mobile Computing and Networking (MobiCom), 2007, 310-313.
- [CaCS06] Capkun, S.; Cagalj, M.; Srivastava, M.: Secure Localization with Hidden and Mobile Base Stations. IEEE International Conference on Computer Communications (INFOCOM), 2006, 1-10.
- [CBGo06] Cho, Y.; Bao, L.; Goodrich, M.: LAAC: A Location-Aware Access Control Protocol. Third Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services, 2006, 1-7.
- [DaBP07] Damiani, M.L.; Bertino, E.; Perlasca, P.: Data security in location-aware applications. An approach based on RBAC. International Journal on Information and Computer Security, 1(1/2), 2007, 5-38.
- [DeMa96] Denning, D.; MacDoran, P.: Location-Based Authentication: Grounding Cyberspace for Better Security. Computer Fraud & Security, Elsevier, February 1996, 12-16.

- [DoHä04] Dodel, H.; Häupler, D.: Satellitennavigation. Hüthig-Verlag, Bonn, 2004.
- [GaWo98] Gabber, E.; Wool, A.: How to Prove Where You Are: Tracking the Location of Customer Equipment. Proceedings of the 5th ACM Conference on Computer and Communications Security, 1998, 142-147.
- [Küpp05] Küpper, A.: Location-based Services. Fundamentals and Operations. Wiley & Sons, Chichester, U.K., 2005.
- [Lang99] Langley, R. B.: The Integrity of GPS. GPS World, March, 1999, 60-63.
- [Mala07] Malaney, R.A.: Securing Wi-Fi Networks with Position Verification (Extended Version). International Journal of Security Networks, 2(1-2), 2007, 27-36.
- [Mich02] Michalakakis, N.: PAC: Location Aware Access Control for Pervasive Computing Environments. Proceedings of the Second Student Oxygen Workshop, Gloucester, MA, USA, 2002.
- [Mund05] Mundt, T.: Location Dependent Digital Rights Management. Proceedings of the 10th IEEE Symposium on Computers and Communications (ISCC 2005), 2005.
- [PiSM82] Pickholtz, R. L.; Schilling, D. L.; Milstein, L. B.: Theory of Spread-Spectrum Communications — A Tutorial. IEEE Transactions on Communications, 30(5), 1982, 855-884.
- [SaSW03] Sastry, N.; Shankar, U.; Wagner, D.: Secure Verification of Location Claims. Proceedings of the Conference on Wireless Security (WiSe), 2003, 1-10.
- [TuPo04] Turowski, K.; Pousttchi, K.: Mobile Commerce — Grundlagen und Technik. Springer-Verlag, Berlin et al., 2004.
- [VoNe06] Vora, A.; Nesterenko, M.: Secure Location Verification Using Radio Broadcast. IEEE Transactions on Dependable and Secure Computing, 3(4), 2006, 377-385.
- [WaFe03] Waters, B. R.; Felten, E. W.: Secure, Private Proofs of Location. Technical Report TR-665-03, Dep. of Computer Science, Princeton University, 2003.
- [WaJo03] Warner, S. W.; Johnston, R.G.: GPS Spoofing Countermeasures. Technical Report LAUR-03-6163, Los Alamos National Laboratory, U.S.A., 2003.
- [WuLC03] Wullems, C.; Looi, M.; Clark, A.: Enhancing the Security of Internet Applications using Location: A New Model for Tamper-resistant GSM Location. Proceedings of the 8th IEEE International Symposium on Computers and Communications (ISCC' 03), 2003, 1251-1258.

Endkundenakzeptanz mobilkommunikationsbasierter Zahlungssysteme

Klaus Kornmeier

mayato GmbH
Am Borsigturm 9
13507 Berlin

klaus.kornmeier@mayato.com

Abstract: Angesichts der Fülle an gescheiterten Versuchen, mobilkommunikationsbasierte Zahlungssysteme (MBZS) als Ergänzung oder Alternative zu gängigen Zahlungsverfahren im Endkundenmarkt zu etablieren, beabsichtigt die Studie einen Beitrag zur Erklärung endkundenseitiger Akzeptanzprozesse innerhalb des Mobile Commerce und zur Erschließung des betriebswirtschaftlichen Potentials von MBZS zu leisten. Den Ausgangspunkt bildet eine von theoretischen und sachlogischen Überlegungen geleitete Formulierung von 16 Hypothesen zu Einflussfaktoren der MBZS-Akzeptanz. Diese werden in ein Akzeptanzmodell überführt, welches genutzt werden kann, um Präferenzen von Endkunden zu erkennen, Gründe für unterschiedliche Akzeptanzniveaus zu verstehen sowie Determinanten des wahrgenommenen Risikos zu identifizieren. Die strukturanalytische Überprüfung des Akzeptanzmodells erfolgt mittels PLS unter Verwendung von im Internet erhobenen Primärdaten.

1 Einleitung

Infolge der Ausweitung der kommerziellen Nutzungsmöglichkeiten des Mobilfunks sind in den letzten Jahren sogenannte *mobilkommunikationsbasierte Zahlungssysteme (MBZS)* immer stärker in den Blickpunkt von Forschung und Praxis gerückt. Mittels dieser, auch als *M(obile)-Payment* bezeichneten, Ausprägungsform elektronischer Zahlungssysteme werden Endkunden in die Lage versetzt, unter Rückgriff auf ein mobiles Endgerät, für den Erwerb von Produkten, die Inanspruchnahme von Dienstleistungen oder den Bezug elektronischer Inhalte bezahlen zu können. Nutzungsintensität und Transaktionsumsätze der bislang am Markt verfügbaren MBZS haben sich jedoch nicht so schnell und umfassend entwickelt, wie von vielen Experten erwartet [KSS08, Ma07]. Dies obwohl (oder gerade weil) es in den vergangenen Jahren eine kaum überschaubare Anzahl an Versuchen gab, MBZS als Ergänzung oder Alternative zu gängigen Zahlungsverfahren im Endkundenmarkt zu etablieren.¹

¹ Laut [Ch08] existier(t)en allein in Europa über 180 verschiedene MBZS.

Aber auch wenn MBZS bislang auf eine nur eine sehr geringe Akzeptanz stoßen, belegt eine Vielzahl empirischer Untersuchungen übereinstimmend ein nach wie vor großes generelles Interesse der Endkunden an einer MBZS-Nutzung [SCA07]. Vor diesem Hintergrund kann argumentiert werden, dass für MBZS-Anbieter die Notwendigkeit besteht, ein tieferes Verständnis der Akzeptanzbildung zu erlangen. Infolgedessen beabsichtigt die vorliegende Studie für die betriebswirtschaftliche Forschung einen Beitrag zur Erklärung endkundenseitiger Akzeptanzprozesse innerhalb des Mobile Commerce und für die Unternehmenspraxis einen Beitrag zur Erschließung des betriebswirtschaftlichen Potentials von MBZS zu leisten. Um die skizzierten Ziele zu erreichen bietet sich folgende Vorgehensweise an: Zunächst werden zentrale Ergebnisse früherer empirischer Studien zur MBZS-Akzeptanz analysiert. Darauf basierend werden für die Akzeptanzbildung relevante Konstrukte identifiziert und Hypothesen zu Wirkungsbeziehungen zwischen diesen Konstrukten postuliert. Die erarbeiteten Zusammenhänge werden anschließend zu einem Pfadmodell der MBZS-Akzeptanz verknüpft, welches mittels des Partial Least Squares (PLS-)Verfahrens empirisch ausgewertet wird. Der Beitrag endet mit einer Diskussion der wesentlichen Befunde der Analyse sowie einem Ausblick auf offene Forschungsfragen.

2 Zentrale Ergebnisse empirischer Arbeiten zur MBZS-Akzeptanz

Trotz der Lücke zwischen der erwarteten Nachfrage einer- und den tatsächlichen MBZS-Nutzerzahlen andererseits, existieren bislang nur wenige wissenschaftlich fundierte Erkenntnisse über die Bereitschaft von Endkunden zu einer MBZS-Nutzung und insbesondere zu Faktoren, welche die Nutzungsentscheidung signifikant fördern oder hemmen [Ch08]. Ein Großteil der veröffentlichten empirischen Arbeiten ist auf eine rein deskriptive Auswertung der erhobenen Daten beschränkt. Nur selten wird eine Modellierung und Analyse komplexer Wirkungszusammenhänge vorgenommen.

Abbildung 1 gibt einen Überblick über die wichtigsten Resultate von 11 MBZS-Studien, die auf Primärdaten basieren und auf eine empirische Identifikation von Akzeptanzfaktoren aus der Endkundenperspektive abzielen. Im Ergebnis ist festzustellen, dass unabhängig von der konkreten technischen Realisierung ein Bündel von *Leistungsmerkmalen*, *Sicherheitsbedenken* und *Aufwandsüberlegungen* von zentraler Bedeutung für die Akzeptanz eines MBZS sind. Darüber hinaus hängt die Nutzungsbereitschaft stark von der jeweiligen *Zahlungssituation* ab. Speziell mit Blick auf akzeptanzfördernde Leistungsmerkmale lassen sich mehrere Eigenschaften identifizieren, die als relativer Vorteil von MBZS zu werten sind [Ma07, PW05]: Vermeidung von Wartezeiten/-schlangen, zeit- und ortsunabhängige Bezahlung, Zahlungen in Situationen, in denen herkömmliche Zahlungssysteme nicht verfügbar sind, Bargeldersatz (und damit auch Vermeidung von Wechselgeld- und gegebenenfalls von Währungsproblemen), kürzere Dauer des Bezahlvorgangs sowie die Möglichkeit zu Zusatzangeboten.

Autor(en)	Geographische Ausrichtung	Methodik	Ergebnisse	Anmerkungen
[PNT02]	Norwegen	<ul style="list-style-type: none"> • Internetbefragung: 320 Teilnehmer • E-Mail Befragung: 175 Teilnehmer • Strukturgleichungsmodell auf Basis der TPB* 	<ul style="list-style-type: none"> • Erklärung der Nutzungsabsicht durch Konstrukte Vergnügen, Ausdrucksfähigkeit und Verhaltenskontrolle • Kein Einfluß der Konstrukte Nützlichkeit, Einstellung und subjektive Norm auf Nutzungsabsicht 	<ul style="list-style-type: none"> • Keine Ableitung von konkreten Empfehlungen für eine akzeptanzfördernde MBZS-Gestaltung
[KPW03]	Deutschland, Österreich, Schweiz	<ul style="list-style-type: none"> • Internetbefragung • 4.432 Teilnehmer • Deskriptive Auswertung 	<ul style="list-style-type: none"> • Wichtigste MBZS-Nutzungsgründe sind Ersatz anderer Zahlungssysteme (38,0%), betragsspezifische Ausrichtung (17,7%), einfache Bedienung (17,4%) • Wichtigste Ablehnungsgründe sind subjektive Unsicherheit (33,2%), Bevorzugung alternativer Zahlungsverfahren (18,7%), Unkenntnis (12,2%) • Akzeptanz ist am höchsten bei Transaktionen im Mobile Commerce und beim Automatenverkauf • Wichtigste Akzeptanzfaktoren sind vertraulicher Umgang mit persönlichen Daten (96,2%), einfache Handhabung (93,0%), keine bzw. geringe Kosten (91,6%) • Mit EC-Karte vergleichbare Kosten und Vorgangsdauern werden akzeptiert • Banken sind bevorzugte MBZS-Betreiber 	<ul style="list-style-type: none"> • Betrachtung mehrerer Transaktions-situationen
[CP04]	Korea	<ul style="list-style-type: none"> • Internetbefragung • 1.034 Teilnehmer • Strukturgleichungsmodell auf Basis TAM* 	<ul style="list-style-type: none"> • Einstellung, Attraktivität von Alternativen und Lernkosten sind wichtigste Determinanten der Nutzungsabsicht • Nutzen und nutzungserleichternde Bedingungen sind wichtigste Determinanten der Einstellung 	<ul style="list-style-type: none"> • Teilnehmer überwiegend junge Männer mit höherer Schulbildung
[ELP04]	Deutschland, Österreich, Schweiz	<ul style="list-style-type: none"> • Internetbefragung • 6.343 Teilnehmer • Deskriptive Auswertung 	<ul style="list-style-type: none"> • Akzeptanz ist am höchsten bei Transaktionen im Electronic Commerce und beim Automatenverkauf • Wichtigste Akzeptanzfaktoren sind vertraulicher Umgang mit persönlichen Daten (97,1%), keine bzw. geringe Kosten pro Transaktion (94,6%) sowie keine bzw. geringe monatliche Grundgebühr (93,7%) 	<ul style="list-style-type: none"> • Betrachtung mehrerer Transaktions-situationen
[KSS06]	Norwegen	<ul style="list-style-type: none"> • Internetbefragung • 462 Teilnehmer • Deskriptive Auswertung 	<ul style="list-style-type: none"> • Nutzung wird gefördert durch möglichst viele Akzeptanzstellen (69,9%), verständlichere Zahlungsbelege (56,6%) und höhere Bedienungsfreundlichkeit (53,7%) 	<ul style="list-style-type: none"> • 44,2% der Teilnehmer nutzen elektronische Zahlungssysteme und MBZS
[Ma06]	Finnland	<ul style="list-style-type: none"> • Postalische Befragung • 360 Einwohner Helsinkis • Strukturgleichungsmodell auf Basis TAM* 	<ul style="list-style-type: none"> • Zahlungssituation beeinflusst Nutzungsabsicht • Benutzungsfreundlichkeit und Kompatibilität fördern Nutzungsabsicht • Mobilität und Nützlichkeit beeinflussen Nutzungssituation 	<ul style="list-style-type: none"> • Betrachtung der Nutzung von Mobile Ticketing
[Ma07]	Finnland	<ul style="list-style-type: none"> • Fokusgruppendifkussion • 6 Gruppen mit insgesamt 46 Teilnehmern • Deskriptive Auswertung 	<ul style="list-style-type: none"> • MBZS-spezifische Operationalisierung und Ergänzung der Rogers-Kriterien: <ul style="list-style-type: none"> – Relativer Vorteil: Zeit- und ortsunabhängige Bezahlung, Vermeidung von Warteschlangen, Bargeldersatz – Kompatibilität: Bei digitalen Inhalten und Kleinbetragszahlungen am POS gegeben – Komplexität: Verschiedene Konten, aufwendige Registrierungsverfahren – Netzefekte: Fehlende Akzeptanzstellen, Proprietäre Dienste – Vertrauen: Vertrauen in Händler, in Mobilfunknetzbetreiber und in Finanzdienstleister – Wahrgenommenes Risiko: Unberechtigte Nutzung, Übertragungsfehler, Datenschutz • Situationsabhängigkeit des relativen Vorteils • Situationsabhängige Bereitschaft zur Nutzung von MBZS: E-Tickets, mobile Inhalte, Automatenzahlungen, Kleinbetragszahlungen am POS 	<ul style="list-style-type: none"> • Kleine Stichprobe

Tabelle 1: Ergebnisse empirischer Untersuchungen zur MBZS-Akzeptanz

Autor(en)	Geographische Ausrichtung	Methodik	Ergebnisse	Anmerkungen
[PW07]	Deutschland	<ul style="list-style-type: none"> • Internetbefragung • 1.104 Teilnehmer • Regressionsanalyse eines TAM-basierten Pfadmodells 	<ul style="list-style-type: none"> • Nützlichkeit und Benutzungsfreundlichkeit fördern Nutzungsabsicht • Benutzungsfreundlichkeit fördert Nützlichkeit • Aufgabe-Technologie Übereinstimmung fördert Nützlichkeit • Kein Einfluß von Vertrauenswürdigkeit und Vertraulichkeit auf Nutzungsabsicht 	<ul style="list-style-type: none"> • Relativ große Stichprobe • Transaktionssituationen: Supermarkt und Parkticket • Keine Ableitung von konkreten Empfehlungen für eine akzeptanzfördernde MBZS-Gestaltung
[Ch08]	USA	<ul style="list-style-type: none"> • Internetbefragung • 299 Teilnehmer • Strukturgleichungsmodell auf Basis TAM 	<ul style="list-style-type: none"> • Kompatibilität fördert Nutzungsabsicht am stärksten • Benutzungsfreundlichkeit und Nützlichkeit fördern, wahrgenommenes Risiko hemmt Nutzungsabsicht • Wahrgenommenes Risiko wird stärker durch sicherheits- als durch datenschutzbezogene Bedenken beeinflusst • Nützlichkeit wird stärker durch Einfachheit als durch Geschwindigkeit des MBZS beeinflusst 	<ul style="list-style-type: none"> • Keine Differenzierung nach Transaktionssituationen und Betragshöhen
[KSS08]	Norwegen	<ul style="list-style-type: none"> • Internetbefragung • 462 Teilnehmer • Regressionsanalyse 	<ul style="list-style-type: none"> • Nichtvorhandensein von MBZS stellt keine Barriere für die Nutzung von M-Commerce Diensten dar • Der Faktor Mobilität beeinflusst die Adoption nur gering • Anonymität ist für die Mehrzahl der Teilnehmer nicht wichtig • Teilnehmer bevorzugen ihnen vertraute Zahlungssysteme 	<ul style="list-style-type: none"> • 44,2% der Teilnehmer nutzen elektronische Zahlungssysteme und MBZS
[WGP 08]	Deutschland	<ul style="list-style-type: none"> • Internetbefragung • 965 Teilnehmer • Deskriptive Auswertung 	<ul style="list-style-type: none"> • Hausbank und „renommierte“ Banken werden als MBZS-Anbieter präferiert • MBSZ sind am attraktivsten für Beträge bis maximal 25 Euro • Bei 42,8% der Teilnehmer würde ein attraktiv gestaltetes Bonusprogramm zu einer erhöhten Nutzung von MBZS führen 	<ul style="list-style-type: none"> • Stichprobe enthält nur Teilnehmer, die eine MBZS-Nutzung in Betracht ziehen

* TAM = Technology Acceptance Model, TPB = Theory of Planned Behavior.

Tabelle 1: Ergebnisse empirischer Untersuchungen zur MBZS-Akzeptanz (Forts.)

3 Hypothesengerüst zur MBZS-Akzeptanz

Der in der betriebswirtschaftlichen Akzeptanzforschung häufig festzustellende schwerpunktmäßige Rückgriff auf sozio-psychographische Konstrukte führt dazu, dass nur selten Faktoren identifiziert werden, aus denen sich direkte Handlungsempfehlungen für die Praxis ableiten lassen. Das in diesem Abschnitt beschriebene Hypothesengerüst versucht diesem Kritikpunkt zu begegnen, indem grundlegende Erkenntnisse der Akzeptanzforschung mit den im vorangegangenen Abschnitt dargestellten Ergebnissen bisheriger empirischer Forschungsarbeiten zusammengeführt werden.

Ausgangspunkt der Modellbildung ist die sogenannte Einstellungs-Verhaltenshypothese, welche besagt, dass Einstellungen das Verhalten bestimmen und die Nutzungswahrscheinlichkeit eines MBZS somit von der Stärke der positiven Einstellung abhängig ist. Die in mehrdimensionalen Einstellungskonzepten enthaltene konative Komponente wird als eigenständiges Konstrukt Nutzungsabsicht ausgegliedert, welches sich von der Einstellung durch eine stärker zeitlich- und situationsorientierte Ausrichtung unterscheidet und angibt, für wie wahrscheinlich ein Nachfrager eine MBZS-Nutzung hält. Gestützt wird diese Vorgehensweise durch Untersuchungen, in denen die grundsätzliche Eignung der Nutzungsabsicht als Prädiktor einer tatsächlichen Nutzung bestätigt wird [KRK06]. Somit kann folgende Hypothese formuliert werden:

H₁: Je positiver die Einstellung einer Person gegenüber einer Nutzung von MBZS ausgeprägt ist, desto höher ist ihre Absicht, MBZS zu nutzen.

Bevor Endkunden ein MBZS tatsächlich nutzen können, muss ein Systemanschluss erfolgen. Dieser kann mit einem gewissen Aufwand, wie etwa dem Kauf und der Installation neuer Hardware oder dem Durchlaufen eines Registrierungsprozesses verbunden sein [Ge03]. Zur Beantwortung der Frage, inwieweit sich die Nutzungsabsicht in einer Bereitschaft zur Übernahme monetärer und nicht-monetärer Anschluss- und Nutzungskosten niederschlägt wird das Konstrukt Aufwandsbereitschaft eingeführt, das als von der Nutzungsabsicht determiniert angenommen wird. Die entsprechende Hypothese lautet:

H₂: Je höher die Absicht einer Person zur Nutzung von MBZS ausgeprägt ist, desto höher ist ihre Aufwandsbereitschaft.

Gemäß den Überlegungen von Rogers [Ro03] zu Adoptionsfaktoren wird des Weiteren ein Einfluss der Leistungsmerkmale eines MBZS auf die Einstellung unterstellt. Ein wesentlicher Vorteil bei der Integration von Leistungsmerkmalen in die Hypothesen- bzw. Modellbildung liegt darin, dass mittels einer empirischen Erfassung subjektiver Merkmalswichtigkeiten explizite Eingriffspunkte für eine akzeptanzfördernde Gestaltung von MBZS identifiziert werden können. Folglich ist als Hypothese zu formulieren:

H₃: Je mehr ein MBZS die von einer Person subjektiv als wichtig beurteilten Leistungsmerkmale aufweist, desto wahrscheinlicher besitzt diese Person eine positive Einstellung gegenüber der Nutzung.

Aufgrund der eingangs erwähnten Situationsabhängigkeit der Einstellung kann davon ausgegangen werden, dass diese in Abhängigkeit von den Zahlungssituationen, für die ein MBZS ausgelegt ist, variiert [Ma07]. Somit lautet die Hypothese H₄:

H₄: Je mehr ein MBZS zur Zahlungsabwicklung in von einer Person als subjektiv wichtig beurteilten Zahlungssituationen geeignet ist, desto wahrscheinlicher besitzt diese Person eine positive Einstellung gegenüber der Nutzung.

Konstrukte zur Erfassung sozialer Einflüsse sind Bestandteil vieler Akzeptanzmodelle. In der Regel finden sie in Form der subjektiven Norm Eingang in der Modellbildung, d.h. des von einer Person wahrgenommenen sozialen Drucks, der sich daraus ergibt, dass andere, für sie wichtige Personen die Ansicht vertreten, sie solle ein bestimmtes Verhalten ausführen [LYY05]. In Übereinstimmung mit der *Theory of Planned Behavior* wird in der eigenen Studie ein Zusammenhang zwischen subjektiver Norm und Einstellung postuliert:

H₅: Je deutlicher eine, der MBZS-Nutzung positiv gegenüberstehende, subjektive Norm von einer Person wahrgenommen wird, desto wahrscheinlicher besitzt diese Person eine positive Einstellung gegenüber der Nutzung.

Den Erkenntnissen der Diffusionsforschung folgend, ist bei der Frage nach sozialen Einflüssen auch die Wirkung der aus interpersonellen und medial vermittelten Einflüssen bestehenden Kommunikationskanäle einer Person zu berücksichtigen [LYY05].

Empirische Untersuchungen belegen, dass interpersonelle und medial vermittelte Einflüsse über das Konstrukt der subjektiven Norm auf die Einstellung wirken [Pe05]. Aus diesem Grund werden zwei Einzelhypothesen formuliert:

H₆: Je stärker positive MBZS-bezogene interpersonelle Einflüsse von einer Person wahrgenommen werden, desto stärker ist ihre subjektive Norm gegenüber einer Nutzung ausgeprägt.

H₇: Je stärker positive MBZS-bezogene medial vermittelte Einflüsse von einer Person wahrgenommen werden, desto stärker ist ihre subjektive Norm gegenüber einer Nutzung ausgeprägt.

Neben der Einstellung wird in der Literatur auch dem Involvement, d.h. dem inneren Engagement, mit dem sich Endkunden einem Produkt zuwenden, eine maßgebliche Wirkung auf die Nutzungsabsicht zugeschrieben. Den, aus einer empirischen Analyse der Adoption von Online-Banking gewonnenen, Ergebnissen von Harms [Ha03] folgend, wird das Involvement als ein die Einstellung ergänzender Einflussfaktor auf die Verhaltensabsicht in die eigene Hypothesenbildung integriert:

H₈: Je stärker das Involvement einer Person in Bezug auf MBZS ausgeprägt ist, desto stärker ist ihre Nutzungsabsicht.

Unter Bezugnahme auf vorgelagerte, das Involvement auslösende Stimuli werden im Schrifttum verschiedene Involvementdeterminanten unterschieden. Als personenspezifischer Faktor wird das Konstrukt der persönlichen Innovationsneigung mit in das Hypothesengerüst aufgenommen. Risikoüberlegungen der Endkunden finden durch das Konstrukt des wahrgenommenen Risikos Berücksichtigung, welches den Einfluss der in empirischen Untersuchungen vielfach geäußerten Sicherheitsbedenken bei der Nutzung von MBZS abbilden soll. Die damit korrespondierenden Hypothesen lauten:

H₉: Je stärker die MBZS-spezifische persönliche Innovationsneigung einer Person ausgeprägt ist, desto stärker ist ihr Involvement in Bezug auf MBZS.

H₁₀: Je mehr eine Person die Nutzung von MBZS als riskant wahrnimmt, desto niedriger ist ihr Involvement in Bezug auf MBZS.

Hinsichtlich einer Betrachtung von Faktoren, die zu einer Reduzierung des wahrgenommenen Risikos beitragen, wird vielfach auf die entscheidende Rolle des Vertrauenskonstrukts hingewiesen. Empirische Untersuchungen belegen insbesondere die Bedeutung des Vertrauens in den Leistungsanbieter für die individuelle Akzeptanzbildung [Ca08]. Neben Banken, Kreditkartenunternehmen oder MBZS-Spezialisten erscheinen Mobilfunknetzbetreiber besonders geeignet, die Rolle eines MBZS-Anbieters einzunehmen [Ge03]. Unter Bezugnahme auf letztere wird die Hypothese formuliert:

H₁₁: Je stärker das Vertrauen einer Person in Mobilfunknetzbetreiber als Anbieter von MBZS ausgeprägt ist, desto weniger nimmt sie die Nutzung von MBZS als riskant wahr.

Auch wenn die empirische Datenlage uneinheitlich ist, wird in der Regel angenommen, dass eine hohe Risikowahrnehmung mit einem verstärkten Informationsbedürfnis einhergeht.

Dieses kann zum Einen über Hinweise befriedigt werden, die ein potentieller Nutzer im direkten Kontakt durch Referenzpersonen im eigenen Umfeld erhält und denen üblicherweise eine hohe Glaubwürdigkeit beigemessen wird. Des Weiteren kann dem Bedürfnis Rechnung getragen werden, indem medial vermittelte Inhalte verarbeitet werden. Somit werden zwei Hypothesen postuliert:

H₁₂: Je stärker positive MBZS-bezogene interpersonelle Einflüsse von einer Person wahrgenommen werden, desto weniger nimmt sie die Nutzung von MBZS als riskant war.

H₁₃: Je stärker positive MBZS-bezogene medial vermittelte Einflüsse von einer Person wahrgenommen werden, desto weniger nimmt sie die Nutzung von MBZS als riskant war.

Darüber hinaus soll davon ausgegangen werden, dass ein Einfluss auf die Risikowahrnehmung auch von der Gestaltung der Leistungsmerkmale sowie durch die Zahlungssituationen, in denen ein MBZS genutzt werden kann, ausgeht. Dies führt zu:

H₁₄: Je mehr ein MBZS die von einer Person subjektiv als wichtig beurteilten Leistungsmerkmale aufweist, desto weniger nimmt sie die Nutzung von MBZS als riskant war.

H₁₅: Je mehr ein MBZS zur Zahlungsabwicklung in von einer Person als subjektiv wichtig beurteilten Zahlungssituationen geeignet ist, desto weniger nimmt sie die Nutzung von MBZS als riskant war.

Da es sich bei dem Risikokonstrukt nicht nur um ein produkt- sondern auch um ein personenspezifisches Phänomen handelt, hängt der Grad der Risikowahrnehmung auch von der individuellen Neigung zur Nutzung innovativer Mobilfunkdienste und Zahlungsverfahren. Dieser Befund spiegelt sich in folgender Hypothese wider:

H₁₆: Je stärker die MBZS-spezifische persönliche Innovationsneigung einer Person ausgeprägt ist, desto weniger nimmt sie die Nutzung von MBZS als riskant war.

4 Untersuchungsmethodik

Ziel der weiteren Ausführungen ist es, die oben postulierten Wirkungsbeziehungen empirisch zu überprüfen. Hierbei ist zu berücksichtigen, dass es sich bei den Konstrukten, auf welche die einzelnen Hypothesen Bezug nehmen, um sogenannte latente Variable handelt, die per Definition a priori nicht direkt beobacht- und messbar sind. Um dennoch eine quantitative Auswertung vornehmen zu können, sind die einzelnen Konstrukte mittels geeigneter Hilfsgrößen messbar zu machen [Hi08]. Diese Operationalisierung erfolgt in der vorliegenden Arbeit, soweit möglich, anhand von in der Literatur beschriebenen und validierten Indikatoren. Für die Konstrukte Einstellung, Nutzungsabsicht, subjektive Norm, interpersonelle Einflüsse, medial vermittelte Einflüsse und persönliche Innovationsneigung liegen für die eigene Erhebung geeignete, *reflektive* Indikatoren vor. Ebenso für die Konstrukte Vertrauen in Mobilfunknetzbetreiber, wahrgenommenes Risiko und Involvement. Da die drei letztgenannten Konstrukte m.E. jedoch verschiedene Dimensionen umfassen, wird *nicht* der im Schrifttum üblicherweise vorgeschlagenen

reflektiven Operationalisierung gefolgt, sondern einer *formativen* Messung der Vorzug gegeben.² Ebenfalls formativ operationalisiert werden die neu entwickelten Konstrukte Leistungsmerkmale, Zahlungssituation, und Aufwandsbereitschaft. Eine formative Operationalisierung besitzt an dieser Stelle den Vorteil, dass es damit möglich wird, von potentiellen MBZS-Anbietern konkret beeinflussbare Akzeptanztreiber zu identifizieren. Die Formulierungen der einzelnen Indikatoren können Tabelle 2 entnommen werden.

Die Erhebung der empirischen Daten erfolgte mittels einer Internetbefragung. Hierbei wurde ausdrücklich darauf hingewiesen, dass sich der Fragebogen auch an Personen richtet, die noch keine eigenen Erfahrungen mit MBZS gemacht haben. Am Ende des Befragungszeitraums lagen 311 verwertbare Datensätze als Ausgangspunkt für die weiteren Analysen vor. Zur Beschreibung der soziodemographischen Struktur der Teilnehmer wurden die vier Merkmale *Geschlecht*, *Alter*, *Wohnsitz* und *Bildungsabschluss* erhoben. Mit einem Anteil von 65,9% bestand in der Stichprobe ein deutliches Übergewicht von Personen männlichen Geschlechts. Bezüglich der Altersstruktur gehörte die überwiegende Mehrheit der Untersuchungsteilnehmer (92,6%) zu der aus Anbietersicht interessanten Altersgruppe bis 45 Jahre. Meistgenannter Wohnsitz war Deutschland (94,2%). Das Bildungsniveau der Teilnehmer kann als überdurchschnittlich hoch bezeichnet werden: 31,5% gaben das Abitur und 56,9% eine Universitäts- oder Fachhochschulausbildung als Abschluss an. Insgesamt ist festzustellen, dass die die Untersuchungsteilnehmer einen Personenkreis repräsentieren, der für MBZS-Anbieter eine wichtige Zielgruppe für die erfolgreiche Gestaltung der Markteinführung ihrer Lösungen darstellt, da sich primär aus ihm Innovatoren und frühe Übernehmer innovativer Systeme rekrutieren [KSS08, WGP08].

5 Untersuchungsergebnisse

Für die in diesem Abschnitt beschriebene strukturelle analytische Überprüfung wurde die Software *PLS-Graph 3.0* verwendet. Die in Tabelle 2 präsentierten Ergebnisse belegen, dass die reflektiven und formativen Messmodelle die zentralen Güteanforderungen an Partial Least Squares-Analysen [Hu07] erfüllen. Abbildung 1 ist zu entnehmen, dass — mit Ausnahme der durch die Hypothesen H_{12} , H_{13} und H_{15} postulierten Wirkungsbeziehungen — die Betragswerte der Pfadkoeffizienten des Strukturmodells durchgängig die kritische Untergrenze von 0,2 überschreiten. Alle über dieser Grenze liegenden Beziehungshypothesen sind darüber hinaus mit einer Irrtumswahrscheinlichkeit von höchstens 1% signifikant.

Eine Betrachtung des Strukturmodells hinsichtlich der R^2 -Bestimmtheitsmaße der latenten Variablen zeigt, dass das geforderte Mindestniveau von 0,3 ebenfalls fast durchweg erreicht wird. Einzige Ausnahme ist das Konstrukt Aufwandsbereitschaft.

² Für Beschreibungen der unterschiedlichen Prämissen und Vorgehensweisen einer reflektiven bzw. formativen Operationalisierung s. etwa [Hu07, Ha06].

Reflektiv operationalisierte Konstrukte/Indikatoren	Kennzahl (n = 311)				
	Ladung	t-Wert ^a	IR ^b	KR ^b	DEV ^b
Subjektive Norm:				0,88	0,71
Personen, die mir wichtig sind, halten es für eine gute Idee, MBZS zu nutzen.	0,91	72,31***	0,82		
Personen, die mein Verhalten beeinflussen, halten es für eine gute Idee, MBZS zu nutzen	0,86	36,68***	0,74		
Die Nutzung eines MBZS ist in meinem sozialen Umfeld ein positiv besetztes Zeichen für Fortschrittlichkeit.	0,74	22,89***	0,55		
Interpersonelle Einflüsse:				0,83	0,71
Die Mehrheit meiner Freunde und Arbeitskollegen beurteilt die Verwendung von MBZS positiv.	0,90	64,18***	0,81		
Einige meiner Freunde und Bekannte haben mir empfohlen, ein MBZS auszuprobieren.	0,78	21,95***	0,62		
Medial vermittelte Einflüsse:				0,88	0,78
In den Medien sind zahlreiche Beiträge zu finden, die positiv über MBZS berichten.	0,89	53,30***	0,80		
Medien und Werbung empfehlen durchweg die Nutzung von MBZS.	0,88	34,28***	0,77		
Innovationsneigung:				0,91	0,71
In meinem Freundeskreis bin ich gewöhnlich der Erste, der neue Mobilfunkdienste nutzt.	0,86	43,71***	0,74		
In meinem Freundeskreis bin ich gewöhnlich der Erste, der neue Zahlungsverfahren nutzt.	0,84	38,07***	0,70		
Meistens wenn ich von einem neuen Mobilfunkdienst höre, möchte ich diesen auch einmal ausprobieren.	0,87	60,72***	0,76		
Meistens wenn ich von einem neuen Zahlungsverfahren höre, möchte ich dieses auch einmal ausprobieren.	0,80	29,28***	0,63		
Einstellung:				0,96	0,89
Ein MBZS zu nutzen ist eine gute Idee.	0,95	141,38***	0,90		
Ein MBZS zu nutzen ist sinnvoll.	0,95	136,41***	0,90		
Ich finde Gefallen an der Vorstellung, ein MBZS zu nutzen.	0,93	98,03***	0,86		
Nutzungsabsicht:				0,93	0,83
Es ist sehr wahrscheinlich, dass ich ein MBZS verwenden werde.	0,91	56,70***	0,84		
Ich werde auf jeden Fall einmal ausprobieren, mit einem MBZS zu bezahlen.	0,91	95,23***	0,84		
Sobald die Möglichkeit dazu besteht, werde ich ein MBZS nutzen.	0,90	55,10***	0,81		
				Kennzahl (n = 311)	
Formativ operationalisierte Konstrukte/Indikatoren				Gewicht	t-Wert
Vertrauen: Mobilfunknetzbetreiber ...					
sind ehrlich.				0,36	2,06**
kümmern sich um ihre Kunden.				0,33	2,09**
kennen ihre Kunden.				0,25	1,68**
sind vertrauenswürdige Unternehmen.				0,43	2,44***
besitzen die zur Abwicklung von Zahlungen erforderlichen Kenntnisse und Fähigkeiten.				0,48	3,59***
Wahrgenommenes Risiko: Meiner Meinung nach ...					
ist die Nutzung eines MBZS mit einem höheren finanziellen Risiko verbunden als die Nutzung einer EC- oder Kreditkarte.				0,04	0,43 ^a
ist die Nutzung eines MBZS mit einem höheren Betrugsrisiko verbunden als die Nutzung einer EC- oder Kreditkarte.				0,24	2,36***
besteht das Risiko, dass es zu wenig Verkäufer gibt, bei denen ich mit MBZS bezahlen kann.				0,05	0,79 ^a
besteht das Risiko, dass MBZS aufgrund fehlender Netzabdeckung oder entladener Batterie des Endgerätes unzuverlässig funktionieren.				0,02	0,32 ^a
besteht das Risiko, dass bei der Nutzung von MBZS persönliche Informationen ohne mein Wissen mißbraucht werden könnten.				0,06	0,73 ^a
hätte die Nutzung eines MBZS negative Auswirkungen auf die Meinung meiner Freunde und Verwandten über mich.				0,07	0,88 ^a
passt die Nutzung eines MBZS nicht zum Bild, das ich von mir selbst habe.				0,37	4,10***

Tabelle 2: Indikatoren und Kennzahlen zur Operationalisierung der Modellkonstrukte

Formativ operationalisierte Konstrukte/Indikatoren	Kennzahl (n = 311)	
	Gewicht	t-Wert ^a
besteht das Risiko, dass unberechtigte Personen mit meinem Mobiltelefon bezahlen, falls ich es verliere oder es mir gestohlen wird.	0,03	0,30 ^{n.s.}
besteht das Risiko, dass aufgrund von mir selbst oder dem MBZS verursachten Übertragungsfehlern falsche Zahlungsdaten (z.B. Betragssumme, Kontonummer) verarbeitet werden.	0,06	0,74 ^{n.s.}
wäre die Nutzung eines MBZS mit einem Verlust an Bequemlichkeit verbunden, da ich viel Zeit mit dem Überprüfen von Transaktion oder der Korrektur von Zahlungsfehlern verbringen würde.	0,20	2,58***
ist die Nutzung eines MBZS insgesamt mit einem hohen Risiko verbunden.	0,35	3,24***
Involvement:		
MBZS interessieren mich nicht so sehr, weil ich mit den verfügbaren elektronischen und traditionellen Möglichkeiten der Bezahlung zufrieden bin. ^c	0,31	4,77***
MBZS finde ich nicht wichtig, weil ich keine Veranlassung sehe, zu einem neuen Zahlungssystem zu wechseln. ^c	0,20	2,70***
Meiner Meinung nach sind Mobiltelefone für die Abwicklung von Zahlungsvorgängen ungeeignet. ^c	0,39	5,82***
MBZS interessieren mich, weil ich die Möglichkeit mit Handys mehr als nur zu telefonieren, spannend finde.	0,40	6,21***
Leistungsmerkmale: Von einem MBZS erwarte ich, dass ...		
ich rund um die Uhr damit bezahlen kann.	0,23	5,26***
ich im Ausland damit bezahlen kann.	0,13	2,78***
es von vielen Verkäufern akzeptiert wird.	0,23	5,59***
ich es anstelle einer EC-Karte benutzen kann.	0,24	6,17***
ich es anstelle von Bargeld benutzen kann.	0,33	7,24***
ich es anstelle einer Kreditkarte benutzen kann.	0,17	3,66***
ich den Bezahlvorgang wesentlich schneller als mit EC- oder Kreditkarten durchführen kann.	0,04	0,64 ^{n.s.}
ich Zahlungen bis zu sechs Wochen nach der Transaktionseinleitung stornieren kann.	0,24	3,26***
ich mir jederzeit einen Überblick über die von mir getätigten MBZS-Zahlungen verschaffen kann.	0,00	0,03 ^{n.s.}
Zusatzleistungen angeboten werden, die mit der Zahlungstransaktion gekoppelt sind (z.B. Bonusprogramme).	0,07	1,10 ^{n.s.}
ich einen monatlichen Höchstbetrag angeben kann, den ich maximal ausgeben möchte.	0,12	1,75**
es einfach zu bedienen ist.	0,18	3,52***
Zahlungssituation: Von einem MBZS erwarte ich, dass ...		
ich an Automaten (z.B. Fahrtscheine, Zigaretten oder Getränke) bezahlen kann.	0,24	1,91**
ich mobil bezogene Inhalte (z.B. Klingeltöne, Musikdownloads, Parkgebühren) bezahlen kann.	0,16	1,47**
ich im Distanzhandel/Internet bezahlen kann.	0,07	0,63 ^{n.s.}
ich Zahlungen an Privatpersonen (z.B. zum anteiligen Begleichen einer Restaurantrechnung) tätigen kann.	0,04	0,33 ^{n.s.}
ich im stationären Präsenzhandel (z.B. Kaufhaus, Restaurant) bezahlen kann.	0,20	1,29*
ich im beweglichen Präsenzhandel (z.B. Taxi, Pizzaservice) bezahlen kann.	0,42	3,64***
ich Beträge bis 5 Euro bezahlen kann.	0,41	3,70***
ich Beträge zwischen 5 und 50 Euro bezahlen kann.	0,10	0,82 ^{n.s.}
ich Beträge über 50 Euro bezahlen kann.	0,19	1,56*
Aufwandsbereitschaft: Um ein MBZS nutzen zu können, bin ich prinzipiell bereit ...		
ein neues Mobiltelefon zu erwerben.	0,11	0,75 ^{n.s.}
zusätzliche Software auf mein Mobiltelefon zu laden.	0,34	1,87**
mich vor der Nutzung bei dessen Betreiber anzumelden.	0,51	3,21***
mit einer EC-Karte vergleichbare Nutzungskosten von ca. 5 Euro pro Jahr zu bezahlen.	0,45	2,76**
eine einmalige Anmeldegebühr in Höhe von 2 Euro zu bezahlen.	0,10	0,54 ^{n.s.}
zeitlich etwas länger dauernde Zahlungsvorgänge als bei einer Zahlung mit Bargeld oder Karte zu akzeptieren.	0,08	0,47 ^{n.s.}

- a) * = signifikant mit 10%iger Irrtumswahrscheinlichkeit, ** = signifikant mit 5%iger Irrtumswahrscheinlichkeit, *** = signifikant mit 1%iger Irrtumswahrscheinlichkeit, n.s. = nicht signifikant (Irrtumswahrscheinlichkeit für einseitigen t-Test).
- b) IR = Indikatorreliabilität, KR = Konstruktreliabilität, DEV = Durchschnittlich erfasste Varianz.
- c) Um eine einheitliche Ausrichtung der Indikatoren zu erzielen, wurden die Antwortstufen dieses Indikators vor der empirischen Auswertung rekodiert.

Tabelle 2: Indikatoren und Kennzahlen zur Operationalisierung der Modellkonstrukte (Forts.)

Insgesamt offenbaren die Berechnungen leichte Schwächen des Modells bei der Erklärung der Aufwandsbereitschaft durch die Nutzungsabsicht, wobei eine abschließende Bewertung allerdings insofern erschwert wird, als dass bislang keine Vergleichsstudien vorliegen, in denen diese Wirkungsbeziehung untersucht wurde. Angesichts der Neuartigkeit und Komplexität des Konstrukts sowie des hochsignifikanten Pfadkoeffizienten soll der niedrige R^2 -Wert jedoch nicht zur Ablehnung des Gesamtmodells führen. Bemerkenswert ist der mit 65% hohe Anteil der erklärten Varianz für die Nutzungsabsicht. Gängige Akzeptanzmodelle weisen für dieses Konstrukt in der Regel, bei einer teilweise deutlich höheren Zahl an Antezedenten, ein deutlich geringeres R^2 von typischerweise ungefähr 40% auf [Ve03]. Die Erklärungskraft des Modells kann somit als sehr gut bewertet werden.

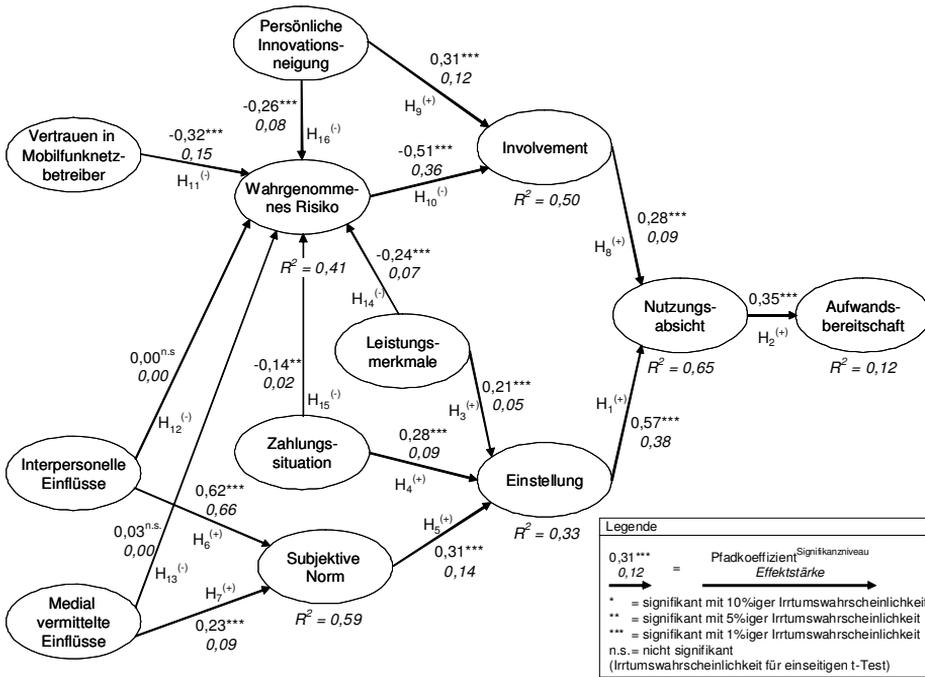


Abbildung 1: Pfadkoeffizienten, Effektstärken und Bestimmtheitsmaße (n = 311)

Die den Einfluss einer erklärenden latenten Variable auf die zu erklärende latente Variable widerspiegelnde Effektstärke ist am stärksten zwischen interpersonellen Einflüssen und subjektiver Norm ($f^2 = 0,66$), Einstellung und Nutzungsabsicht ($f^2 = 0,38$), wahrgenommenen Risiko und Involvement ($f^2 = 0,36$) sowie Vertrauen und wahrgenommenen Risiko ($f^2 = 0,15$). Die geringste Wirkung weisen die medial vermittelten und die interpersonellen Einflüsse auf das wahrgenommene Risiko ($f^2 = 0,00$) sowie die Zahlungssituation auf das wahrgenommene Risiko auf ($f^2 = 0,02$).

Zusammenfassend kann aus den Ergebnissen das folgende Resümee gezogen werden: Den höchsten Bestätigungsgrad weisen die Hypothesen zu einer Wirkung zwischen Einstellung und Nutzungsabsicht (H_1), interpersonellen Einflüssen und subjektiver Norm (H_6), Risiko und Involvement (H_{10}) sowie Vertrauen in Mobilfunknetzbetreiber und wahrgenommenem Risiko (H_{11}) auf. Die Hypothese H_2 eines Zusammenhangs zwischen Nutzungsabsicht und Aufwandsbereitschaft wird nur schwach bestätigt. Eine gemäß der Hypothesen H_3 und H_4 von Leistungsmerkmalen und Zahlungssituation ausgehende Wirkung auf die Einstellung wird mit einem mittleren Bestätigungsgrad nachgewiesen. Ebenso die postulierten Beziehungen zwischen subjektiver Norm und Einstellung (H_5), medial vermittelten Einflüssen und subjektiver Norm (H_7), Involvement und Nutzungsabsicht (H_8) sowie Innovationsneigung und Involvement (H_9). Die von den Leistungsmerkmalen gemäß H_{14} sowie der persönlichen Innovationsneigung gemäß H_{16} ausgehenden Wirkungen auf das wahrgenommene Risiko werden ebenfalls mit einem mittleren Bestätigungsgrad nachgewiesen. Der in der Hypothese H_{15} postulierte Einfluss der Zahlungssituation auf das wahrgenommene Risiko kann aufgrund des unter der kritischen Grenze von 0,2 liegenden Pfadkoeffizienten nicht bestätigt werden. Gleichermaßen nicht nachgewiesen werden können Effekte der interpersonellen (H_{12}) und medial vermittelten Einflüsse (H_{13}) auf das wahrgenommene Risiko.

6 Diskussion

Das wahrgenommene Risiko stellt eine zentrale Determinante für die MBZS-Akzeptanzbildung dar. Eine Betrachtung der Konstruktindikatoren zeigt, dass sich die Risikobeurteilung der Befragungsteilnehmer vor allem aus einer, auf ein sehr stark undifferenziertes Risikoempfinden hindeutende, unspezifischen Gesamtrisikowahrnehmung sowie einer psychologischen und einer Betrugsrisikokomponente zusammensetzt. Eine risikoreduzierende Wirkung geht in erster Linie vom Vertrauen in Mobilfunknetzbetreiber, der persönlichen Innovationsneigung des Endkunden sowie den Leistungsmerkmalen des MBZS aus.

Die deutliche Bestätigung des Wirkungspfades zwischen dem Vertrauen in Mobilfunknetzbetreiber und dem wahrgenommenen Risiko belegt die Bedeutung, welche dem Anbieter eines MBZS für die Risikowahrnehmung der Endkunden zukommen kann. Die Befragungsteilnehmer bewerteten Mobilfunknetzbetreiber als vertrauenswürdige Unternehmen, die auch die Kompetenz zur Abwicklung von Zahlungsvorgängen besitzen. Besteht zu einem Mobilfunknetzbetreiber, der einem Endkunden (z.B. aus einer bereits bestehenden Geschäftsbeziehung) bekannt ist, eine als positiv wahrgenommene Beziehung wird das mit dieser verbundene Vertrauen aller Voraussicht nach auf das MBZS übertragen [PI05].

Hinsichtlich der Leistungsmerkmale und Zahlungssituation präferieren die Befragungsteilnehmer MBZS, die als Ersatz von Bargeld und Karten für Bezahlungen kleinerer Beträge bis 5 Euro an Automaten und im stationären und beweglichen Präsenzhandel verwendet werden können.

Angesichts dieser Ergebnisse kann vor allem MBZS-Varianten, die auf Technologien zur drahtlosen Nahbereichskommunikation (wie etwa NFC) basieren, ein hohes Akzeptanzpotential zugeschrieben werden. Die in der Presse teilweise prognostizierten Erfolgchancen von MBZS-Zahlungen zwischen Privatpersonen sind für den deutschen Markt hingegen eher kritisch zu beurteilen. Darüber hinaus legen die Teilnehmer Wert auf eine hohe Zahl von Akzeptanzstellen und darauf, Zahlungen gegebenenfalls stornieren zu können.

Mit der Aufwandsbereitschaft wurde ein Konstrukt in die Modellbildung aufgenommen, aus dem sich, obgleich seines eher geringen R^2 -Wertes, Erkenntnisse darüber ableiten lassen, inwieweit die Nutzungsabsicht die Bereitschaft zur Übernahme monetärer und nicht-monetärer Kosten beeinflusst. Wie die Ergebnisse zeigen, sind die Befragungsteilnehmer prinzipiell dazu bereit, sich vor einer MBZS-Nutzung beim Betreiber anzumelden. Auffällig ist, dass ein Laden von zusätzlicher Software auf das Mobiltelefon nicht generell zurückgewiesen wird. Gering ausgeprägt sind hingegen die Bereitschaft zur Zahlung einer einmaligen Anmeldegebühr sowie die Akzeptanz von im Vergleich zu Bargeld- und Kartenzahlungen zeitlich länger dauernden Zahlungsvorgängen.

Bei Betrachtung der Determinanten der subjektiven Norm fällt die dominierende Wirkung von über interpersonelle Kanäle erhaltenen Informationen auf. Um bereits im Vorfeld einer tatsächlichen Nutzung den interpersonellen Informationsaustausch unter den Endkunden positiv zu beeinflussen und die Abgabe negativer Informationen zu verringern, können MBZS-Anbieter sowohl aufmerksamkeitssteigernde klassische Werbe- und Promotionsaktionen als auch das Internet („virales Marketing“) nutzen. Darüber hinaus besteht die Möglichkeit, vorhandene Endkunden zu kontaktieren, um diese zu motivieren, durch die Verbreitung positiver Informationen Neukunden aus ihrem persönlichen Umfeld zu werben [Ni07]. Ein wichtiger Grund für die vergleichsweise schwache Wirkung medial vermittelter Einflüsse dürfte darin liegen, dass MBZS-Anbieter massenmediale Kommunikationsinstrumente bislang kaum genutzt haben. Gleichwohl können diese Instrumente aber große Aufmerksamkeit generieren und wesentlich zur Akzeptanz beitragen. Im Rahmen der Gestaltung medial vermittelter Einflüsse ist auf eine realistische und stringente Kommunikation zu achten [PI05]. Hauptziel sollte es sein, breite Bevölkerungsschichten über das Bezahlen mit dem Handy zu informieren, um so positiven Einfluss auf die Einstellungsbildung zu nehmen. Zu empfehlen ist eine Darstellung von konkreten Anwendungsszenarien, in denen endkundengerechte Verwendungsmöglichkeiten von MBZS aufgezeigt werden. Im Mittelpunkt sollten dabei klar nutzwert orientierte Aussagen (z.B. „Bargeldersatz“) stehen.

7 Fazit

Den Ausgangspunkt des vorliegenden Beitrags bildete die Feststellung, dass für MBZS, speziell in Deutschland, eine Diskrepanz zwischen in der Vergangenheit prognostizierten Marktpotentialen und der derzeitigen Bedeutung in der unternehmerischen Praxis zu beobachten ist. Dementsprechend bestand die zentrale Zielsetzung darin, die MBZS-Akzeptanz theoretisch und empirisch zu untersuchen und darauf aufbauend Schlussfolgerungen für akzeptanzfördernde Gestaltungsmaßnahmen zu erarbeiten.

Das auf der Basis theoretischer und sachlogischer Überlegungen entwickelte Hypothesengerüst wurde empirisch weitgehend bestätigt. Trotz des Versuchs, das Phänomen der MBZS-Akzeptanz möglichst umfassend zu untersuchen, stellt jedes Modell gleichwohl immer nur ein vereinfachtes Abbild der Realität dar. Die gewonnenen Erkenntnisse unterliegen daher Restriktionen, die aber zugleich Anknüpfungspunkte für weitere Forschungsaktivitäten sind. So wäre etwa, mit Blick auf eventuelle nationale Unterschiede in den Zahlungsgewohnheiten, eine Untersuchung der Generalisierbarkeit der Befunde anhand einer Überprüfung des entwickelten MBZS-Akzeptanzmodells mit Datensätzen aus anderen Ländern eine mögliche Aufgabe für sich anschließende Arbeiten. Mit Blick auf die Modellkonzeption ist zu erwähnen, dass die in der Arbeit verwendeten Vertrauensindikatoren nur das Vertrauen in Mobilfunknetzbetreiber als potentielle Anbieter von MBZS erfassen. Nicht berücksichtigt wurde das Vertrauen der Endkunden in Leistungsanbieter, die Zahlungen mit MBZS akzeptieren. Um eine ganzheitliche Informationsbasis für die Ableitung von Maßnahmen zur Förderung der Ausbreitung von MBZS im Massenmarkt zu erhalten, sollten daher weiterführende Studien durchgeführt werden, in denen Einflussgrößen der MBZS-Akzeptanz von Leistungsanbietern im Mittelpunkt stehen. Weiterer Forschungsbedarf besteht in diesem Zusammenhang auch zur konkreten Ausgestaltung von Kooperationsstrategien im MBZS-Markt, insbesondere zu Fragen des Marken- und Erlösmanagements.

Zusammenfassend lässt sich feststellen, dass Bezahlen weder Selbstzweck ist noch ein Produkt, das sich gut verkaufen lässt. Technologische Veränderungen an sich interessieren die Endkunden üblicherweise nicht. Vielmehr müssen neue Produkte konkrete Verbesserungen oder nutzensteigernde Neuerungen anbieten. Ausschlaggebend für eine hohe Akzeptanz auf Seiten der Endkunden sind aber nicht nur an deren Anforderungen ausgerichtete Produktgestaltungsmerkmale sondern auch die Produkte und Dienstleistungen selbst, die per MBZS erworben bzw. in Anspruch genommen werden können. Wie die vorangegangenen Ausführungen gezeigt haben, besitzen MBZS aber insbesondere bei Präsenzzahlungen überzeugende Vorteile, die sie zu einer Ergänzung oder gar einem Ersatz bestehender Zahlungsmöglichkeiten machen können.

Literaturverzeichnis

- [Ca08] Carr, M. (2008): Adoption and diffusion of Internet banking. In: Ravi, V. (Hrsg.): *Advances in Banking Technology and Management*. New York: Hershey: 33-52.
- [Ch08] Chen, L.-D. (2008): A model of consumer acceptance of mobile payment. In: *International Journal of Mobile Communications*, 6: 32-52.
- [CP04] Cheong, J.H./Park, M.-C. (2004): Mobile payment adoption in Korea: URL: userpage.fu-berlin.de/~jmueller/its/conf/berlin04/Papers/Hwang_paper.pdf, Abruf am 01.02.2009.
- [ELP04] Eisenmann, M./Linck, K./Pousttchi, K. (2004): Nutzungsszenarien für mobile Bezahlfverfahren. In: Pousttchi, K./Turowski, K. (Hrsg.): *Mobile Economy*. Bonn: Gesellschaft für Informatik: 50-62.
- [Ge03] Gerpott, T.J. (2003): Diffusion mobilfunkbasierter Zahlungssysteme. In: Kruse, J. (Hrsg.): *Multi Media Mobil*. München: R. Fischer: 173-209.
- [Ha03] Harms, A.-K. (2003): Die Bedeutung von Inhibitoren im Adoptionsprozess technologiebasierter Self-Service-Innovationen. In: *Marketing ZFP*, 25: 257-272.

- [Hu07] Huber, F./Herrmann, A./Meyer, F./Vogel, J./Vollhardt, K. (2007): Kausalmodellierung mit Partial Least Squares. Wiesbaden: Gabler.
- [Hi08] Hildebrandt, L. (2008): Hypothesenbildung und empirische Überprüfung. In: Herrmann, A. et al. (Hrsg.): *Handbuch Marktforschung*. 3. Aufl., Wiesbaden: Gabler: 81-105.
- [KPW03] Khodawandi, D./Pousttchi, K./Wiedemann, D.G. (2003): Akzeptanz mobiler Bezahlverfahren in Deutschland. In: Pousttchi, K./Turowski, K. (Hrsg.): *Mobile Commerce*. Bonn: Gesellschaft für Informatik: 42-57.
- [KRK06] Koivumäki, T./Ristola, A./Kesti, M. (2006): Predicting consumer acceptance in mobile services. In: *International Journal of Mobile Communications*, 4: 418-435.
- [KSS06] Kristoffersen, S./Synstad, A./Sørli, K. (2006): What do users think of mobile payment? URL: heim.ifi.uio.no/%7Eesteinkri/Papers/isoneworld.pdf, Abruf am 01.02.2008.
- [KSS08] Kristoffersen, S./Synstad, A./Sørli, K. (2008): User perception of mobile payment. In: *International Journal of Knowledge Management Studies*, 2(1): 74-95.
- [LYY05] Lu, J./Yao, J.E./Yu, C.-S. (2005): Personal innovativeness, social influences and adoption of wireless Internet services via mobile technology. In: *Journal of Strategic Information Systems*, 14: 245-268.
- [Ma06] Mallat, N./Rossi, M./Tuunainen, V.K./Öörmi, A. (2006): The impact of use situation and mobility on the acceptance of mobile ticketing services. URL: csdl2.computer.org/comp/proceedings/hicss/2006/2507/02/250720042b.pdf, Abruf am 01.09.2006.
- [Ma07] Mallat, N. (2007): Exploring consumer adoption of mobile payments: A qualitative study. In: *Journal of Strategic Information Systems*, 16: 413-432.
- [Ni07] Nießing, D. (2007): Kunden-werben-Kunden-Kampagnen. Wiesbaden: DUV.
- [Pe05] Pedersen, P.E. (2005): Adoption of mobile Internet services: An exploratory study of mobile commerce early adopters. In: *Journal of Organizational Computing & Electronic Commerce*, 15: 203-222.
- [PI05] Pleil, T. (2005): Anmerkungen und strategische Ansätze zur Kommunikation von M-Payment. In: Hampe, F. et al. (Hrsg.): *Mobile Business*. Bonn: GI: 74-86.
- [PNT02] Pedersen, P.E./Nysveen, H./Thorbjørnsen, H. (2002): The adoption of mobile services: A cross service study. URL: ikt.hia.no/perep/rep3102.pdf, Abruf am 01.02. 2008.
- [PW05] Pousttchi, K./Wiedemann, D.G. (2005): Relativer Vorteil bei mobilen Bezahlverfahren: Mobiles Bezahlen aus dem Blickwinkel der Diffusionstheorie. In: Stucky, W./Schiefer, G. (Hrsg.): *Perspektiven des Mobile Business*. Wiesbaden: DUV: 35-50.
- [PW07] Pousttchi, K./Wiedemann, D.G. (2007): What influences consumers' intention to use mobile payments? URL: www.marshall.usc.edu/assets/025/7534.pdf, Abruf am 12.01.2009.
- [Ro03] Rogers, E.M. (2003): Diffusion of Innovations. 5. Aufl., New York: Free Press.
- [SCA07] Smart Card Alliance (2007): Proximity mobile payments. URL: www.smartcardalliance.org/download/pdf/Proximity_Mobile_Payments_200709.pdf, Abruf am 01.02.2008.
- [Ve03] Venkatesh, V./Morris, M.G./Davis, G.B./Davis, F.D. (2003): User acceptance of information technology: Toward a unified view. In: *MIS Quarterly*, 27: 425-478.
- [WGP08] Wiedemann, D.G./Goetze, L./Pousttchi, K. (2008): Ausgestaltung mobiler Bezahlverfahren: Ergebnisse der Studie MP3. In: Breitner, M.H. (Hrsg.): *Mobile und Ubiquitäre Informationssysteme*. Bonn: Gesellschaft für Informatik: 94-107.
- [WGP08] Wiedemann, D.G./Goetze, L./Pousttchi, K. (2008): Ausgestaltung mobiler Bezahlverfahren: Ergebnisse der Studie MP3. In: Breitner, M.H. (Hrsg.): *Mobile und Ubiquitäre Informationssysteme*. Bonn: Gesellschaft für Informatik: 94-107.

Kausalanalytische Untersuchung von Akzeptanzproblemen ambienter Technologien zur Vermeidung von Behandlungsfehlern in deutschen Krankenhäusern

Tyge-F. Kummer, Markus Bick

Wirtschaftsinformatik
ESCP-EAP Europäische Wirtschaftshochschule Berlin
Heubnerweg 6
14059 Berlin
tyge.kummer@escp-eap.de
markus.bick@escp-eap.de

Abstract: Die Besonderheit ambienter Technologien liegt vor allem in der Datenermittlung mittels Sensoren und mobiler Geräte, die in die Umgebung eingebettet sind und die gesammelten Informationen weiterleiten und auswerten. Die Einführung derartiger Technologien bringt häufig Akzeptanzprobleme mit sich. Ambiente Technologien werden von den Anwendern unter anderem wegen der damit verbundenen Überwachungsmöglichkeiten ihrer Tätigkeiten leicht als Eingriff in die Intimsphäre gewertet und in der Folge häufig abgelehnt. Im vorliegenden Beitrag werden die Wirkungszusammenhänge von Akzeptanzproblemen bezüglich ambienter Technologien in deutschen Krankenhäusern mittels empirischer Daten herausgearbeitet.

1 Einführung

Ambiente Systeme bieten zahlreiche Möglichkeiten zur Verbesserung der Effizienz und Effektivität von Arbeitsprozessen in Krankenhäusern [BKR08]. Durch den Einsatz von Sensoren bieten solche Systeme eine kontextspezifische Unterstützung aus dem Hintergrund. So kann ein ambientes System das Operationsteam in einem Krankenhaus bereits bei den Operationsvorbereitungen unterstützen, indem die entsprechenden OP-Siebe auf eine korrekte Bestückung und Vollständigkeit überprüft werden. Dies würde einen reibungslosen Ablauf gewährleisten, zumal nach erfolgter Operation festgestellt werden kann, ob Instrumente im Patienten vergessen wurden bzw. verblieben sind (vgl. z.B. [MMM06]). Die damit verbundenen vermeidbaren Fehler führen zu über 17.000 Todesfällen pro Jahr in Deutschland [Sa07]. Der Einsatz moderner Informations- und Kommunikationstechnologien kann das Risiko für den Patienten wesentlich reduzieren.

Dennoch führt der Einsatz von Ambient Intelligence auch zu erheblichen Herausforderungen. So können Sensoren, die eine Schlüsseltechnologie im Zusammenhang mit Ambient Intelligence darstellen, erhebliche Probleme mit sich bringen. Dies wird beispiels-

weise beim Einsatz von Radio Frequency Identification (RFID-) Tags im Einzelhandel deutlich. Die Kunden können zu der Auffassung gelangen, dass der Einsatz dieser Technologie einen potentiellen Angriff auf ihre Privatsphäre darstellt. Dies führt dann unter Umständen zur Ablehnung der Produkte oder des Anbieters bis hin zu Boykotten [Sp08]. Zudem sind Akzeptanzprobleme aufgrund der möglichen Überwachung und Kontrolle des Personals in seinem Arbeitsumfeld zu erwarten. Auch dies kann zu negativen Assoziationen bzw. Ängsten und Ablehnung führen. Hinsichtlich einer erfolgreichen Einführung und Etablierung von Ambient Intelligence steht somit die Akzeptanz derartiger Technologien im Vordergrund. Da die Akzeptanz eng mit der wahrgenommenen Nützlichkeit in Verbindung steht, lautet die Forschungsfrage des Beitrags: *Bestehen beim Einsatz ambienter Technologien zur Unterstützung von Medikationsprozessen Wirkungsbeziehungen zwischen den Ängsten und Befürchtungen sowie der wahrgenommenen Nützlichkeit und der Nutzungsintention?*

Dementsprechend werden im vorliegenden Beitrag potentielle Wirkungsbeziehungen zwischen Einflussfaktoren auf die Akzeptanz ambienter Technologien im Krankenhausumfeld herausgearbeitet. Dabei gilt es zu berücksichtigen, dass es aufgrund der besonderen Merkmale ambienter Systeme nicht möglich ist, die derzeit existierenden Akzeptanzmodelle zur Analyse zu nutzen [Sp08]. Als konkretes Anwendungsfeld wird dabei die ambiente Unterstützung von Medikationsprozessen betrachtet.

Nach einer Einführung zum derzeitigen Stand der Forschung sowie dem Untersuchungsgegenstand (Kap. 2) werden das Forschungsdesign sowie die Methodik zur Datenerhebung und Datenauswertung vorgestellt (Kap. 3). Darauf aufbauend werden die Ergebnisse unserer Untersuchung von Akzeptanzproblemen ambienter Technologien in deutschen Krankenhäusern analysiert und strukturiert (Kap. 4). Der Beitrag schließt mit einer Zusammenfassung sowie einem Ausblick auf weitere Forschungsfragen (Kap. 5).

2 Stand der Forschung

Ambiente Intelligence Systeme bzw. ambiente Systeme können in zahlreichen Anwendungsbereichen unterstützend eingesetzt werden. Im Rahmen einer Studie der EU-Kommission wurden verschiedene Bereiche identifiziert, die wesentliche Vorteile durch den verstärkten Einsatz von Sensoren aufweisen. Die hierbei genannten Anwendungsfelder sind die Logistik, das Gesundheitswesen, öffentliche Verwaltungen oder Produktions- und Fertigungsprozesse. Zudem werden der öffentliche Personennahverkehr oder Bibliotheken als Anwendungsgebiete angeführt [EU06].

Dabei werden ambiente Technologien häufig im Sinne der ursprünglichen Einführung des Begriffs Ambient Intelligence durch die European Union's Information Society Technologies Program Advisory Group (ISTAG) verstanden [IS03]: *Ambiente Technologien sind sämtliche Informations- und Kommunikationstechnologien, die selbstständig Informationen erfassen und verwerten können, um den Nutzer aus der Umgebung heraus aktiv zu unterstützen. Mobile Geräte und ein hoher Grad an Vernetzung der verwendeten Systeme und Technologien sind dabei Voraussetzung für den effizienten Einsatz ambienter Technologien.* Diesem Verständnis folgt auch der vorliegende Beitrag.

Im Gesundheitswesen werden komplexe Dienstleistungen erbracht, die sich in der Höhe des möglichen Unterstützungsgrades gegenüber den anderen Anwendungsfeldern deutlich abheben. Somit kommt der Forschung in diesem Bereich eine wesentliche Bedeutung zu. Für den Bereich Gesundheitswesen im Allgemeinen und den Einsatz ambienter Technologien im Krankenhausumfeld im Speziellen können drei Einsatzfelder ambienter Technologien identifiziert werden [BKR08]: *medizinische Geräte*, *Telemedizin* sowie eine *generelle Prozessunterstützung*. Hinsichtlich der Verbesserung der Effizienz und Effektivität von Arbeitsabläufen in Krankenhäusern und einer damit verbundenen Steigerung der Behandlungsqualität bei gleichzeitiger Reduzierung von Fehlbehandlungen rückt die generelle Prozessunterstützung in den Vordergrund. Mögliche Einsatzgebiete ambienter Technologien in diesem Einsatzfeld sind die *Patientenidentifikation*, das *Patientenmonitoring*, die *Authentifikation*, die *Adaptive Logistik* sowie die *Lokalisierung von Objekten und Personen* (vgl. z. B. [Ba03; BBA05; BBF06; Ri03; Ro05; Mo06]).

Trotz dieser vielfältigen Anwendungsbeispiele bleibt die Frage der Akzeptanz ambienter Technologien im Krankenhausumfeld und daraus resultierender Probleme weitestgehend unbeantwortet. Derzeit liegen keine umfassenden Untersuchungen der Akzeptanz ambienter Technologien im Gesundheitswesen vor. Einzig die Akzeptanz mobiler Technologien wurde betrachtet [Ra07; Wu06]. Dabei wurde die Benutzerakzeptanz neuer Technologien im Gesundheitswesen häufig unter Rückgriff auf das Technology Acceptance Model (TAM) [Da89] bzw. TAM2 [VD00], die Theory of Planned Behavior (TBP) [Aj91] oder die Innovation Diffusion Theory (IDT) [AP98] erhoben. Im Vordergrund steht die Analyse der Entscheidung für oder gegen die Nutzung einer neuen Technologie, vor allem auf der Basis von Einschätzungen zur Nützlichkeit der neuen Technologien und deren Bedienungsfreundlichkeit.

Diese beiden Faktoren – wahrgenommene Nützlichkeit und Bedienungsfreundlichkeit – sind auch für ambiente Technologien von großer Bedeutung. Sie reichen allerdings nicht aus, um die Ablehnung oder Annäherung an diese Technologien zu erklären. Dies liegt an den spezifischen Charakteristika ambienter Systeme, die für den Nutzer mehr oder minder unsichtbar Daten erheben und weiterverarbeiten. Bei ambienten Systemen sind somit Fragen hinsichtlich des Datenschutzes, des Schutzes der Privatsphäre und der eigenen Entscheidungsfreiheit von großer Bedeutung [Sp08].

3 Forschungsdesign und Methodik

Die mit dem Einsatz ambienter Technologien verbundenen impliziten Überwachungspotentiale können mithilfe klassischer Akzeptanzmodelle nicht adäquat abgebildet werden (Kap. 2). Vor diesem Hintergrund untersuchen wir, welche Akzeptanzprobleme bezüglich ambienter Technologien im Krankenhaus von besonderer Relevanz sind. Ein erster Schritt hierzu ist herauszuarbeiten, welche Akzeptanzprobleme existieren, bevor ambiente Technologien implementiert werden, und ob diese Probleme einen Einfluss auf die Absicht zur Nutzung dieser Technologie haben.

Der Untersuchung liegt ein Mixed-Methods-Ansatz zugrunde. Dabei wurden zunächst verschiedene Akzeptanzprobleme entlang von Experteninterviews identifiziert [BKR08].

Dieser qualitative Zugang führte zu einem tieferen Verständnis des Einsatzes ambierter Technologie durch das medizinische Personal deutscher Krankenhäuser. Im Sinne eines Mixed-Methods-Ansatzes (vgl. [TT02]) bilden die geführten Interviews die Grundlage der quantitativen Erhebung. Die abgeleiteten Forschungshypothesen (Tabelle 1) werden im Folgenden mithilfe eines Fragebogens getestet.

H1	Ängste vor neuen Technologien haben einen signifikanten Einfluss auf Überwachungsängste.
H2	Ängste vor neuen Technologien haben einen signifikanten Einfluss auf die Angst vor einer Verschlechterung der Arbeitsbedingungen auf.
H3	Ängste vor neuen Technologien haben einen signifikanten Einfluss auf die wahrgenommene Nützlichkeit.
H4	Ängste im Zusammenhang mit neuen Technologien haben einen signifikanten Einfluss auf die Nutzungsintention.
H5	Überwachungsängste haben einen signifikanten Einfluss auf die Angst vor veränderten Arbeitsbedingungen.
H6	Überwachungsängste haben einen signifikanten Einfluss auf die wahrgenommene Nützlichkeit auf.
H7	Überwachungsängste haben einen signifikanten Einfluss auf die Nutzungsintention.
H8	Ängste vor einer Verschlechterung der Arbeitsbedingungen haben einen signifikant Einfluss auf die wahrgenommene Nützlichkeit.
H9	Ängste vor veränderten Arbeitsbedingungen haben einen signifikant Einfluss auf die Nutzungsintention.
H10	Die wahrgenommene Nützlichkeit hat einen signifikanten Einfluss auf die Nutzungsintention.

Tabelle 1: Übersicht der aufgestellten Forschungshypothesen

Im Vordergrund standen drei Fragenbereiche, die zu einer Ablehnung ambierter Technologien führen können, sowie zwei klassische Fragenbereiche der Akzeptanzforschung.

- *Ängste vor neuen Technologien:* Mit diesem Konstrukt wurden jene Ängste fokussiert, die aus dem zunehmenden Einsatz von Technologien resultieren. Dies sind zum einen die generelle Ablehnung bzw. ein grundsätzliches Vertrauen in Technologien. Zum anderen sind es aber auch Befürchtungen, dass ambiante Technologien die persönliche Komponente der medizinischen Behandlung negativ beeinflussen.
- *Ängste vor Verschlechterung der Arbeitssituation:* Mithilfe dieses Konstrukts wurden mögliche Veränderungen innerhalb des Arbeitsumfelds bzw. der Arbeitsprozesse hinterfragt. Dies ist vor allem die Angst vor einem Arbeitsplatzverlust, aber auch die Befürchtung zusätzlicher Arbeitsbelastungen. Zudem wurden Einschätzungen hinsichtlich möglicher Einschränkungen, wie beispielsweise individueller Entscheidungsfreiräume, getestet.
- *Überwachungsängste:* Hiermit wurde insbesondere auf die zuvor erwähnte Besonderheit ambierter Technologien eingegangen, die seitens der Anwender als Eingriff in die Privatsphäre gewertet werden können (Kap. 2).
- *Wahrgenommene Nützlichkeit:* Dieses Konstrukt gibt an, inwieweit das System als nützlich und für die Berufsausübung relevant angesehen wird.

- *Nutzungsintention*: Die Nutzungsintention gibt an, inwieweit die Befragten bereit sind, das System tatsächlich freiwillig zu nutzen.

Die beiden letzten Fragenbereiche wurden innerhalb unserer Erhebung gegenüber deren Messung mithilfe klassischer Akzeptanzmodelle, wie TAM2 (Kap. 2), abgeändert. Dabei wurde die *Wahrgenommene Nützlichkeit* mit dem Konstrukt der *Relevanz* aus dem TAM2 zusammengefasst. Zudem wurde die *Nutzungsintention* um das Konstrukt der *Freiwilligkeit* ergänzt. Dies war im ursprünglich geplanten Design nicht vorgesehen. Allerdings erforderte eine deutlich zu geringe Diskriminanzvalidität (Kap. 4.1) aufgrund der Korrelation zwischen diesen Konstrukten eine entsprechende Anpassung der Konstrukte.

Als Konkretisierung des Gegenstandsbereich Fehlbearbeitungen wurde die Vermeidung von Medikationsfehlern ausgewählt, da dieses Anwendungsszenario zahlreiche Aspekte ambienter Prozessunterstützung wie beispielsweise die Identifikation von Menschen und Objekten vereint und somit auch Ängste im Zusammenhang mit diesen Technologien impliziert [BKR08]. Zudem handelt es sich um einen zentralen Behandlungsprozess in Krankenhäusern. Als Zielpersonen für die Datenerhebung wurden Pflegekräfte ausgewählt, da diese oftmals für die Medikation in Krankenhäusern zuständig sind und eine homogenere Gruppe darstellen als beispielsweise Patienten und Ärzten verschiedener Fachrichtungen. Gegenstand des Fragebogens war somit ein Szenario, in dem der Einsatz ambienter Technologien in einer für Pflegekräfte typischen Arbeitssituation beschrieben wird: *Mithilfe eines mobilen Gerätes und Sensoren wird dieser Prozess derart unterstützt, dass zunächst geprüft wird, ob das richtige Medikament gewählt wurde. Zudem wird festgestellt, ob die Pflegekraft für diese Tätigkeit berechtigt ist bzw. ob sie ihr aufgetragen wurde. Abschließend wird die Medikation durch die Pflegekraft bestätigt. Die entsprechenden Daten werden zu Dokumentationszwecken gespeichert, sodass weitere Schreibarbeiten entfallen. Sollten die Faktoren Patient, Medikament und Pflegekraft nicht übereinstimmen erfolgt ein Alarm, der ebenfalls gespeichert wird.*

Sämtliche Frageitems sind in Tabelle 2 abgebildet und wurden anhand einer siebenstufigen Likert-Skala gemessen. Die Items wurden innerhalb eines Pretests geprüft. Dabei wurde der Fragebogen von sieben Mitarbeitern des medizinischen Personals eines Krankenhauses ausgefüllt und anschließend mit diesen diskutiert. Darauf aufbauend wurden kleinere Änderungen vorgenommen, um die Verständlichkeit und somit letztendlich die Validität zu erhöhen.

An der tatsächlichen Befragung nahmen insgesamt 113 Pflegekräfte teil, wobei 112 Fragebögen ausgewertet werden konnten. Bei den Teilnehmerinnen und Teilnehmern handelt es sich um auszubildende Pflegekräfte von vier verschiedenen deutschen Ausbildungsstätten, die sich mindestens im zweiten Lehrjahr befinden und über praktische Erfahrung im Krankenhaus verfügen. Somit wurde sichergestellt, dass die Befragten mit den zentralen Arbeitsabläufen eines Krankenhauses vertraut sind. Diese Form des Feldzuganges schränkt die Generalisierbarkeit der Ergebnisse im Vergleich zu einer Erhebung bei examinierten Pflegekräften ein, da diese über eine umfassendere Berufserfahrung verfügen. Das durchschnittliche Alter der Teilnehmerinnen und Teilnehmer ist 21,80 Jahre. Der Anteil befragter Frauen beträgt 81,25 % (18,75 % Männer). Dieses

Verhältnis spiegelt die Realität in Pflegeberufen wider, weshalb geschlechtsspezifische Verzerrungen in der externen Validität der Ergebnisse auszuschließen sind.

Die Hypothesen wurden mithilfe der Partial Least Squares (PLS)-Methode getestet, welche den sogenannten Strukturgleichungsmodellen zugerechnet wird. Der Einsatz dieses Verfahrens erlaubt die empirische Auswertung entlang eines *Messmodells* und eines *Strukturmodells*. Das *Strukturmodell* besteht aus einem Netzwerk, welches die endogenen und exogenen Variablen in Beziehung zueinander setzt, wohingegen das *Messmodell* die Konstrukte mit einer Reihe von Indikatoren verbindet [Wo74; Wo82]. Im Gegensatz zu klassischen statistischen Verfahren, wie der Faktorenanalyse oder der Regressionsanalyse, bewertet PLS Mess- und Strukturmodell gemeinsam. Dabei wird ein iterativer Algorithmus verwendet, der einerseits die Faktorladungen bezüglich der Konstrukte schätzt und andererseits die kausalen Beziehungen zwischen den Konstrukten bewertet [FL81]. Ein Vorteil von PLS ist, dass keine normalverteilten Werte vorausgesetzt werden und die Methode bei vergleichsweise kleinen Stichproben eingesetzt werden kann. Darüber hinaus kann PLS auch zum Testen von Theorien herangezogen werden, die sich noch in einem frühen Entwicklungsstadium befinden [Ha98]. Dabei ist auch PLS nicht frei von Kritik. So wird insbesondere der frühe Einsatz von PLS bei zu geringen Stichprobengrößen problematisiert [MS06]. Aus diesem Grund sind ergänzende Konstrukte, wie die Teststärke heranzuziehen, um die Güte der Analyse beurteilen zu können (Kap. 4.2).

4 Ergebnisse

Für die im Folgenden beschriebene PLS-Pfadanalyse wurde die Software SmartPLS 2 [RWW05] verwendet. Bei der Verwendung von PLS empfiehlt es sich, bei der Ergebnisbetrachtung Messmodell (Kap. 4.1) und Strukturmodell (Kap. 4.2) zu unterscheiden. Zusätzlich zu den Ergebnissen ist jeweils deren Güte zu überprüfen, um somit Rückschlüsse auf die Qualität der Untersuchung zu erzielen. Eine Interpretation der Ergebnisse erfolgt in Kap. 4.3.

4.1 Messmodell

Das der Untersuchung zugrundeliegende Messmodell ist ausschließlich reflektiv, da die jeweiligen Angstkategorien durch die verschiedenen Frageitems abgebildet werden. Sämtliche Pfadbeziehungen im Messmodell verlaufen daher von den latenten zu den manifesten Variablen (Tabelle 2). So wird beispielsweise die latente Variable *Ängste vor neuen Technologien* durch vier manifeste Variablen (die vier konkreten Frageitems) erklärt. Die Zuordnung der negativen Assoziationen als Frageitems zu den latenten Variablen erfolgte in einem Pretest. Die Güte eines reflektiven Messmodells ergibt sich aus der Konvergenz- und der Diskriminanzvalidität [Ha98], wobei sich erstere aus der Indikator- und Konstruktreliabilität sowie der durchschnittlich erfassten Varianz zusammensetzt [FL81].

	Arithmetisches Mittel	Standard Abweichung	Faktorladung
Ängste vor neuen Technologien			
Man sollte sich besser auf Menschen verlassen als auf derartige Systeme.	3.321	1.590	0.743
Ich sehe es als problematisch an derartigen Systemen zu vertrauen.	3.532	1.686	0.790
Ein derartiges System steht im Widerspruch zu ethischen Werten.	3.866	1.901	0.735
Ich würde ungern mit noch mehr technischen Geräten zu tun haben.	4.277	1.955	0.821
Ängste vor Verschlechterung der Arbeitssituation			
Ich befürchte, dass mich derartige Systeme auf Dauer ersetzen könnten.	3.712	2.124	0.606
Die Nutzung des Systems wird für mich negative Konsequenzen haben.	4.718	1.612	0.722
Ein derartiges System wird zu Überstunden führen.	4.469	1.659	0.772
Das System wird eine zusätzliche Belastung darstellen.	3.864	1.732	0.812
Überwachungsängste			
Die Vorstellung, mich der Überwachung durch das System nicht entziehen zu können, bereitet mir Angst.	4.330	1.862	0.844
Es ist mir unangenehm, wenn ich nicht weiß, was aufgezeichnet wird.	3.232	1.831	0.760
Durch ein derartiges System würde ich fürchten, bloßgestellt zu werden.	4.402	2.016	0.801
Es ist mir unangenehm, dass ich die über mich gespeicherten Daten nicht verändern kann.	4.153	1.942	0.767
Wahrgenommene Nützlichkeit			
Die Nutzung eines derartigen Systems würde meine Arbeitsleistung steigern.	4.376	1.728	0.872
Für meine Arbeit ist die Verwendung eines derartigen Systems von hoher Relevanz.	4.643	1.565	0.829
Die Nutzung eines derartigen Systems würde mir helfen, meine Arbeit besser zu erledigen.	4.469	1.825	0.891
Durch die Nutzung des Systems könnte das Krankenhaus produktiver werden.	3.946	1.825	0.691
Nutzungsintention			
Vorausgesetzt das System würde wie beschrieben funktionieren, würde ich es auch nutzen.	2.847	1.861	0.900
Ein derartiges System würde ich auch freiwillig verwenden.	3.634	1.870	0.881
Es wäre nicht notwendig, dass Vorschriften das System anordnen, damit ich es nutze.	3.846	1.846	0.591
Wenn das mobile Endgerät klein ist, würde ich es gerne mit mir führen.	3.268	2.127	0.866

Tabelle 2: Deskriptive Ergebnisse und Faktorladungen

Die *Indikatorreliabilität* ergibt sich aus der Korrelation zwischen den Indikatoren und dem Konstrukt und wird über die Faktorladungen ausgewiesen. Dabei sollten die Faktorladungen den Wert 0,7 übersteigen. Faktorladungen unter 0,5 gelten als nicht akzeptabel und sind von der Analyse auszuschließen [Ch98]. Die *Konstruktrelabilität* gibt an, inwieweit eine latente Variable durch die Indikatoren beschrieben wird. Hierbei sollte ein Wert von mindestens 0,6 erreicht werden. Zur Bestimmung der internen Konsistenz kann zudem *Cronbachs Alpha* herangezogen werden. Als Grenzwert für akzeptable Ergebnisse gilt hierbei ein Wert größer 0,7. Allerdings wird die Verwendung von Cronbachs Alpha bei PLS aufgrund der nicht zwangsweise vorhandenen τ -Äquivalenz der Indikatoren mitunter kritisiert [Ch98]. In der vorliegenden Untersuchung wurden beide Erfordernisse zur Bestimmung der internen Konsistenz erfüllt (Tabelle 2).

Mit Hilfe der durchschnittlich erfassten Varianz (*DEV*) wird ermittelt, in welchem Ausmaß die latente Variable den erklärten Varianzanteil in Relation zum Messfehler wiedergibt. Dabei sollte die DEV größer 0,5 sein. Um zu überprüfen, ob eine ausreichende Diskriminanzvalidität vorliegt, schlagen [FL81] zudem vor, dass die DEV stets größer als die größte quadrierte Konstruktkorrelation der latenten Variablen sein sollte. Wie in Tabelle 3 dargestellt, werden diese Anforderungen erfüllt.

	Cronbachs Alpha	Konstruktreliabilität	DEV	max. quadrierte Korrelation
Ängste vor neuen Technologien	0,776	0,856	0,598	0,450
Ängste vor Verschlechterung der Arbeitssituation	0,714	0,821	0,536	0,442
Überwachungsängste	0,808	0,872	0,630	0,211
Wahrgenommene Nützlichkeit	0,840	0,894	0,680	0,500
Nutzungsintention	0,831	0,889	0,672	0,500

Tabelle 3: Validitätsergebnisse des Messmodells

4.2 Strukturmodell

Varianzbasierte Ansätze wie PLS tendieren dazu, bei einer zu geringen Stichprobengröße die Faktorenladungen im Messmodell generell zu hoch und die Pfadkoeffizienten im Strukturmodell zu niedrig zu schätzen. Zur Bestimmung der erforderlichen Stichprobengröße sollte der zehnfache Wert der größten Anzahl von Indikatoren der komplexesten latenten Variable oder der höchsten Anzahl von Pfaden zu einer endogenen Variablen herangezogen werden [Ch98]. Im vorliegenden Beitrag sind bei maximal vier Indikatoren und maximal fünf Pfaden zu einer endogenen Variablen 40 bzw. 50 Fragebögen erforderlich. Da stets die strengere der beiden Regeln Anwendung findet, setzt das Modell eine Stichprobengröße von mindestens 50 ausgewerteten Fragebögen voraus.

Zusätzlich muss die Teststärke (statistical power) zur Bestimmung der erforderlichen Stichprobengröße beachtet werden, da sonst möglicherweise signifikante Pfadkoeffizienten als nicht signifikant geschätzt werden [Co88]. Hierbei wird von einem Wert von mindestens 0,8 ausgegangen [Co88, Ha98]. Unter Verwendung von G*Power 3.0 [Fa07], einer moderaten Effektstärke von 0,15, einem α von 0,05 sowie vier exogenen

latenten Variablen erfordert das hier untersuchte Modell für eine Teststärke von 0,8 eine Stichprobengröße von mindestens 85. Da die tatsächliche Stichprobengröße 112 beträgt (Teststärke 0,913), ist diese Anforderung erfüllt.

Nach der Schätzung der Pfadkoeffizienten in dem Strukturmodell wurde das Bootstrapping-Verfahren eingesetzt, um die korrespondierenden t-Werte zu ermitteln. Die Hypothesen H1 bis H10 entsprechen dabei den Pfaden im Strukturmodell. Jede Hypothese wird durch einen Pfadkoeffizienten determiniert, der positive und negative Werte annehmen kann, sowie durch einen t-Wert. Mittels des t-Tests können die verschiedenen Signifikanzniveaus 0,05, 0,01 und 0,001 der Ergebnisse abgeleitet werden, die dann zur Annahme oder Ablehnung der Forschungshypothesen führen [ET93]. In Abbildung 1 sind sämtliche Pfadkoeffizienten und Signifikanzniveaus dargestellt.

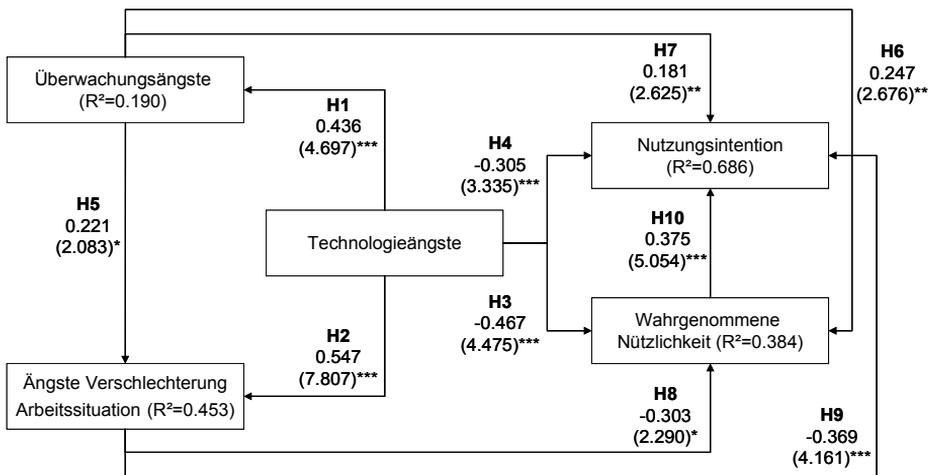


Abbildung 1: Ergebnisse des Strukturmodells
 (***) = $p < 0.001$, (**) = $p < 0.01$, (*) = $p < 0.05$)

Ebenso wie bei der Regressionsanalyse gibt das Bestimmtheitsmaß (R^2) die Erklärungskraft einer latent endogenen Variablen an. Hierbei beurteilt [Ch98] die Werte 0,67, 0,33 und 0,19 als substantiell, adäquat und gering. Diese Werte werden zur Evaluierung des Ergebnisses herangezogen (Abbildung 1). Es zeigt sich, dass das Bestimmtheitsmaß der Nutzungsintention substantiell ist. Die wahrgenommene Nützlichkeit und die Ängste in Verbindung mit einer Verschlechterung der Arbeit sind als adäquat zu beurteilen. Die Überwachungsangst kann zumindest in einem niedrigen Maß durch die Ängste vor neuen Technologien erklärt werden.

Um Aussagen zur Prognoserelevanz (Q^2) treffen zu können, wurde zudem das Stone-Grosser-Kriterium in Form einer Blindfolding-Prozedur angewendet [Te05]. Ein Wert größer 0 gibt an, dass das Modell Prognoserelevanz aufweist. In dem untersuchten Modell sind die Q^2 -Werte für sämtliche Konstrukte größer 0.

4.3 Interpretation der Ergebnisse

Wie aus den Ergebnissen hervorgeht, weisen sämtliche Konstrukte einen signifikanten Einfluss auf. Alle getesteten Hypothesen (Tabelle 1) wurden bestätigt. Durch die Korrelationen zwischen den verschiedenen Ängsten wird deutlich, dass die Befragten ambiante Technologien auf verschiedenen Ebenen ablehnen. Sofern diese Abneigung gegenüber der Technologie auftritt, spiegelt sich diese Ablehnung in den verschiedenen Ängsten wider. Daher scheint es eine generelle Ablehnung hinsichtlich der untersuchten Technologie zu geben, die die Akzeptanz erheblich negativ beeinflusst. Die *Ängste in Zusammenhang mit neuen Technologien* wurden in dem vorgestellten Modell als Ausgangspunkt gewählt. Personen, die Technologien generell ablehnen und diese konkret als ethisch verwerflich empfinden, werden eher andere Ängste aufweisen als Personen, die derartigen Technologien offen gegenüberstehen und keine ethischen Bedenken aufzeigen. Die *Überwachungsängste* verstärken wiederum signifikant die Ängste vor einer Verschlechterung der Arbeitsbedingungen (Abbildung 1). Dies erscheint ebenfalls nachvollziehbar, da, sofern die permanente Überwachung am Arbeitsplatz als störend empfunden wird, dies für den Befragten auch eine Verschlechterung der Arbeitsbedingungen darstellt. Allerdings ist dieser Einfluss deutlich schwächer als der Einfluss der Ängste in Verbindung mit neuen Technologien.

Bei den Einflüssen auf die *wahrgenommene Nützlichkeit* und die *Nutzungsintention* sind insbesondere die Vorzeichen der Pfadkoeffizienten von Interesse (Abbildung 1). Die *Ängste im Zusammenhang mit neuen Technologien* sowie die *Ängste vor einer Verschlechterung der Arbeitssituation* weisen einen signifikant negativen Einfluss auf beide Konstrukte auf. Die Ablehnung in Form von Ängsten schlägt folglich auf die Nutzungseinschätzung durch und das konkrete ambiante System wird als nicht sinnvoll bzw. nicht relevant erachtet. Ebenso weisen beide Konstrukte einen signifikant negativen Einfluss auf die Nutzungsintention auf. Dies ist vor dem Hintergrund einer Implementierung des Systems von erheblicher Bedeutung, da – sofern diese Ängste nicht berücksichtigt werden – die Nutzungsintention verringert wird, wenn diesen Ängsten nicht entgegenge wirkt wird. Umgekehrt kann allerdings bei einer aktiven Reduzierung der Ängste die spätere Nutzungsintention erhöht werden. Dieser Umstand wird durch den indirekten Einfluss der Ängste über die wahrgenommene Nützlichkeit auf die Nutzungsintention weiter verstärkt.

Überraschend ist jedoch der positive Einfluss der *Überwachungsängste* auf den wahrgenommenen Nutzen und die Nutzungsintention, da hier aufgrund der vorgelagerten qualitativen Datenerhebung ebenfalls von einer negativen Auswirkung ausgegangen wurde [BKR08]. Allerdings scheinen hinsichtlich dieser Angst bei den Befragten die Vorteile des Einsatzes ambienter Technologien zu überwiegen. Die Transparenz der Handlungen und die damit verbundene Bloßstellung werden dem Ziel der Verbesserung der Medikation und somit einer Steigerung der Behandlungsqualität untergeordnet. Auch dieses Ergebnis hat erhebliche Implikationen auf die Einführung ambienter Technologien, da in Hinblick auf diese Ängste deutlich weniger Sensibilität erforderlich ist als dies beispielsweise Publikationen aus dem Endkundenbereich nahelegen (vgl. [Sp08]). Überwachung stellt im Krankenhausumfeld somit keine Ursache für eine generelle Ablehnung ambienter Technologien in diesem Anwendungsbereich dar.

Insgesamt konnte das Modell 68,6 % der Varianz für die *Nutzungsintention* erklären. Auch bei Bestimmung des korrigierten Bestimmtheitsmaßes beträgt dieser Wert noch 67,44 % und ist daher vom Erklärungsgehalt mit anderen Modellen im medizinischen Umfeld vergleichbar, wie z. B. [WWL07]. Die negativen Assoziationen im Zusammenhang mit Ambient Intelligence beeinflussen die Nutzungsintention erheblich. Die wahrgenommene Nützlichkeit zeigt einen ähnlich starken Einfluss auf die Nutzungsintention, wie die Ängste vor einer Verschlechterung der Arbeitssituation (Abbildung 1). Aufgrund dieser Ergebnisse erscheint es sinnvoll die Einführung und Etablierung derartiger Technologien anhand der Ergebnisse anzupassen und die negativen Assoziationen explizit zu berücksichtigen. Zudem sollten Akzeptanzmodelle im Zusammenhang mit Ambient Intelligence stets die Assoziationen der Nutzer berücksichtigen.

5 Zusammenfassung und Ausblick

Ziel des Beitrags war es zum einen, zu untersuchen, ob ambiante Technologien aufgrund ihrer besonderen Charakteristika speziell angepasste Akzeptanzmodelle erfordern. Zum anderen wurden die konkreten Wirkungsweisen der Ängste im Zusammenhang eines konkreten Anwendungsszenarios näher untersucht. Es wurde deutlich, dass die besonderen Assoziationen im Zusammenhang mit Ambient Intelligence einen erheblichen Einfluss auf die Nutzungsentscheidung ausüben, der in den bisherigen Akzeptanzmodellen nicht ausreichend berücksichtigt wurde. Derartige Assoziationen bzw. Ängste spielen bei ambienten Technologien eine entscheidende Rolle, die bei einer nachhaltigen Einführung zu beachten sind. Diese Ängste können negative Auswirkungen haben, wie dies bei *Ängsten in Zusammenhang mit neuen Technologien* und bei *Ängsten vor einer Verschlechterung der Arbeitsbedingungen* der Fall ist. Zudem können Überwachungsängste, sofern diese gerechtfertigt sind, auch positive Auswirkungen auf eine freiwillige Nutzung haben. Als Einschränkung muss allerdings betont werden, dass diese Ergebnisse auf Basis einer Befragung von Auszubildenden gewonnen wurden, die vergleichsweise jung sind und eine andere Einstellung gegenüber modernen Technologien aufweisen könnten als ältere Personen mit mehr Erfahrung. Aufgrund der fortschreitenden Technologisierung wäre daher bei jüngeren Befragten eine geringere Ablehnung von Technologien zu erwarten.

Um die Akzeptanzprobleme ambienter Technologien im Krankenhausumfeld detaillierter zu untersuchen, sollten in einem nächsten Schritt einzelne Einführungsprojekte in Form von Fallstudien in diesem Bereich analysiert werden. Damit verbunden ist die Frage, inwieweit die Ausprägungen und Wirkungsweisen in Zusammenhang mit ambienten Technologien in diesem Umfeld kulturabhängig sind. Daher sollte die Forschung auf weitere und vor allem unterschiedliche Kulturkreise ausgedehnt werden. Zudem sind konkrete Handlungsempfehlungen zu den einzelnen Ängsten zu entwickeln, um den Einführungsprozess zu unterstützen.

Literaturverzeichnis

- [AP98] Agarwal, R.; Prasad, J.: A Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology. In: *Information Systems Research*, 9 (2), 1998; S. 204-215.
- [Aj91] Ajzen, I.: The Theory of Planned Behavior, Organizational Behavior and Human Decision Processes, 50 (2), 1991; S. 179-211.
- [Ba06] Bardram, J. et al.: Experiences from Real-World Deployment of Context-Aware Technologies in a Hospital Environment. In: (Dourish, P.; Friday, A. Hrsg.) *Proceedings of Ubicomp 2006: Ubiquitous Computing*. Springer, Berlin, Heidelberg, 2006.
- [BBA05] Barger, T., Brown, D., Alwan, M.: Health-status monitoring through analysis of behavioural. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 35 (1), 2005.
- [BBF06] Bardram, J.; Baldus, H.; Favela, J.: Pervasive Computing in Hospitals. In (Bardram, J.; Mihailidis, A.; Wan, D. Hrsg.): *Pervasive Computing in Healthcare*. CRC Press, London, 2006; S. 40-78.
- [BKR08] Bick, M.; Kummer, T.-F.; Rössig, W.: Kritische Analyse der Einsatzgebiete von Ambient Intelligence in Krankenhäusern. In: (Breitner, M. et al. Hrsg.): *MMS 2008: Mobilität und Ubiquitäre Informationssysteme*. Köllen Druck+Verlag GmbH, München, 2008; S. 178-191.
- [Ch98] Chin, W. W.: The Partial Least Squares Approach to Structural Equation Modeling. In: *Modern Methods for Business Research* (Marcoulides, G. A. Hrsg.) Lawrence Erlbaum Associates, Mahwah, NJ, 1998; S. 295-336.
- [Co88] Cohen, J.: *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum, Hillsdale; 1988.
- [Da89] Davis, F. D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. In: *MIS Quarterly*, 13 (3), 1989; S. 319-340.
- [ET93] Efron, B.; Tibshirani, R. J.: *An Introduction to the Bootstrap*, Chapman and Hall/CRC, Boca Raton (u.a.), 1993.
- [EU06] The RFID Revolution: Your voice on the Challenges, Opportunities and Threats – Online Public Consultation Preliminary Overview of the Results. European Commission Information Society and Media. [WWW-Dokument: http://www.rfidconsultation.eu/docs/ficheiros/Summary_of_Consultation.pdf (14 Nov. 2008)].
- [Fa07] Faul, F.; Erdfelder, E.; Lang, A.-G.; Buchner, A.: G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 2007; S. 175-191.
- [FL81] Fornell, C.; Larcker, D. F.: Evaluating Structural Equation Models with Unobservable Variables and Measurement Errors. In: *Journal of Marketing Research*, 19 (1), 1981; S. 39-50.
- [Ha98] Hair, J.-F. et al.: *Multivariate Data Analysis*, Prentice Hall, Englewood Cliffs, NJ, 1998.
- [IS03] ISTAG – IST Advisory Group: *Ambient Intelligence: from vision to reality*. 2003. [WWW-Dokument: ftp://ftp.cordis.lu/pub/ist/docs/istag-ist2003_consolidated_report.pdf (24.11.2008)].
- [MMM06] Macario, A.; Morris, D.; Morris, S.: Initial Clinical Evaluation of a Handheld Device for Detecting Retained Surgical Gauze Sponges Using Radiofrequency Identification Technology. In: *Arch Surg*, 141, 2006; S. 659-662.
- [MS06] Marcoulides, G. A.; Saunders, C.: PLS: A Silver Bullet. *MIS Quarterly*, 30 (2), 2006, S. 3-10.
- [Mo06] Moran, E. et al.: Mobility in Hospital Work: Towards a Pervasive Computing Hospital Environment. In: *International Journal of Electronic Healthcare*, 3 (1), 2006; S. 72-89.
- [Ra07] Raitoharju, R.: *Information Technology Acceptance in the Finnish Social and Healthcare Sector: Exploring the Effects of Cultural Factors*. Esa Print, Tampere, 2007.

- [RWW05] Ringle, C. M.; Wende, S.; Will, S.: SmartPLS 2.0 (M3) Beta, <http://www.smartpls.de>, Hamburg, 2005.
- [Ri03] Riva, G.: Ambient Intelligence in Health Care. In: *CyberPsychology & Behavior*, 6 (3), 2003; S. 295-300.
- [Ro05] Rodriguez, M. et al.: Agent-based ambient intelligence for healthcare. In: *AI Communications*, 18 (3), 2005; S. 201-216.
- [Sa07] Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen: Kooperation und Verantwortung – Voraussetzungen einer zielorientierten Gesundheitsversorgung. Gutachten 2007 [WWW Dokument] <http://www.svr-gesundheit.de/Startseite/Startseite.htm> (17. Nov. 2008).
- [SC98] Strauss, A.; Corbin, J.: *Basics of Qualitative Research: Techniques and procedures for developing grounded theory*. 2. Aufl. Sage Publications, London, 1998.
- [Sp08] Spiekermann, S.: *User Control in Ubiquitous Computing: Design Alternatives and User Acceptance*. Shaker Verlag, Aachen, 2008.
- [TT02] Teddlie, C.; Tashakkori, A.: Major Issues and Controversies in the Use of Mixed Methods in the Social and Behavioral Sciences. In: (Tashakkori, A. et al. Hrsg.): *Handbook of Mixed Methods in Social and Behavioral Research*. Sage Publications, Thousand Oaks, 2002; S. 3-50.
- [Te05] Tenenhaus, M.; Vinzi, V. E.; Chatelin, Y.-M.; Lauro, C.: PLS path modelling. *Computational Statistics & Data Analysis*, 48 (1), 2005; S. 159-205.
- [VD00] Venkatesh, V.; Davis, F. D.: A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. In: *Management Science*, 46 (2), 2000; S. 186-204.
- [Wo74] Wold, H.: Causal Flows with Latent Variables: Parting of the Ways in the Light of NIPLAS Modelling. In: *European Economic Review*, 5 (1), 1974; S 67-86.
- [Wo82] Wold, H.: Soft Modelling: The Basic Design and Some Extensions, In *Systems Under Indirect Observation, Part II.*, Vol. 1-54 (Jöreskog K. G.; Wold, H. Hrsg.) Elsevier Science, North-Holland, Amsterdam, New York, Oxford; S. 1-54.
- [WWL07] Wu, J.-H.; Wanga, S.-C.; Lind, L.-M.: Mobile computing acceptance factors in the healthcare industry: A structural equation model. In: *International Journal of Medical Informatics*, 76 (1), 2007; S. 66-77.

Realisierung eines dezentralen Recommender Systems für PDAs

Henrik Mühe, Wolfgang Wörndl, Georg Groh

Lehrstuhl für Angewandte Informatik / Kooperative Systeme
Technische Universität München
Boltzmannstr. 3
85748 Garching
{muehe, woerndl, grohg}@in.tum.de

Abstract: Dezentrale Recommender Systeme erscheinen in mobilen Szenarien sinnvoll, wurden allerdings bisher noch kaum untersucht und erprobt. Wir haben daher in diesem Projekt einen Ansatz zur Empfehlung von Gegenständen (Items) auf Personal Digital Assistants (PDAs) realisiert, der ein gemeinsames Display zusätzlich zu den PDAs einbindet. Dabei werden Bewertungsvektoren zwischen den PDAs ausgetauscht, lokale Matrizen der Item-Ähnlichkeit errechnet und zur Empfehlung ausgewertet. Unsere Neuerungen gegenüber existierenden Ansätzen betreffen insbesondere die Erweiterbarkeit des Modells und die Optimierung des auf einem mobilen Gerät erforderlichen Speicherbedarfs. Die Tauglichkeit unseres Ansatzes wurde in einem kleinen Anwendertest evaluiert. Des Weiteren wurde die Skalierbarkeit mit einem Standard-Datensatz im Bereich Recommender Systeme nachgewiesen.

1 Einführung

Recommender Systeme sind ein mittlerweile etablierter Forschungsbereich der Informatik, der auch erfolgreiche Verwendung in kommerziellen Anwendungen wie etwa bei Amazon oder anderen Online-Shops gefunden hat. Das Ziel dabei ist es, Empfehlungen für Produkte wie Bücher oder andere Gegenstände (engl. *items*) – oft auf Basis von explizit abgegebenen Bewertungen – zu generieren. Zumeist laufen diese Recommender Systeme zentralisiert auf einem Server ab. Eine dezentrale Speicherung und Auswertung von Daten ohne zentralen Server verspricht jedoch einige Vorteile. Dezentrale Recommender können auch ohne Verbindung zu einem Server eingesetzt werden und verbessern somit die Portabilität des Systems [MKR04]. Empfehlungen können lokal abgegeben werden, indem Benutzer, die sich in der Nähe aufhalten, einbezogen werden. Des Weiteren wird eine Dezentralität von Recommender Systemen in der Literatur als Möglichkeit angesehen, die Kontrolle eines Benutzers über seine persönlichen Daten, wie etwa Bewertungen, und damit seine Privacy (Privatheit, Datenschutz), zu verbessern (siehe z.B. [Ko07]). Dezentrale Recommender Systeme wurde bisher jedoch vergleichsweise wenig untersucht oder in Anwendungsszenarien getestet.

Das hier betrachtete Szenario besteht aus Benutzern, die sich Informationen auf ihren privaten mobilen Endgeräten, wie z.B. einem Personal Digital Assistant (PDA), empfehlen lassen wollen. Items sollen dazu auf den PDAs personalisiert verwaltet und angezeigt werden. Ein Beispiel wären Messebesucher, die auf ihrem PDA Informationen zu einer Messe, wie z.B. verfügbare Ausstellungsstände, Produkte und Nachrichten bewerten und abfragen wollen. Zusätzlich sollen in unserem System als Neuerung gemeinsame Displays eingebunden werden die öffentlich für eine Gruppe von Personen – die zu einem Zeitpunkt anwesenden Benutzer – Informationen darstellen.

Das Ziel dieser Arbeit ist es, ein Konzept für die Umsetzung dieses Szenarios zu erarbeiten, zu implementieren und zu testen. Als konkretes Anwendungsbeispiel dient die Bewertung und Empfehlung von Bildern, wobei das Konzept auch auf andere Itemkategorien übertragen werden kann. Der Rest dieses Beitrags ist wie folgt organisiert: Zunächst erläutern wir einige Grundlagen zu Recommender Systemen und stellen einen existierenden Ansatz aus der Literatur im Bereich dezentrale Recommender vor. Im folgenden Kapitel 3 werden dann der Entwurf und die Implementierung unseres Systems mit Benutzerschnittstelle beschrieben. In Kapitel 4 diskutieren wir die Evaluierung des Systems, die in Form eines kleinen Anwendertests sowie einer Untersuchung zur Skalierbarkeit durchgeführt wurde. Dabei gehen wir auch auf verwandte Arbeiten ein. Unser Beitrag schließt mit einer kurzen Zusammenfassung und einem Ausblick.

2 Dezentrale Recommender Systeme

Wir werden in diesem 2. Kapitel zunächst eine Einführung in Recommender Systeme allgemein geben, und einen für die Realisierung auf PDAs geeignet erscheinenden Ansatz genauer betrachten. Dieser Ansatz dient dann als Ausgangspunkt des eigenen Systems.

2.1 Recommender Systeme und kollaboratives Filtern

Die Grundidee von Recommender Systemen ist es, Items wie z.B. Bücher oder CDs für einen aktiven Benutzer zu empfehlen. Dazu wird auf Basis von Informationen über den Benutzer, die Items und ggf. weiterer Daten berechnet, inwieweit ein Item dem Geschmack des Benutzers entspricht. Zu den Merkmalen von Empfehlungsalgorithmen gehören neben der Empfehlungsqualität u.a. Speicher- und Laufzeitkomplexität, sowie Anonymität und Erweiterbarkeit des Modells.

Grundsätzlich unterscheidet man individuelle und kollaborative Recommender Systeme. Individuelle Systeme ermitteln auf Basis des Profils des aktiven Benutzers empfehlenswerte Items. Oftmals geschieht dies mit Hilfe von Metadaten, die die Items beschreiben – daher werden diese Systeme auch inhaltsbasierte Recommender genannt. Zur Realisierung kann z.B. ein Regel-basiertes System eingesetzt werden, das auf Basis von explizit ausgewählten oder implizit erlernten Benutzerprofilattributen Items empfiehlt.

Die zweite Recommender Kategorie ist kollaboratives Filtern (engl. *collaborative filtering*, CF). Dazu werden bei der Empfehlung auch andere Benutzer berücksichtigt. Dies erfolgt mit Hilfe von Bewertungen (engl. *rating*, R) der Benutzer für Items, z.B. auf einer Skala von 1 bis 5. Als Bewertungsvektor wird dabei der Vektor aller Bewertungen eines Benutzers bezeichnet. Bei kollaborativen Filtern gibt es wieder zwei Varianten, nämlich speicherbasiertes (engl. auch *User-based CF* genannt) oder modellbasiertes (engl. auch als *Item-based CF* bezeichnet) kollaboratives Filtern.

Der Empfehlungsprozess bei speicherbasiertem kollaborativen Filtern besteht grundsätzlich aus zwei Schritten: 1. Berechnung einer Menge von k Benutzern, die in der Vergangenheit ähnlich wie der aktive Benutzer bewertet haben (Nachbarschaft), 2. Empfehlung von (neuen) Items für den aktiven Benutzer. Zur Nachbarschaftsberechnung im 1. Schritt wird der Bewertungsvektor des aktiven Benutzers mit denen der anderen verglichen. Dazu wurden in der Literatur verschiedenen Metriken vorgeschlagen, z.B. Euklidische Distanz oder Pearson-Spearman Korrelation [AT05]. Dieses Verfahren wertet somit die User-Item Matrix der Bewertungen aus (vgl. Abbildung 1, links). Im 2. Schritt werden dann Items ausgewählt, die der aktive Benutzer noch nicht beurteilt hat, die aber in seiner Nachbarschaft schon als empfehlenswert bewertet wurden.

	<i>item₁</i>	<i>item₂</i>	...	<i>item_i</i>
<i>user₁</i>	R	R	...	R
<i>user₂</i>	R	-	...	-
⋮	⋮	⋮	⋮	⋮
<i>user_{u-1}</i>	-	-	...	R
<i>user_u</i>	-	R	...	R

	<i>item₁</i>	<i>item₂</i>	...	<i>item_i</i>
<i>item₁</i>	1	...		
<i>item₂</i>	S	1	...	
⋮	⋮	⋮	⋮	
<i>item_i</i>	S	S	...	1

Abbildung 1: User-Item und Item-Item Matrix

Bei modellbasiertem CF wird nicht die Ähnlichkeit der Benutzer betrachtet, sondern die der Items [SKKR01]. Die User-Item Matrix wird also nicht zeilen-, sondern spaltenweise ausgewertet. Ein prinzipieller Unterschied ist, dass es zunächst unerheblich ist, für welchen Benutzer eine Empfehlung erfolgen soll, und somit die Berechnung im Voraus erfolgen kann. Als „Modell“ dieses Verfahrens ergibt sich eine Item-Item Matrix, wobei ein Element $S_{i,j}$ dieser Matrix die aus den Bewertungen berechnete Ähnlichkeit zwischen Item i und Item j ausdrückt (vgl. Abbildung 1, rechts). Für die eigentliche Empfehlung wird dann schließlich der Bewertungsvektor des aktiven Benutzers herangezogen, und Items vorgeschlagen, die ähnlich zu in der Vergangenheit positiv bewerteten Items sind.

Zu beachten ist, dass modellbasiertes kollaboratives Filtern wenig mit individuellem, inhaltsbasiertem Filtern gemein hat, da die Item-Ähnlichkeit nur auf Basis der Bewertungen der Benutzer errechnet wird – Metadaten von Items spielen keine Rolle. Modellbasiertes CF hat gegenüber Speicher-basiertem CF den Vorteil, dass die Berechnung der Empfehlungen weitaus weniger Rechenkapazität erfordert, weil die Item-Item Matrix als „Zwischenspeicher“ der Item-Ähnlichkeit unabhängig vom aktiven Benutzer berechnet werden kann. Außerdem muss zum Zeitpunkt der Empfehlung nur der Bewertungsvektor des aktiven Benutzers vorliegen, was Vorteile hinsichtlich des Datenschutzes der persönlichen Informationen wie Bewertungen verspricht. Deshalb ist modellbasiertes CF für eine dezentrale Realisierung auf PDAs besonders gut geeignet.

2.2 PocketLens

PocketLens [MKR04] ist ein Ansatz eines dezentralen Recommender Systems aus der Literatur, der als Basis unseres Systems dient. Die grundsätzliche Motivation der Entwicklung von PocketLens liegt in der Tatsache begründet, dass sich die meisten Charakteristika eines modellbasierten Recommenders zwar bereits gut für den dezentralen Einsatz eignen, sich jedoch auch einige Probleme aus der Nutzung einer Item-Item Matrix ergeben. Hierbei ist es insbesondere erforderlich, die Matrix, welche in ihren Feldern die Ähnlichkeit jeweils zweier Items enthält, neu zu generieren, sobald eine neue Bewertung hinzugefügt wird.

Bei diesem Vorgang ist, sofern lediglich der errechnete Ähnlichkeitswert in einem Feld gespeichert wird, eine erneute Betrachtung aller für die Item-Item Ähnlichkeit entscheidender Bewertungen notwendig. Dies verhindert jedoch einen umfassenden Datenschutz, da alle Bewertungsvektoren für den Fall der Neugenerierung des Datenmodells gespeichert werden müssten. Hinzu kommt, dass ein vollständiges Verwerfen der bislang berechneten Distanz bei einer Erweiterung des Modells um lediglich eine Bewertung, die Anforderungen an die genutzte Hardware erhöht.

PocketLens [MKR04] verwendet nun einen modifizierten Algorithmus zur Ermittlung der Ähnlichkeit zweier Items, bei dem Bewertungsvektoren hinzugefügt werden können, ohne dass alle betroffenen Felder der Item-Item Matrix vollständig neu berechnet werden müssen. Hierzu wird in jedem Matrixfeld nicht mehr nur die ermittelte Ähnlichkeit zweier Items gespeichert, sondern es werden die zur Berechnung notwendigen Zwischenergebnisse abgelegt. Aufgrund der Verwendung des Cosinus-Abstandsmaßes speichert PocketLens für jede Kombination zweier Items das Kreuzprodukt ihrer Bewertungsvektoren sowie deren Länge separat. Hierdurch ist es nun möglich, die Matrix um Bewertungen neuer Nutzer zu erweitern, indem ihre Bewertungen lediglich zum zwischengespeicherten Kreuzprodukt und deren Vektorlängen addiert werden, ohne dass das gesamte Datenmodell unter Nutzung aller bisher verwendeten Bewertungsvektoren neu generiert werden muss.

Zusätzlich wird bei PocketLens keine Item-Item Matrix generiert, die eine Empfehlung für beliebige Nutzer erlaubt, sondern lediglich ein Ausschnitt dieser Matrix, der auf einen Anwender fokussiert ist. Da der Algorithmus für die Nutzung auf mobilen Geräten mit zumeist nur einem festen Nutzer konzipiert ist, werden Zeilen und Spalten, die für die Empfehlung eines Items an diesen einzelnen Nutzer unerheblich sind, aus der Matrix entfernt, was die Speicherkomplexität verbessert.

Neben dem geringeren Aufwand bei der Erweiterung der Matrix erlaubt das skizzierte Verfahren außerdem, dass Bewertungsvektoren neuer Nutzer nach der Modellerweiterung verworfen werden können, wodurch eine Verbesserung der Anonymität des Recommender-Algorithmus möglich ist.

3 Entwurf und Implementierung

Auf Basis des im vorherigen Kapitel vorgestellten PocketLens Algorithmus, soll nun auf den Entwurf und die Erweiterungen des in diesem Projekt erarbeiteten Recommenders eingegangen werden [Mu08]. Neben dem verteilten Empfehlungsalgorithmus wird auch erläutert, wie eine Empfehlung an eine Gruppe von Nutzern ausgesprochen werden kann und ein Überblick über die konkrete Implementierung im Projektszenario mit Benutzerschnittstelle geboten.

3.1 Entwurf des Recommenders auf Basis von PocketLens

Der zuvor vorgestellte PocketLens Algorithmus ist bereits zur Nutzung in verteilten Systemen und mit weniger leistungsfähigen Endgeräten vorgesehen. Dennoch zeigt sich bei der Konzeption eines verteilten Recommenders, dass in vielen Szenarien Verbesserungen notwendig sind, um den sinnvollen Einsatz auf PDAs zu ermöglichen.

Zum einen lässt sich leicht erkennen, dass durch die Verwendung eines Modells erheblich mehr Daten gespeichert werden müssen, als dies bei einem speicherbasierten Recommender der Fall wäre. Zum anderen realisiert PocketLens zwar das einfache Hinzufügen neuer Nutzer und ihrer Bewertungsvektoren, erlaubt es aber nicht, eine neue Version eines Bewertungsvektors mit neuen Bewertungen oder gar Änderungen hinzuzufügen. In einem verteilten Szenario jedoch, bei dem Benutzer regelmäßig Daten austauschen, kann dies dazu führen, dass dem Modell ein Benutzer hinzugefügt wird, der erst sehr wenige Items bewertet hat und alle von diesem Nutzer in der Zukunft getätigten Bewertungen nicht mehr in das Modell einfließen.

Um den Algorithmus um die genannten Fähigkeiten zu erweitern, wurde zunächst untersucht, in wieweit eine Optimierung des Datenmodells möglich ist und ein Konzept entwickelt, wie veränderte Bewertungsvektoren eines bekannten Nutzers in das Modell integriert werden können.

3.1.1 Speicherung des Recommender-Modells

Um die Matrix von Ähnlichkeitswerten, die von PocketLens genutzt wird, möglichst kompakt speichern zu können, wurden deren Merkmale untersucht. Hierzu wurde PocketLens implementiert und der Aufbau der Speichermatrix beim Hinzufügen von Daten auf die Eigenschaften Größe im Vergleich zum speicherbasierten Recommender, Besetztheit und Doppeleinträge untersucht.

Mit zunehmender Zahl an hinzugefügten Bewertungsvektoren steigt der Speicherbedarf des Modells stark an. Die Matrix ist bereits nach dem Hinzufügen weniger Vektoren dicht besetzt, so dass die Nutzung eines Algorithmus zur Speicherung dünn-besetzter Matrizen keine Speicherersparnis erzielen würde. Unter den ausgefüllten Feldern befinden sich jedoch viele Duplikate, d.h. Felder, die den exakt gleichen Inhalt haben, so dass an dieser Stelle Speicher gespart werden kann. Ausgehend von einer dicht-besetzten Zeigermatrix ist jeder Feldinhalt bei PocketLens zumindest 20 Byte, jeder Zeiger jedoch lediglich 4 Byte groß, so dass für jedes doppelte Element in der Matrix eine Einsparung von 16 Byte möglich ist, wenn lediglich Zeiger auf einzigartige Feldinhalte abgelegt werden.

Neben der so erzielten Einsparung, kann die Definition gleicher Feldinhalte verändert werden. PocketLens speichert Fließkommawerte für Ähnlichkeit und Vektorlänge als Inhalt der Matrixfelder. Definiert man gleiche Felder als genau die Felder, die sich bis zu einer bestimmten Nachkommastelle und nicht bitgenau nicht unterscheiden, ergibt sich ein geringer Einfluss auf die Empfehlungsfindung, jedoch erhöht sich die Anzahl der nunmehr „ungenauen“ Duplikate erheblich, so dass eine höhere Speicherersparnis möglich ist. Kombiniert man dieses Verfahren mit einer dynamischen Anpassung der Vergleichsgenauigkeit, können unter Verringerung der Empfehlungsqualität erheblich größere Modelle verarbeitet werden, als dies bei einer direkten Speicherung der Fall wäre. Die Beschreibung der Evaluierung unseres Ansatzes mit einem Standard Datensatz im Bereich Recommender Systeme erfolgt in Kapitel 4.2. dieses Beitrags.

3.1.2 Versionierte Bewertungsvektoren

Zur Verbesserung der Erweiterbarkeit von PocketLens ist zunächst festzuhalten, dass diese nicht auf Kosten der Anonymität erfolgen soll, d.h. es soll nicht notwendig sein, die Bewertungsvektoren aller Nutzer vorzuhalten. Außerdem ist darauf zu achten, den Mehrverbrauch an Speicher so gering wie möglich zu halten.

Um die genannten Kriterien erfüllen zu können, wurde ein Algorithmus entwickelt, der es erlaubt, den Einfluss eines Bewertungsvektors auf das Modell zurückzusetzen, ohne dafür sämtliche in das Modell eingeflossenen Bewertungsvektoren zu kennen. Hierzu ist es zunächst notwendig, das Konzept des Bewertungsvektors von einer Menge an Zwei-Tupeln bestehend aus Item-Identifikator und Bewertung so zu erweitern, dass alte Zustände des Bewertungsvektors nachvollzogen werden können. Das daraus entstandene Konzept wird mit dem Namen „Versionierter Bewertungsvektor“ bezeichnet und ist nunmehr eine Menge an Zwei-Tupeln, welche neben dem Item-Identifikator wiederum eine Menge an Versionsnummern und dazugehörigen Bewertungen enthält. Die für jedes Item enthaltenen Versions- und Bewertungs-paare werden als rückwärts-verkettete Liste realisiert, um einen schnellen Zugriff auf die aktuellste Bewertung zu ermöglichen (vgl. Abbildung 2).

Bewertungsvektor		Versionierter Bewertungsvektor	
Subjekt	Wertung	Subjekt	Versionierte Wertungen
A	4	A	(1,5) ← (4,4)
C	1	B	(2,4) ← (5,0)
		C	<u>(3,1)</u>

Abbildung 2: Vergleich der konkreten Ausprägung eines Bewertungsvektors und eines versionierten Bewertungsvektors

Bei der Änderung einer Bewertung wird beim versionierten Bewertungsvektor nun lediglich ein weiterer Eintrag in die Liste des entsprechenden Items eingefügt, so dass alte Zustände rekonstruierbar bleiben. Fügt ein Endgerät den Bewertungsvektor eines Benutzers zu seinem Datenmodell hinzu, so speichert es die Version des hinzugefügten Vektors, die als Maximum alle einzelnen Versionsnummern alle enthaltener Bewertungen definiert ist. Soll eine neue Version mit geänderten oder zusätzlichen Bewertungen in das Modell einfließen, kann der Einfluss des zuvor in PocketLens eingefügten, alten Vektors rückgängig gemacht werden, indem aus dem hinzuzufügenden Vektor zunächst die alte, bereits genutzt Version errechnet, deren Einfluss auf das Modell rückgängig gemacht und dann die aktuelle Version hinzugefügt wird (vgl. [Mu08]).

Insgesamt ist es durch die Einführung versionierter Bewertungsvektoren möglich, PocketLens um neue Bewertungen bekannter Nutzer zu erweitern und Änderungen an bekannten Bewertungen vorzunehmen. Hierbei wird der erhöhte Speicheraufwand für geänderte Bewertungen und für die Speicherung des komplexeren, modifizierten Bewertungsvektors jeweils vom verursachenden Nutzer getragen. Insbesondere wird hierdurch der Aufwand zur Speicherung der Ähnlichkeitsmatrix nicht erhöht und es muss pro hinzugefügten Bewertungsvektor lediglich die genutzte Version, jedoch nicht etwa der gesamte Bewertungsvektor oder pro modifizierten Matrix Feld ein Wert gespeichert werden.

Schließlich gewährleistet der skizzierte Algorithmus immer noch den gleichen Grad an Anonymität, wie ihn auch PocketLens selbst bietet, so dass vor allem keine Speicherung der konkreten Bewertungsvektoren jedes Nutzers notwendig ist und sich diese auch nicht aus dem gespeicherten Modell errechnen lassen.

3.2 Vorgehensweise zur Gruppenempfehlung

In unserem Szenario soll das dezentrale Recommender System nun auch genutzt werden, um für eine zu einem bestimmten Zeitpunkt vor einem gemeinsamen Display anwesende Gruppe von Personen, Items empfehlen zu können. Die Gruppenbildung kommt dadurch zustande, dass der Rechner, der das gemeinsame Display steuert, auch die Bewertungsvektoren der anwesenden PDAs empfängt. Eine Gruppe – und damit der gemeinsame Kontext – besteht dann aus der Menge der PDAs (beziehungsweise deren Benutzer), die in einem Zeitfenster ihren Vektor geschickt haben und damit als aktiv vor den Display erkannt wurden.

Für eine Empfehlung für eine Gruppe von Nutzern muss das vorgestellte, auf einen einzelnen Nutzer bezogene Verfahren erweitert werden. Es ist insbesondere nötig, die Repräsentation einer Gruppe mit mehreren Benutzern, die jeweils verschiedene Items verschieden bewertet haben, derart zu modifizieren, dass diese Gruppe für den Recommender-Algorithmus als ein kombinierter „Pseudouser“ wirkt [JS07].

Um einen kombinierten Bewertungsvektor für eine Gruppe zu erhalten, verwenden wir die Bildung eines Mittelwertes der Bewertungsvektoren aller beteiligten Gruppenmitglieder. Dazu wird zunächst eine Menge G aller von mindestens einem Gruppenmitglied bewerteten Items ermittelt. Der Bewertungsvektor der Gruppe kann dann gebildet werden, indem für jedes Element des Vektors, welches ein von mindestens einem Gruppenmitglied bewertetes Item repräsentiert, die Bewertungen der Nutzer, die diesen tatsächlich bewertet haben, aufsummiert und durch die Anzahl dieser Benutzer geteilt werden. Der so erhaltene Vektor repräsentiert nun den Meinungsdurchschnitt der Gruppe. Dieser Bewertungsvektor kann dann im vorgestellten Algorithmus genutzt werden, um Empfehlungen zu generieren. Der Vorteil der Berechnung eines kombinierten Bewertungsvektors für eine Nutzergruppe liegt somit darin, dass der eigentliche Recommender-Algorithmus im Vergleich zur Empfehlung für einen einzelnen Benutzer unverändert verwendet werden kann.

Neben der Anzeige von Empfehlungen an eine Gruppe, können auch am gemeinsamen Display Bewertungen abgegeben werden. Dazu wird für eine anwesende Gruppe ein neuer, anonymer, Benutzer beziehungsweise Bewertungsvektor angelegt, und abgegebene Bewertungen diesem Benutzer zugeordnet. Erfolgen mehrere Bewertungen in geringem zeitlichem Abstand, werden diese gemeinsam dem anonymen Bewertungsvektor hinzugefügt. Der Bewertungsvektor wird dann – analog zum Verhalten der mobilen Endgeräte – an alle anderen in der Nähe befindlichen Geräte übertragen, damit auf den PDAs die Item-Item Matrizen entsprechend aktualisiert werden können. Hierdurch ist es sowohl möglich, ohne eigenes Endgerät an der Empfehlungsbildung der anderen, mobilen Nutzer teilzunehmen, als auch, die Gruppenempfehlung für eine festgelegte Zeit mit zu beeinflussen.

3.3 Implementierung

Der erweiterte Algorithmus wurde zur Generierung von Bildempfehlungen für PDAs mit dem Betriebssystem Windows Mobile implementiert [Mu08]. Hierzu erfolgte zunächst eine Umsetzung des in diesem Beitrag beschriebenen Recommender-Algorithmus, welcher dann für die Empfehlung von Bildern spezialisiert wurde. Die Matrix, in der Item-Item Ähnlichkeiten gespeichert werden, ist hierbei als Referenzmatrix auf konkrete Feldinhalte implementiert, wobei eine Duplikatsuche und Elimination aus Gründen der Laufzeit nicht beim Einfügen in die Matrix sondern asynchron beim Erreichen der Speichergrenze des mobilen Endgeräts erfolgt.

Da ein möglichst realistisches Szenario abgedeckt werden sollte, wurde ein Protokoll für den Austausch von Bewertungsvektoren entwickelt, welches lediglich eine bestehende Verbindung in einem gemeinsamen Netzwerk voraussetzt. Konkret wurde ein auf UDP Multicast basierendes Netzwerkprotokoll entwickelt, welches es ermöglicht auch Datenmengen jenseits der Limitierungen für einzelne UDP Pakete auszutauschen, ohne dabei jedoch eine Aussage über den Erfolg der Übermittlung treffen zu können.

Die Nutzung eines Multicast Protokolls bietet sich insbesondere deshalb an, weil die Konstruktion des Algorithmus keine bidirektionale Kommunikation erfordert, und der eigene versionierte Bewertungsvektor lediglich anderen Teilnehmern zur Verfügung gestellt werden muss. Hat ein anderer Nutzer einen solchen Vektor empfangen, kann dieser unabhängig von der Frage, ob eine ältere Version bereits in das Modell eingeflossen ist oder nicht in das Empfehlungsmodell integriert werden. Für die Speicherung des aktuellen Modellzustands sowie für den Austausch zwischen den mobilen Geräten untereinander sowie mit dem Touchscreen Display werden die Datenstrukturen in eine XML Repräsentation umgewandelt und mittels „bzip2“ komprimiert.



Abbildung 3: Mobile Benutzerschnittstelle

3.4 Benutzerschnittstelle

Abbildung 3 zeigt die Benutzerschnittstelle am Mobilgerät. Angezeigt wird ein Bild mit der Möglichkeit, dieses von 1 bis 5 (Sternen) zu bewerten, oder gegebenenfalls eine bereits abgegebene Bewertung anzuzeigen und zu ändern. Es ist möglich in eine „Landscape“ Ansicht zu wechseln (vgl. Abbildung 3, rechts), um die Bilder besser betrachten zu können.

In einem Menü kann man einstellen, ob man zufällige oder empfohlene Bilder anzeigen lassen will. Zufallsbilder sind sinnvoll, um Anfang einige Bewertungen abgeben zu können, auf deren Basis dann neue empfohlene Items ermittelt werden können, wie oben erläutert. Nach Auswahl der Funktion „Weiter“ wird ein weiteres Bild angezeigt. Die Benutzerschnittstelle wurde im Hinblick auf eine möglichst einfache Bedienbarkeit entworfen.

Als gemeinsames Display wird ein Monitor mit Touchscreen-Funktion verwendet. Dabei wird zunächst ein für die aktuell anwesende Gruppe von Personen empfohlenes Bild ohne Bewertungsleiste angezeigt. Berührt ein Benutzer den Touchscreen, findet ein Moduswechsel statt und ein Benutzer kann mittels einer Bewertungsleiste – analog zu der im mobilen Endgerät verwendeten Leiste – eine Bewertung für das Bild abgeben (vgl. Abbildung 4). Zusätzlich werden in unserem Prototypen rechts Informationen über aktuell als anwesende erfasste Benutzer bzw. PDAs, sowie Debug-Informationen, angezeigt. Dies würde bei einem realen Einsatz ausgeblendet werden. Zu bemerken ist, dass das Touchscreen Interface in unserem System keinen Server darstellt, sondern im Wesentlichen auch nur ein Peer mit angepasster Benutzerschnittstelle ist.



Abbildung 4: Szenario mit Touchscreen Monitor

4 Evaluierung

Das vorgestellte System wurde in einem (kleinen) Anwendertest in einem realistischen Szenario evaluiert [Mu08]. Dieser Test soll nicht eine formale Evaluierung des Systems darstellen, sondern die praktische Einsetzbarkeit zeigen und durch einen Fragebogen einige Tendenzen bezüglich interessanter Fragestellungen ermitteln. Des Weiteren wurden die Skalierbarkeit und Performance anhand eines Standard-Datensatzes für Recommender Systeme getestet. Schließlich diskutieren wir verwandte Arbeiten.

4.1 Anwendertest

Der Anwendertest wurde mit 13 Personen und Windows Mobile PDAs von HTC und HP durchgeführt, z.B. einem HTC P3600 PDA Phone. Als gemeinsames Display kam ein 32“ Monitor mit Touchscreen Funktionalität zum Einsatz (vgl. Abbildung 4). Die Itemmenge bestand aus 63 vorgegebenen Bildern. Zunächst sollten die Benutzer, Bewertungen für etwa 15-20 zufällig gewählte Bilder auf ihren PDAs abgeben. Anschließend wurden die so generierten Bewertungsvektoren ausgetauscht, wodurch vom Touchscreen Display ein Modell für die Gruppenempfehlung und von jedem einzelnen Endgerät ein Modell für die persönliche Empfehlung ermittelt werden konnte. Die so erstellten Empfehlungsmodelle wurden anschließend getestet. Hierzu betrachteten die Benutzer zum einen jeder für sich die ihnen persönlich vorgeschlagenen, neuen Bilder, zum anderen die in verschiedenen Gruppenkonstellationen der jeweiligen Gruppe vorgeschlagenen Bilder. Es befand sich immer nur eine Teilmenge der Benutzer vor dem gemeinsamen Display, die maximale gleichzeitige Gruppengröße in unserem Test war 6 Personen.

Der Test zeigte zunächst, dass die Anwendung praxistauglich ist und im dargestellten Szenario funktionierte. Sowohl die Berechnung der Item-Item Matrix als auch die Ermittlung der Empfehlungen lief ohne Performanceprobleme. Die Teilnehmer waren im Anschluss an den Test aufgefordert, einen Fragebogen auszufüllen. Bei den meisten Fragen sollte ein Teilnehmer Aussagen auf einer Skala der Übereinstimmung von -2 für „trifft nicht zu“ bis +2 für „trifft zu“ mit Zwischenstufen beurteilen. Der verwendete Fragebogen sowie die detaillierteren Ergebnisse sind in [Mu08] zu finden.

Zunächst wurde abgefragt, inwieweit den Testteilnehmern der Schutz ihrer persönlichen Daten allgemein wichtig ist. Dies war mit einer mittleren Übereinstimmung von +1,38 der Fall. Außerdem wurde eine Aussage zur Beurteilung vorgelegt, ob die Nutzung eines Bewertungsprofils zur Generierung von Empfehlungen prinzipiell einen Eingriff in die Privatsphäre darstellt. Dabei ergab sich kein einheitliches Bild, die Tester nutzen die ganze Bandbreite der Skala relativ gleichmäßig aus. Eine weitere Frage, ob ein Recommender System, welches garantiert, dass die konkreten Bewertungen und Daten nicht offengelegt werden, die Bereitschaft zur Nutzung erhöhen würde, wurde mit vielen positiven Einschätzungen beurteilt (Mittelwert +0,92).

Bei der Evaluierung des getesteten Systems an sich, wurde zunächst eine Beurteilung der Dezentralität des Systems untersucht. Dabei erkannten fast alle Benutzer die Vorteile der Unabhängigkeit von einem Server im getesteten Szenario. Zur Beurteilung der Empfehlungsqualität könnten die Tester zunächst beurteilen, ob ihnen die vorgeschlagenen Bilder angemessen und sinnvoll – in Bezug auf ihre Bewertungen – erschien. Dies wurde mit einem Mittelwert von +0,54 (schwach) positiv beurteilt. Als weitere Frage sollten die Tester dann die Empfehlungsqualität der persönlichen Empfehlungen nach Austausch der Bewertungsvektoren begutachten. Dabei ergab sich mit einem Mittelwert von +0,69 auch eine positive Tendenz. Im Nachhinein betrachtet war für eine noch differenzierte Analyse der Empfehlungsqualität aus Nutzersicht unsere Itemmenge von 63 Fotos wahrscheinlich zu homogen.

Im letzten Teil des Fragebogens konnten die Tester die Einsetzbarkeit der Anwendung begutachten. Dabei wurden sowohl die Frage „Ein System dieser Art für die Bewertung und Empfehlung von Bildern/Musik/Videos halte ich für sinnvoll“ wie auch „Ich könnte mir vorstellen, ein solches System aktiv zu nutzen.“ mit einem Mittelwert von +0,92 bzw. +0,69 positiv beantwortet. Abschließend konnten die Teilnehmer noch konkrete Anwendungsgebiete für das System angeben, dabei wurde neben „Messen“ und „Konferenzen“ auch „Einkauf“, „Dezentrale Filesharing Netzwerke“, „Parties“ und „Museen“ genannt.

4.2 Test der Skalierbarkeit

Neben dem praktischen Test haben wir auch die Skalierbarkeit des implementierten und in Abschnitt 3.1. genauer erläuterten Algorithmus‘ näher untersucht. Dazu haben wir den oft für die Evaluierung von Recommender System herangezogenen „MovieLens“ Datensatz des GroupLens Projektes (siehe <http://www.grouplens.de>) der University of Minnesota verwendet [HKBR99]. Der Datensatz umfasst in der von uns benutzten Variante 100.000 Bewertungen, die von 943 Benutzern für 1682 Filme abgegeben wurden. Jeder Benutzer hat dabei mindestens 20 Bewertungen durchgeführt.

Das Datenmodell von PocketLens kann durch Elimination von echten Duplikaten von einer Größe von etwa 45 Megabyte auf eine Größe von 32 Megabyte reduziert werden. Vergleicht man die einzelnen Feldinhalte mit einer Genauigkeit von lediglich 0,05, so reduziert sich der erforderliche Speicher auf unter 20 Megabyte, so dass der gesamte Datensatz – verlustbehaftet – in den Speicher des PDAs geladen werden konnte.

Um eine möglichst geringe Laufzeit zu erreichen, wird eine Duplikatsuche lediglich dann ausgeführt, wenn eine bestimmte Speicherschranke überschritten wird. Im Mittel benötigten die Testgeräte etwa 1,5 Sekunden zum Einfügen eines Bewertungsvektors, was das Verfahren für den skizzierten Anwendungsfall tauglich macht. Der Großteil der Laufzeit wird hierbei für die nebenläufige Duplikatelimination benötigt, während nur ein Bruchteil der Laufzeit auf das Einfügen eines einzelnen Bewertungsvektors entfällt.

4.3 Verwandte Arbeiten

Neben dem ausführlichen besprochenen PocketLens gibt es einige andere Ansätze in eine ähnliche Richtung. [BKR07] stellt ein verteiltes Recommender System vor, das eine Partitionierung der User-Item Matrix auf Basis einer Domänen-spezifischen Aufteilung der Items vornimmt. Es sind dazu allerdings zusätzliche Informationen über die Anwendungsdomäne nötig. Einige Beiträge wie [WPLR06] diskutieren dezentrales kollaboratives Filtern in Peer-to-Peer (P2P) Netzwerken, wobei allerdings keine mobilen Peers wie PDAs berücksichtigt werden. [JRH06] bespricht einen CF Ansatz für mobile „Media Agents“. Dabei werden den Items, wie z.B. Musik Dateien, Informationen und Regeln in einem Profil mitgegeben. Die Items sollen dann selbstständig als Software-Agenten interagieren und ein CF-ähnliches Verhalten simulieren.

Bezüglich der Empfehlung für Gruppen von Personen gibt [JS07] einen Überblick über den aktuellen Stand der Forschung. Ein frühes und bekanntes Beispiel ist das „PolyLens“ System [OCKR01], was zur Unterstützung einer Gruppe von Personen beim gemeinsamen Kinobesuch entwickelt wurde. Dabei wird kollaboratives Filtern verwendet. Die empirischen Untersuchungen in diesem Projekt haben gezeigt, dass Benutzer nicht nur die Empfehlungen des Systems als wertvoll eingeschätzt haben, sondern auch bereit waren, einen Teil ihrer Privacy für die Gruppenempfehlungen aufzugeben [OCKR01]. PolyLens ist allerdings kein dezentrales Recommender System. Ein mit unserem System direkt vergleichbarer und auf PDAs realisierter Ansatz ist uns nicht bekannt.

5 Fazit

In diesen Beitrag haben wir eine Realisierung eines dezentralen Recommender Systems für PDAs vorgestellt. Dabei haben wir insbesondere eine private Empfehlung und Anzeige auf einem PDA mit einem gemeinsamen Display für eine Empfehlung für Gruppen von Personen integriert. Darüberhinaus haben wir gegenüber dem existierenden PocketLens Ansatz eine verbesserte Erweiterbarkeit durch versionierte Bewertungsvektoren eingeführt, sowie eine Optimierung des auf einem mobilen Gerät erforderlichen Speicherbedarfs erzielt.

Neben einem weitergehenden Praxistest unseres Systems, auch mit einer anderen Itemmenge, sind verschiedene Erweiterungen für zukünftige Arbeiten denkbar. Zum einen könnte die Effizienz beim Prozess der Generierung von Empfehlungen noch weiter erhöht werden. Dazu könnten eine Speicherung relevanter Zwischenergebnisse und Änderungsmarkierungen eingesetzt werden, um auch bei einem großen Datenmodell oder einem sehr schwachen Mobilgerät Empfehlungen ableiten zu können. Auch kann das Verfahren zur Ermittlung von Empfehlungen für eine Gruppe von Personen insbesondere durch eine Prüfung weiterer Methoden zur Generierung des repräsentativen Bewertungsvektors einer Gruppe verbessert werden.

Schließlich wäre es von Interesse, den Datenaustausch zwischen einzelnen Benutzern weniger explizit zu gestalten und mit geeigneten Methoden, wie etwa asymmetrischen Verschlüsselungsverfahren oder einem impliziten Austausch von Bewertungen und Modellteilen, einen höheren Grad an Anonymität zu erreichen.

Literaturverzeichnis

- [AT05] Adomavicius, G.; Tuzhilin, A.: Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 6, 2005.
- [BKR07] Berkovsky, S.; Kuflik, T.; Ricci, F.: Distributed collaborative filtering with domain specialization. In *Proceedings of the 2007 ACM Conference on Recommender Systems*, Minneapolis, MN, 2007.
- [HKBR99] Herlocker, J.L.; Konstan, J.A.; Borchers, A.; Riedl, J.: An algorithmic framework for performing collaborative filtering. *Proceedings of the 22nd Annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley, CA, 1999.
- [JRH06] Jacobsson, M.; Rost, M.; Holmquist, L. E.: When Media Gets Wise: collaborative filtering with mobile media agents. In *Proceedings of the 11th International Conference on Intelligent User Interfaces (IUI '06)*, Sydney, Australia, 2006.
- [JS07] Jameson, A.; Smyth, B.: Recommendation to Groups. In (Brusilovsky, P.; Kobsa, A.; Nejdl, W., Eds.): *The Adaptive Web*, LNCS 4321, Springer Verlag, Berlin/Heidelberg, 2007; S. 596-627.
- [Ko07] Kobsa, A.: Privacy-Enhanced Personalization. *Communications of the ACM*, Vol. 50, No. 8, 2007; S. 24-33.
- [MKR04] Miller, B.N.; Konstan, J.A.; Riedl, J.T.: PocketLens: Toward a Personal Recommender System. *ACM Transactions on Information Systems*, Vol. 22, No. 3, 2004; S. 437-476.
- [Mu08] Mühe, H.: Entwurf und Realisierung eines dezentralen Recommender Systems für die Empfehlung von Items für Gruppen. Bachelorarbeit in Informatik, Technische Universität München, 2008.
- [OCKR01] O'Connor, M.; Cosley, D.; Konstan, J. A.; Riedl, J.: PolyLens: a recommender system for groups of users. In *Proceedings of the Seventh Conference on European Conference on Computer Supported Cooperative Work (ECSCW)*, Bonn, 2001.
- [SKKR01] Sarwar, B.; Karypis, G.; Konstan, J.; Riedl, J.: Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on the World Wide Web*, Hong Kong, China, 2001.
- [WPLR06] Wang, J.; Pouwelse, J.; Lagendijk, R. L.; Reinders, M. J.: 2006. Distributed collaborative filtering for peer-to-peer file sharing systems. In *Proceedings of the 2006 ACM Symposium on Applied Computing (SAC 06)*, Dijon, France, 2006.

Blended Shopping

Britta Fuchs, Thomas Ritz

Fachbereich Elektrotechnik und Informationstechnik
Fachhochschule Aachen
Eupener Straße 70
52066 Aachen
b.fuchs@fh-aachen.de
ritz@fh-aachen.de

Abstract: Neben dem klassischen Ladenverkauf hat sich insbesondere für Konsumgüter der Verkauf von Waren über das Internet (eCommerce) als Verkaufsweg etabliert. Der Artikel startet mit der These, dass diese Vertriebswege, insbesondere in den kundennahen Prozessen, völlig getrennt voneinander ablaufen. Diese These wird durch eine Feldstudie belegt. Die Anreicherung des Ladenverkaufs durch eCommerce-Ansätze, also die bisher fehlende Integration von traditionellem Ladenverkauf und eCommerce, wird als Blended Shopping eingeführt, die Potenziale erarbeitet und schließlich anhand von zwei prototypischen Implementierungen visualisiert und belegt. Der Artikel schließt mit einem Ausblick, welche technischen wie betriebswirtschaftlichen Vorarbeiten notwendig sind, um Blended Shopping effektiv und effizient in die Praxis bringen zu können.

1 Einleitung

Im Konsumgüterbereich hat sich neben traditionellem Handel in den letzten Jahren eCommerce durch die starke Verbreitung des Internet etabliert, wie wachsende Umsätze belegen [Ro07]. Nach einer ersten Hype-Phase der Überbewertung von eCommerce [KAS08] konnten inzwischen weitreichende Erfahrungen über Möglichkeiten und Einschränkungen sowohl im Ladenverkauf als auch im eCommerce gewonnen werden.

Aus Forschungssicht wäre es durchaus interessant zu evaluieren, ob man durch die Kombination beider Vertriebswege im Rahmen eines Verkaufsprozesses eines Gutes bisher noch ungenutzte Potentiale wecken kann. Die Arbeit geht von der These aus, dass eCommerce und traditioneller Ladenverkauf bisher organisatorisch wie technisch getrennt betrachtet werden und belegt diese These im weiteren Verlauf. Im Rahmen dieser Arbeit werden daher zunächst die bekannten Vertriebswege im Detail betrachtet, um basierend darauf eine Definition für Blended Shopping, also die Anreicherung von traditionellem Verkauf im Ladenlokal um eCommerce-Elemente, einzuführen. Mit einer Feldstudie wird die These belegt, dass eCommerce und Ladenverkauf bisher technisch wie organisatorisch getrennt ablaufen, und untersucht, ob und wie mit den bisherigen Online-Angeboten Blended Shopping realisiert werden kann.

Aus der Feldstudie werden die aus Sicht der Autoren entscheidenden Faktoren abgeleitet, die bisher Blended Shopping verhindern. Auf dieser Grundlage werden Prototypen entwickelt, die die benannten Barrieren überwinden. Der Artikel schließt mit einem Ausblick auf notwendige technische und organisatorische Maßnahmen, um Blended Shopping effizient und effektiv einführen zu können.

2 Verkaufsmodele

Die etablierten Verkaufskanäle - Ladenverkauf und eCommerce - unterscheiden sich in Ausprägungen der einzelnen Verkaufsphasen (Information, Vermittlung, Verhandlung, Vertragsabschluss, Aftersales [Me02]). In diesem Kapitel werden beide Kanäle beleuchtet, um diese Unterschiede herauszustellen.

2.1. Ladenverkauf

Der traditionelle Handel ist der – bis vor einigen Jahren – übliche Absatzkanal im Konsumgüterbereich [AK07]. Die Waren werden in einem Ladenlokal präsentiert, der interessierte Kunde kann sie anfassen, an- und ausprobieren und sie stehen ihm unmittelbar nach der Bezahlung zur Verfügung. Außerdem steht in der Regel ein Verkäufer bereit, um Informationen zu den Produkten zu geben oder Beratungsleistung und passende weitere Produkte aus der Produktpalette anzubieten. Dem Händler entstehen u.a. durch das Ladenlokal, das Personal, die Lagerhaltung und die Kassenführung Kosten. Dafür hat er die Möglichkeit, persönlichen Kontakt zum Kunden aufzunehmen, auf den Verkaufsprozess einzuwirken und weitere Kaufanstöße zu geben [KKB07]. Bezahlt der Kunde bar, findet der gesamte Prozess trotz eines direkten Kundenkontakts anonym statt und es werden keine veredelbaren digitalen Kundendaten erhoben. Strategien, um dies aufzulösen, sind etwa Kundenkarten [MB08].

Für den Kunden ist die Informationsphase relativ aufwendig. Zur Informationssammlung müssen u. U. verschiedene, örtlich getrennte, Ladenlokale aufgesucht, Verkäufer gesucht und Gespräche geführt werden. Es besteht jedoch eher die Möglichkeit, über die Kaufmodalitäten zu verhandeln als beim eCommerce. Grundsätzlich geht man davon aus, dass Ladenlokale aufgrund ihrer örtlichen Präsenz und der persönlichen Bindung eher über einen höheren Anteil von Stammkunden [In07] verfügen als über Laufkundschaft, d.h. Neukunden ohne wiederholte Kaufabsicht. Der traditionelle Handel probiert mittlerweile den Einsatz moderner Technologien aus. Z.B. werden Anwendungen getestet (etwa Mobiler Einkaufsassistent MEA im Real Future Store [Me08], [St08]), die dem Kunden im Ladenlokal u. a. eine digitale Verkaufsberatung zur Verfügung stellen, den selbständigen Scan von Barcodes ermöglichen und eine schnellere und einfachere Zahlungsabwicklung versprechen. Bisher sind die getesteten Lösungen noch nicht ausgereift und auf einen kleinen Problembereich beschränkt. Der Erfolg solcher Neuerungen ist vor allem abhängig von der Akzeptanz durch die Kunden. Nur wenn der Kunde vom tatsächlich vorhandenen Nutzen überzeugt werden kann, wird er Neuerungen aufgreifen. Hier ist eine umfassende Kommunikationsstrategie [Th99] gefragt.

2.2. eCommerce

eCommerce wird hier verstanden als “Handel auf elektronischem Weg” [Me02], also die digitale Durchführung von digitalisierbaren Verkaufsaktivitäten. Der Erfolg beruht u.a. auf der weiten Verbreitung des Internet (42,7 Millionen [AR08] Nutzer in Deutschland in 2008). Die angebotenen Waren werden per Webshop mit einer Beschreibung, häufig illustriert durch Fotos oder neuerlich auch Videos und Animation, angeboten. Den größten Anteil am eCommerce haben die Produktgruppen Bücher (77%), DVD/CD/Video (52%) und Elektrogeräte (44%) [[KW08], es handelt sich also um leicht vergleichbare und preissensible Produkte, die Kaufentscheidung erfolgt gezielt (82%) [Mo08] und eher rational.

Eine veränderte Mediennutzung drückt sich auch in den Ansätzen des Web 2.0 [Or05] aus. Die interaktive Gestaltung von Inhalten durch die Nutzer („user generated content“) [GKT07] hat u. a. eine Verlagerung der Beratungsfunktion bewirkt. Kunden stellen ihre bereits gesammelten Erfahrungen mit Produkten und Dienstleistungen im Internet anderen (potentiellen) Kunden zur Verfügung. Es handelt sich dabei um eine Form von Social Navigation [BWW04], die bereits von 17 Mio. Deutschen genutzt wird [Sc08a]. Einerseits wird also weniger Beratungsleistung durch den Händler nötig, was sich in den Personalkosten niederschlägt, andererseits verliert er die Möglichkeit, Einfluss auf die Kaufentscheidung durch eine differenzierte Produktdarstellung zu nehmen, was sich u.a. in einer höheren Preissensibilität der Nachfrage ausdrückt.

Die hohe Preissensibilität der Kundschaft und die Preistransparenz durch Vernetzung ist sicher ein Grund für die schwer zu etablierende Bindung des Kunden an den Webshop. ECommerce bietet durch relativ einfach und kostengünstig zu sammelnde Kundendaten gute Möglichkeiten, Kunden gezielt anzusprechen. Offensichtlich ist hier der Preis in höherem Maße für die Kaufentscheidung ausschlaggebend, als beim klassischen Ladenverkauf [Be92], [GKT07]. Der Aufbau eines treuen Kundenstamms fällt reinen Webshops eher schwer [Th06].

2.3 Vor- und Nachteile der Kanäle bezogen auf die Verkaufsphasen

Auf der Basis der vorgestellten Verkaufskanäle, Ladenverkauf und eCommerce, werden nachfolgend die allgemeinen Vor- und Nachteile je Kanal bezogen auf die einzelnen Verkaufsphasen [Me02] systematisiert. Diese allgemeinen Vor- und Nachteile werden im Rahmen der Feldstudie branchenbezogen betrachtet. Darauf aufbauend wird versucht, die positiven Aspekte beider Vertriebswege zu integrieren, indem traditioneller Verkauf und eCommerce miteinander kombiniert werden. Diese Kombination wird nachfolgend als Blended Shopping eingeführt.

TRANS-AKTIONS-PHASE	TRADITIONELLER VERKAUF	eCOMMERCE
Information 	+ Verkäufer kann Einfluss auf die Kaufentscheidung nehmen + Anfassen / Ausprobieren möglich + Beratung möglich + Verkaufserlebnis - aufwändige Informationssuche - personalintensiv für den Verkäufer - kostenintensiv (Verkaufsraum)	+ hohe Markttransparenz + einfache / objektive Informationssammlung + sehr strukturiert + keine personalintensive Beratung durch den Verkäufer nötig + kein Raum nötig + user generated content-Beratung - kein Anfassen / Ausprobieren möglich - keine klassische Beratung - kein kollektives Erlebnis - Kundenpräferenzen schwer erfassbar
Vermittlung 	+ Anfassen / Ausprobieren möglich + Beratung möglich - schwer zu strukturierende Informationen	+ breite Datenbasis steht dem Kunden zur Verfügung + Beratung durch user generated content - kein Anfassen / Ausprobieren möglich - keine klassische Beratung
Verhandlung 	+ persönliche Interaktion möglich + Spielraum für individuelle Lösungen - personalintensiv / zeitintensiv	+ sehr standardisiert, dadurch effizient - wenig Spielraum für individuelle Lösungen - keine persönliche Interaktion möglich
Vertragsabschluss 	+ für den Kunden einfach + Teil der Lebenserfahrung eines großen Kundenkreises	+ digitale Prozesse + weniger Medienbrüche - nicht standardisiert - für den Kunden aufwendig - untransparente Rechtslage - fehlendes Vertrauen in Datensicherheit, Sicherung der Privatsphäre
Abwicklung 	+ für den Verkäufer einfache Übergabe der Ware an den Käufer + Ware sofort verfügbar - Ware muss vom Verkäufer lokal vorrätig gehalten werden - Bezahlung erfordert Kassensystem - Kunde muss selbst für den Transport der Ware sorgen	+ Kein Kassensystem erforderlich + Kunde muss die Ware nicht transportieren + Lagerhaltung im Zentrallager + Etablierte Prozesse - Ware steht dem Kunden nicht sofort zur Verfügung - untransparente Rechtslage - fehlendes Vertrauen in Datensicherheit, Sicherung der Privatsphäre
After-Sales 	+ Möglichkeit zur persönlichen Nachbetreuung - separate Erhebung Kundendaten	+ Nutzung bereits digitalisierter Kundenkontakte + Nutzung digitalisierter Kundenerfahrungen (user generated content)

Tabelle 1: Vor- und Nachteile der Absatzkanäle bezogen auf die Verkaufsphasen

3 Blended Shopping

Im vorliegenden Artikel wird die These vertreten und nachfolgend belegt, dass bisher keine systematische Verknüpfung zwischen Ladenverkauf und eCommerce etabliert ist, auch wenn viele größere Handelsketten mittlerweile über beide Vertriebswege verfügen. Dabei können Kombinationen sowohl für den Händler als auch für den Konsumenten Vorteile haben.

Das sogenannte Multi-Channel-Verhalten der Konsumenten, also die gezielte Nutzung eines Absatzkanals zur Kaufvorbereitung oder –anbahnung in einem anderen Kanal, kann als grundsätzliche Bereitschaft, Verknüpfungen zwischen Verkaufskanälen zu nutzen, interpretiert werden. So wurde in 2008 z.B. für ca. $\frac{1}{4}$ der Käufe [BH08] der jeweils andere Kanal als kaufvorbereitende Informationsquelle genutzt. Mittlerweile befassen sich erste Ansätze mit einer erweiterten Form des realen Einkaufs, dem Augmented Shopping [KMO06], [LLO06], dieser Begriff ergibt sich jedoch direkt aus der Anwendung von Augmented Reality im Rahmen des Verkaufsprozesses und stellt somit unmittelbar auf die Nutzung von Virtual Reality-Technologien ab. Der Ansatz ist also eher technologiegetrieben. Der vorliegende Beitrag regt aber zur Entwicklung integrierter Konzepte zur Verknüpfung zweier bisher separater Verkaufskanäle und somit zur ganzheitlichen Betrachtung des Verkaufsprozesses an, weshalb hier der Begriff des Blended Shopping eingeführt wird. Unter Blended Shopping wird im Folgenden eine effiziente und sinnvolle flexible Verknüpfung von Prozessen verschiedener Absatzkanäle im Verkauf verstanden. Wir definieren Blended Shopping wie folgt:

DEFINITION 1: Unter Blended Shopping wird die Durchführung der Informations-, Vermittlungs-, Verhandlungs-, Vertragsabschluss-, Abwicklungs- und After-Sales-Phase unter Einbezug sowohl realer Verkaufs- und Präsentationsmechanismen als auch netzwerkgestützter Verkaufsfunktionalitäten verstanden.

Blended Shopping hat somit zwei Zielrichtungen. Einerseits geht es darum, Informationen und Funktionen, die im Internet für eCommerce etabliert sind, auch im Ladenlokal verfügbar zu machen und so den Kunden in der Kaufentscheidung zu unterstützen und dem Händler zu ermöglichen, Umsatz zu generieren. Andererseits sollen real existierende Produktpräsentationen, etwa im Schaufenster oder Showroom, genutzt werden, Produkte haptisch und optisch zu erleben und damit den „technischen Ansatz“ des eCommerce aufzulösen. Dabei ist zu beachten, dass wir hier mit mobilen Einsatzszenarien zu tun haben, da der Ladenverkauf für den Konsumenten durch Mobilität – entweder der Endgeräte oder der Benutzer [Ro05] - gekennzeichnet ist. Die vorgestellte Definition stellt die Basis für die im folgenden Kapitel beschriebene Feldstudie dar.

4 Feldstudie

In einer Feldstudie soll die Ausgangsthese des Artikels belegt und zusätzlich untersucht werden, inwieweit mit den existierenden Angeboten der Händler Blended Shopping realisiert werden kann. Für die Feldstudie sind also Unternehmen zu identifizieren, die Konsumgüter sowohl über klassischen Ladenverkauf als auch über eCommerce – als Voraussetzung für den Einsatz von Blended Shopping-Modellen – anbieten. Im Einzelhandel ist in Deutschland zu beobachten, dass sich Handelsketten gegenüber einzelnen Händlern vermehrt durchsetzen, um sich besser gegenüber Konkurrenten und anderen Marktteilnehmern behaupten zu können [KAS08].

Es ist davon auszugehen, dass Handelsketten am ehesten versuchen werden, auf neue Trends im Einzelhandel zu reagieren, da sie gegenüber einzelnen Händlern überregional oder sogar international agieren und die Entwicklung des gesamten Unternehmens im Fokus steht [KAS08]. Aus diesem Grund werden Handelsketten für die Feldstudie herangezogen. Die Filialen einer Handelskette sind in der Regel sehr ähnlich strukturiert, gestaltet und ausgestattet. Das bedeutet, dass die Ergebnisse eines Filialbesuchs relativ aussagekräftig sind in Bezug auf die grundsätzliche Ausstattung der anderen Filialen in Deutschland. Unterschiedliche Produkte und Dienstleistungen bedingen unterschiedliche Anforderungen an die einzelnen Verkaufsphasen [BH08] und die Gestaltung der Prozesse, deshalb erfolgt die Betrachtung branchenbezogen. Zur besseren Strukturierung und Übersichtlichkeit werden hier drei Branchen betrachtet. Zusätzlich wird eine separate Handelskette als Best-Practice-Beispiel für Blended Shopping einbezogen.

4.1 Branchenauswahl

Es wurden drei Branchen ausgewählt, in denen mehr als eine Handelskette in Deutschland vertreten sind, die Konsumgüter sowohl über ein bundesweites Filialnetz als auch über eCommerce anbieten. Das trifft für die Bereiche Buchhandel (Mayersche [Ma08c], Hugendubel [Hu08], Thalia [Th08]), Drogerien (dm Drogeriemarkt [Dm08], Rossmann [Ro08], Ihr Platz/Drosipa [Ih08], Schlecker [Sc08b]) und Kaufhäuser (Kaufhof [Ka08a], Karstadt [Ka08b]) zu. Zusätzlich zu diesen Branchen wurde der Real Future Store als Best Practice-Beispiel für moderne Verkaufskonzepte im Konsumgütereinzelhandel in die Untersuchung einbezogen. Im Real Future Store werden von der Metro Handelsgruppe Neuentwicklungen aus dem Metro Future Store am Markt getestet [Me08]. Insgesamt wurden zehn Handelsketten untersucht. Dabei wurden die Verkaufsphasen im eCommerce und im Ladenlokal anhand von jeweils drei unterschiedlichen Produkten untersucht, insgesamt also 60 Testkäufe durchgeführt.

4.2 Vorgehen

Die Erhebung wurde in drei Schritte durchgeführt. Im ersten Schritt wurden die Webshops der ausgewählten Handelsketten anhand der Verkaufsphasen [Me02] untersucht. Ausgehend von der Annahme, dass Informationen aus dem Internet beim realen Einkaufen den Kunden unterstützen können (z.B. Wegbeschreibung), wurde im zweiten Schritt die technische Verwendbarkeit untersucht. Dazu wurde etwa die Nutzbarkeit der Webshops auf mobilen Endgeräten evaluiert. Es wurde die URL der jeweiligen Handelskette auf einem Palm Treo Pro (Windows Mobile 6.1) und einem iPod Touch (iPhone) aufgerufen und die Navigierbarkeit in einer realen Einkaufssituation und die wahrgenommene Informationsqualität getestet. Mobile Geräte kämen im Ladenlokal zur Anwendung, um auf Informationen oder Funktionalitäten, die im Webshop zur Verfügung stehen, im Sinne des Blended Shopping auch im Ladenlokal zurück greifen zu können (Geräte mobilität [Ro05]). In einem solchen Szenario können auch weitere Informationen interessant sein, etwa genaue Standortangaben zu bestimmten Produkten, die bei der herkömmlichen Nutzung eines Webshops keine Relevanz besitzen.

Im dritten Schritt schließlich wurde untersucht, welche Unterstützung in den Ladenlokalen selbst durch stationäre Terminals zur Verfügung steht (Benutzermobilität [Ro05]). In den Filialen wurden die einzelnen Verkaufsphasen [Me02] durchlaufen und auf Möglichkeiten zur Vernetzung mit eCommerce auf mobilen Geräten bzw. stationären Terminals hin untersucht.

4.2 Qualitative Auswertung

Handelsketten bieten eCommerce in unterschiedlichen Ausprägungen – von reiner Produktpräsentation bis zu integrierten Online-Shops – an. Die für die Ausgangsthese relevanten Resultate der Untersuchung werden nachfolgend anhand der Verkaufsphasen (vgl. [Me02]) zusammen gefasst.

4.2.1 Information, Vermittlung, Verhandlung

Über alle Branchen verteilt stellen die Handelsketten mehrheitlich detaillierte Informationen sowohl zu ihrem Produkt- und Service-Sortiment als auch zu den Filialen im Internet zur Verfügung. Die Überprüfung der Warenverfügbarkeit bezieht (möglich bei 3 Buchhandlungen, 2 Drogerien, 1 Warenhaus) nicht die Bestände in den Filialen mit ein. Dementsprechend gibt es auch nicht die Möglichkeit, vor dem Besuch der Filiale zu überprüfen, ob die gewünschte Ware vorrätig ist oder sie zu reservieren, um sich unnötige Wege zu ersparen. Die filialbezogene Vorratsabfrage ist beispielsweise bei IKEA möglich [IK08]. Auch der Real Future Store nutzt diese Möglichkeiten bisher nicht. Allerdings kann hier ein Marktplan aufgerufen werden, der beim Einkauf selbst bei der Orientierung unterstützt (sonst noch in 1 Buchhandlung, 1 Warenhaus). Der Real Future Store bietet die Möglichkeit eine Einkaufsliste anzulegen. Allerdings führt das Scannen der Produkte, die in den virtuellen Warenkorb gelegt werden, um später automatisch abrechnen zu können, nicht zu einem Abarbeiten der vorher angelegten Einkaufsliste.

4.2.2 Vertragsabschluss, Abwicklung, After-Sales

Bestelloptionen im Webshop bieten alle untersuchten Anbieter an. Einige (2 Drogerien, 1 Warenhaus, Real Future Store) nur auf einen Teil des Sortiments beschränkt. Im Buchhandel und in den Drogerien (für einen Teil des Sortiments) gibt es bereits einen Verknüpfungsansatz zwischen Webshop und Filialnetz. Lieferungen bestellter Ware sind wahlweise an eine angegebene Adresse oder in die Filialen möglich. Vereinzelt kann auch die (Voraus-)Zahlung von Webshop-Bestellungen in den Filialen vorgenommen werden (2 Buchhandlungen, 1 Warenhaus). Hier gibt es Ansätze, die Prozesse beider Verkaufsformen nach den Anforderungen des Kunden auszuwählen. Unklar ist, inwieweit das Personal über eine entsprechende Informationspolitik der Konzerne eingebunden ist und diese Möglichkeiten gegenüber den Kunden kommuniziert. Im Bereich After-Sales ist die Warenrückgabe – sofern die Ware nicht personalisiert ist – bei allen Anbietern per Rücksendung möglich, allerdings werden auch hier mehrheitlich die Filialen als Anlaufstelle für Reklamationen nicht einbezogen.

Dabei würde dies die Möglichkeit eröffnen, z.B. Probleme bei der Nutzung zu beheben, ohne den ganzen Kaufprozess rückabwickeln zu müssen.

4.2.3 Mobile Nutzung des eCommerce-Angebots

Zur Einbindung von eCommerce-Prozessen in den Verkauf im Ladenlokal bietet sich neben Terminals, die in den Filialen vorgehalten werden müssen, die Nutzung mobiler Endgeräte (internetfähige Mobiltelefone, Smartphones, etc.) an [HMN03]. Die zur Verfügung stehende Bildschirmfläche ist im Vergleich zu stationären PC's deutlich kleiner und stellt besondere Anforderungen an die Gestaltung der abrufbaren Inhalte. Die Prüfung der Nutzbarkeit der Webshops ergab, dass keine der betrachteten Handelsketten über auf mobile Geräte angepasste Funktionalitäten verfügt. Die Webshops sind einerseits umständlich bis kaum navigierbar und andererseits fehlen zusätzliche Informationen, die direkt im Ladenlokal von Nutzen sein können (z.B. Übersicht über die Warenanordnung, Unterstützung bei der Produktsuche).

4.2.4 Ladenlokal

Im dritten Schritt wurden die Ladenlokale der Handelsketten besucht. Die zur Verfügung stehenden Terminals in der Filiale wurden aufgesucht und die gebotenen Funktionen ausprobiert. Da die Filialen einer Handelskette in der Regel gleich aufgebaut und ausgestattet sind, wurde jeweils eine Filiale einer Handelskette (Großraum Aachen) besucht. Die Untersuchung ergab, dass bisher kaum Informationsterminals eingesetzt werden und auch keine konsequente Einbindung der Webshops erfolgt. Bei der Drogeriekette dm ist im Bereich Gesundheitsprodukte ein Terminal im Einsatz, mit dem durch Einscannen eines Produkts Informationen eingeholt und vergleichbare Produkte aufgerufen werden können. Auch gibt es eine symptomgesteuerte Suchunterstützung mit entsprechender Beratungsleistung. Es findet keine Verknüpfung mit dem Webshop statt, bei dem beispielsweise die Bewertungen der Produkte von Stiftung Warentest hinterlegt sind. Auch ist ein Zugriff auf die Terminal-Informationen über das Internet nicht möglich. Das Terminal ist auf Gesundheitsprodukte beschränkt, für andere Produktgruppen gibt es keine Unterstützung. Seit verganginem Jahr stehen in einigen Filialen von dm auch Bestellterminals für Medikamente und Gesundheitsprodukte bereit. Hier können Produkte mit oder ohne Rezept bei einer kooperierenden Versandapotheke bestellt und in die Filiale oder an eine Wunschadresse geliefert werden. Die von Kaufhof und Karstadt beschriebenen Terminals konnten nur in wenigen Bereichen getestet werden, weil sie in den besuchten Märkten nicht vollumfänglich zur Verfügung standen. Die Möglichkeiten beschränkten sich bei Kaufhof auf den Abruf der Payback-Konten sowie das Einscannen von Produkten zur Preisabfrage. Eine weiterführende Nutzung wie z.B. die Möglichkeit, das eingescannte Produkt am Terminal zu bezahlen, ist bisher nicht implementiert. Davon unterscheidet sich der Real Future Store durch eine breite Palette von Terminals, die entweder zum Informationsabruf für eingescannte Produkte oder zur Analyseunterstützung etwa im Kosmetikbereich (Bestimmung des Hautbildes mit anschließender Empfehlung entsprechender Kosmetika) zur Verfügung stehen. Zur Einkaufsunterstützung befindet sich der Mobile Einkaufsassistent (MEA) in der Erprobungsphase. MEA funktioniert über ein Handy mit Kamera, über die man die Barcodes der Produkte einscannen und in einen virtuellen Warenkorb legen kann.

Der Warenkorb im Handy kann bearbeitet und erweitert werden, zu einigen Produkten sind weiterführende Informationen abrufbar. Der virtuelle Warenkorb kann zum Abschluss des Einkaufs einfach per individuell produziertem Barcode an eine entsprechende Kasse übergeben und abgerechnet werden. Die bereitstehenden Informations- und Analyseterminals sind bisher als Insellösungen nicht z.B. mit dem Warenkorb verknüpft. Die Ware wird separat eingescannt, um in die Schlussabrechnung aufgenommen zu werden. Weiterhin wurde getestet, ob das Filialpersonal den eigenen Webshop bereits in ihre Verkaufsaktivitäten integriert, etwa durch einen Verweis bei nicht vorrätigen Artikeln in der Filiale. In den Testfällen wurde vom Verkaufspersonal zwar angeboten, dass die Filiale die Ware ordert und zur Abholung bereit stellt (1 Buchhandlung, 1 Warenhaus). Eine Bestellung per hauseigenem eCommerce und damit Lieferung nach Hause wurde nicht angeboten. Aus den beschriebenen Ergebnissen lässt sich ableiten, dass Blended Shopping-Modelle bisher von acht Faktoren verhindert werden, die im folgenden Kapitel beschrieben werden.

5 Barrieren von Blended Shopping

Abgeleitet aus den Ergebnissen der Feldstudie wird im Folgenden die These aufgestellt, dass hauptsächlich acht Faktoren die Erschließung der Potentiale, die sich durch Blended Shopping-Modelle ergeben, verhindern. Diese dienen als Basis zur Entwicklung der im nächsten Kapitel vorgestellten Prototypen.

B1	<i>Technischer Zugang:</i> bisher sind viele Angebote nicht effizient mobil nutzbar, weil sie keine auf mobile Geräte angepassten Funktionalitäten bieten. Zwar können die Webshops auf mobilen Geräten aufgerufen werden, der kleinere Bildschirm und die veränderten Eingabemöglichkeiten erschweren die Nutzung deutlich und verhindern somit den effizienten Zugriff auf Informationen und Prozesse im Ladenlokal. Zusätzliche Funktionalitäten über das eigentliche Webangebot hinaus können den Einkauf erleichtern.
B2	Bewusst <i>getrennte Organisation:</i> die Absatzkanäle Filialnetz und Webshop sind organisatorisch voneinander getrennt. Z.B. ist häufig eine Rückgabe von per eCommerce bestellter Ware in den Filialen des gleichen Händlers nicht möglich ist.
B3	<i>Mitarbeiter werden nicht eingebunden:</i> durch die organisatorische Trennung der beiden Absatzkanäle werden die Mitarbeiter in den Filialen nicht trainiert, Absatzpotentiale zu erkennen, die sich aus einer Verbindung mit eCommerce-Prozessen ergeben.
B4	<i>Navigation in Kategorien:</i> das Angebot im Webshop ist nach bestimmten Kategorien geordnet, diese Navigation entspricht aber nicht den Suchbedürfnissen im Ladenlokal und zieht keinen Nutzen aus der tatsächlich vorhandenen Ware (z.B. Produktcodes).
B5	Unterschiedliche Strategien hinsichtlich <i>Preistransparenz:</i> im Netz herrscht aufgrund der leichten Zugänglichkeit von Informationen ein sehr hoher Grad an Preistransparenz. Im Ladenlokal wird Preistransparenz tendenziell eher vermieden. Dabei wird übersehen, dass die Kenntnis der Alternativen auch schnelle Kaufentscheidungen fördern kann (jetzt kaufen und sofort mitnehmen für diesen Preis, bestellen mit Lieferfrist für jenen Preis).

B6	<i>Fehlender Zugriff auf Datenbasis:</i> Auf die breite Datenbasis, die im Internet zur Verfügung steht (z.B. user generated content), hat der Kunde bisher im Ladenlokal keinen Zugriff. Wäre eine Abfrage dieser Informationen vor Ort möglich, könnte dies eine schnelle Kaufentscheidung unterstützen.
B7	<i>Fehlende digitale Handlungsanweisungen,</i> die im Internet zur Verfügung stehen und in der Filiale abgerufen werden können: ein Vorteil in der Informationsphase beim Ladenverkauf stellt die haptische Prüfung der Ware dar. Dieser Vorteil kann noch durch die Beratung des Verkäufers gesteigert werden. Die Anzahl der erklärungsbedürftigen Produkte nimmt weiter zu, die Anzahl der beratenden Verkäufer nimmt tendenziell aufgrund des Kostendrucks aber eher ab. Stünden digitale Handlungsanweisungen zur Verfügung, könnte der Beratungsvorteil trotzdem genutzt werden.
B8	<i>Planvoller Einkauf</i> wird nicht unterstützt: bei einigen Webshops ist es möglich, eine Einkaufsliste anzulegen, die in eine Bestellung überführt werden kann. Diese Prozesse stehen nur für Webshops zur Verfügung. Selbst das Best-Practice-Beispiel Real Future Store bietet bisher keine integrierte Lösung an. Zwar ist es möglich, eine Einkaufsliste für den mobilen Einkaufsassistenten (MEA) anzulegen, diese ist aber nur ein Textdokument, automatisiertes abhaken durch das Einscannen eines Artikels ist nicht möglich.

Tabelle 2: Barrieren von Blended Shopping

6 Prototypen

Aus den in Kapitel 5 vorgestellten Barrieren wurden Szenarien entwickelt und Studierenden mit der Aufgabenstellung vorgelegt, auf dieser Basis benutzerzentrierte Lösungen [Ri07] zu entwickeln. Im Folgenden soll anhand von zwei im Rahmen dieser Aufgabenstellung entstandenen Prototypen gezeigt werden, wie die ermittelten Barrieren überwunden werden können. Beide Prototypen wurden drei Vertretern des Handels präsentiert und anschließend von ihnen bewertet.

6.1 Prototyp 1: Schaufensterverkauf

Der erste Prototyp geht davon aus, dass ein Kunde außerhalb der Ladenöffnungszeiten vor einem Schaufenster steht und von der Auslage zum Kauf animiert wird. Nun könnte man unmittelbar und damit mobil die in der Ausgangstheese angenommenen eCommerce-Angebote nutzen, um das Produkt im Internet zu bestellen und den damit verbundenen Prozess anzustoßen. Von den oben aufgeführten Barrieren finden sich dabei vor allem die fehlende Interaktion mit der ausgestellten Ware (B4) sowie der nicht gegebenen Optimierung für ein Device (B1). In einem ersten Ansatz wurde versucht, die Auslage mit QR-Codes [De08], [BA07] auszustatten und damit mobil einen Eingang in die Shopping-Funktionalität zu finden. Dieser Ansatz wurde nach kurzen Tests verworfen, weil die optischen Bedingungen (Reflektionen) den Einsatz vor dem Schaufenster schwierig gestalteten. Für den angeführten Prototypen wurde hinsichtlich des letzten Punktes konsequent eine Touchscreen-Interaktion gestaltet. Der Touchscreen wird dabei stationär am Schaufenster eingerichtet.

Als Navigationskonzept wurde die Anordnung der Produkte im Schaufenster gewählt und in einer Flash Applikation dem Konsumenten präsentiert. Der Kunde kann die gewählten Produkte direkt bestellen. Neben dem eigentlichen Produktkauf kann der Kunde sich mit Hilfe des Terminals auch eine Erinnerungsmail senden, um später den Einkauf digital am Rechner fortsetzen zu können.

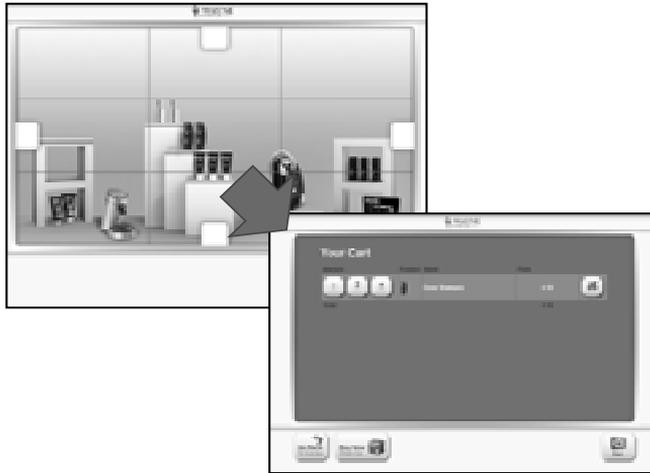


Abbildung 1: Anwendung an einem Schaufenster

Der Prototyp belegt, dass hier weitestgehend auf Standard-Komponenten eines PHP-basierten Webshop zurückgegriffen werden konnte und lediglich die Navigation entsprechend des Nutzungskontextes (Auslage im Schaufenster) gestaltet werden musste.

6.2 Prototyp 2: Web 2.0 im Ladenlokal

In einem zweiten Szenario wird davon ausgegangen, dass ein Konsumgütergeschäft zurzeit primär von den Konsumenten genutzt wird, um einen haptischen Eindruck von High-Tech-Produkten (etwa MP3-Player) zu bekommen. Angenommen wird, dass der eigentliche Verkaufsprozess dann im Internet stattfindet, wo Kunden die preisgünstigeren Angebote vermuten [EH08], [Ma08b]. Barrieren, das Geschäft im Ladenlokal abzuschließen sind demnach die fehlende Preistransparenz (B5), der fehlende Zugriff auf user generated content (B6) sowie der fehlenden Unterstützung eines planvollen Einkaufs (B8).

Hierzu wurde eine iPhone-Applikation entwickelt, die es erlaubt, den Einkauf im Ladenlokal vorzubereiten, auf Testimonials zuzugreifen und schließlich sich im Laden zu versichern, dass das Produkt zwar (unwesentlich) teurer ist als „im Internet“ dafür aber sofort verfügbar. Weiterhin können im Ladenlokal Reviews erstellt und eingesehen werden.



Abbildung 2: iPhone-Applikation für „Web2.0 im Shop“

Im Rahmen einer Evaluation in Form eines Expertenworkshops stellte sich heraus, dass diese Lösung sowohl für den Kunden als Komfortgewinn und für den Anbieter ökonomisches Potential bieten würde. Dies ist künftig noch genauer zu untersuchen. Insbesondere die Schaffung der Preistransparenz war allerdings argumentationsbedürftig, wurde letztendlich aber akzeptiert, da in der Applikation nur gezeigt wird, wenn im Internet ein geringfügig günstigeres Angebot (Grenze vom Händler wählbar) gefunden wird. Weiterhin wurde angeführt, dass die Mitarbeiter intensiv auf die Situation vorbereitet werden müssen, dass Kunden gegenüber dem Verkaufspersonal einen Informationsvorsprung haben (vgl. Barriere B3).

7 Fazit und Ausblick

Der vorliegende Artikel belegt, dass zurzeit eCommerce und der Verkauf im Ladenlokal als getrennte Prozesse gesehen werden. Gleichzeitig konnte gezeigt werden, dass eine Integration, insbesondere das Anreichern von realen Verkaufsszenarien um Internetmechanismen, vielversprechend scheint. Dies wurde als Blended Shopping eingeführt. Es wurde im Folgenden untersucht, ob sich existierende Internetangebote eignen, um in Blended Shopping-Szenarien eingebracht zu werden und es wurden Barrieren identifiziert, die diesen Einsatz unmöglich machen bzw. erschweren. Schließlich wurde anhand von Softwareanwendungen gezeigt, dass diese Barrieren, da nun bekannt, gezielt adressiert und abgestellt werden können.

Um den Nutzen von Blended Shopping zukünftig genauer quantifizieren zu können, wird aktuell an einer Methode zur Kosten-/Nutzenanalyse (vgl. [FR08]) von Blended Shopping-Angeboten gearbeitet. Weiterhin wird erarbeitet, in wie fern Standardwebshops bzw. Frameworks (vgl. etwa [Ma08a]) und die dort implementierten APIs erweitert werden müssen, um die identifizierten technischen Barrieren zukünftig effektiv abstellen zu können.

Literaturverzeichnis

- [AK07] Ahlert, D.; Kenning, P.: Handelsmarketing. Grundlagen der marktorientierten Führung von Handelsbetrieben. Springer-Verlag, Berlin, Heidelberg, 2007.
- [AR08]: ARD-/ZDF-Onlinestudie 2008. <http://www.ard-zdf-onlinestudie.de/>, 24.11.2008.
- [BA07] BARCODAT: 2D-Code-Fibel, Dornstetten, 2007.
- [Be92] Bender, D.: 5., überarb. Aufl. 1992. Vahlen, München, 1992.
- [BH08] Baal, v. S.; Hudetz, K.: Das Multi-Channel-Verhalten der Konsumenten. In: Mehrkanalsystemen des Handels. Ausgewählte Studien des ECC Handel Band 20, Institut für Handelsforschung an der Universität Köln, 2008.
- [BWW04] Baier, T.; Weinreich, H.; Wollenweber, F.: Verbesserung von Social Navigation durch Identitätsmanagement. Oldenbourg Verlag, München, 2004.
- [De08] Detecon International GmbH: Mobile Tagging. <http://www.detecon.com/en>, 24.11.2008.
- [Dm08] dm Drogeriemarkt: Homepage. <http://www.dm-drogeriemarkt.de/>, 17.11.2008.
- [EH08] EHI Retail Institute GmbH: Handel aktuell 2008/2009, Köln, 2008.
- [FR08] Fuchs, B.; Ritz, T.: Kosten-Nutzen-Bewertungen bei mobilen Anwendungen. In: Breitner et. al (Hrsg.): 3. Konferenz Mobile und Ubiquitäre Informationssysteme, 2008
- [GKT07] Gerhards, M.; Klingler, W.; Trump, T.: Web 2.0 - Eine Begriffsdefinition und eine Analyse der Auswirkungen auf das allgemeine Mediennutzungsverhalten. <http://www.competence-site.de/business-networking-web20.nsf/3c26e7f5f24138ac125691800380650/717735902a681010c125738d00417615!OpenDocument>, 24.11.2008.
- [HMN03] Hansemann, U.; Merk, L.; Nicklous, M. S.: Pervasive Computing. Springer-Verlag, Berlin, 2003.
- [Hu08] Hugendubel: Homepage. <http://www.hugendubel.de/>, 21.11.2008.
- [Ih08] Ihr Platz/Drospa: Homepage. <http://www.ihrplatz.de/>, 24.11.2008.
- [IK08] IKEA: Homepage. www.ikea.de, 24.11.2008.
- [In07] Internet World Stats: Homepage. www.internetworldstats.com, 17.11.2008.
- [Ka08a] Kaufhof: Homepage. <http://www.galeria-kaufhof.de/> 21.11.2008.
- [Ka08b] Karstadt: Homepage. <http://www.karstadt.de/>, 21.11.2008.
- [KAS08] Kotler, P. et al.: Grundlagen des Marketing. Pearson Studium, München, 2008.
- [KKB07] Kotler, P.; Keller, K. L.; Bliemel, F.: Marketing-Management. Strategien für wertschaffendes Handeln. Pearson Studium, München, 2007.
- [KMO06] Kato, T.; Miyazaki, T.; Ogino, A.: Framework for Modelling User Interest using Augmented Shopping Store, Taipei, 2006.
- [KW08] Kroeber-Riel, W.; Weinberg, P.: Konsumentenverhalten, Vahlen, München, 2008.
- [LLO06] Lee, J.-H. et al.: Design of the PromoPad: An Automated Augmented Reality Shopping Assistant. <http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/4273787/>

4273788/04273855.pdf?isnumber=4273788&prod=CNF&arnumber=4273855&arSt=358&ared=363&arAuthor=Ogino%2C+A.%3B+Miyazaki%2C+T.%3B+Kato%2C+T, 2006.

- [Ma08a] Magento: Magento API.
http://www.magentocommerce.com/support/magento_core_api, 21.11.2008.
- [Ma08b] Martens, P.: Freund oder Feind? Handel versus Online-Handel.
<http://diepresse.com/home/techscience/hightech/homeentertainment/421467>, 10.10.2008.
- [Ma08c] Mayersche Buchhandlung: Homepage. www.mayersche.de, 21.11.2008.
- [MB08] Meffert, H.; Burmann, C.: Marketing. Grundlagen marktorientierter Unternehmensführung ; Konzepte - Instrumente - Praxisbeispiele. Verlag Gabler, Wiesbaden, 2008.
- [Me02] Merz, M.: E-Commerce und E-Business. Marktmodelle, Anwendungen und Technologien. dpunkt-Verl., Heidelberg, 2002.
- [Me08] Metro Group: Homepage. <http://www.future-store.org/fsi-internet/html/de/375/index.html>, 18.11.2008.
- [Mo08] Mobile Fachgruppe im BVDE: Mobile Kompass 2008, Düsseldorf, 2008.
- [Or05] O'Reilly, T.: What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>, 2005
- [Ri07] Ritz, T.: Die benutzerzentrierte Entwicklung mobiler Unternehmenssoftware. In: König-Ries, Lehner, Malaka, Türker (Hrsg.): MMS 2007: Mobilität und mobile Informationssysteme - 2nd conference of GI-Fachgruppe MMS; 2007
- [Ro05] Roth, J.: Mobile Computing. Grundlagen, Technik, Konzepte. dpunkt-Verl., Heidelberg, 2005.
- [Ro07] Roik, O.: E-Commerce-Umsatz 2008: HDE erwartet 20 Milliarden Euro.
<http://www.einzelhandel.de/servlet/PB/menu/1077365/index.html>, 25.11.2008.
- [Ro08] Rossmann: Homepage. <http://www.rossmann.de>, 17.11.2008.
- [Sc08] Schmidt, H.: Marketing 2.0. In Frankfurter Allgemeine Zeitung, 2008.
- [Sc08b] Schlecker: Homepage. www.schlecker.de, 24.11.2008.
- [St08] Stumm, K.: Entdecke die Unwägbarkeiten. In manager-mMagazin.de, 2008.
- [Th06] Theis, H.-J.: Erfolgreiche Strategien und Instrumente im E-Commerce. Dt. Fachverl., Frankfurt am Main, 2006.
- [Th08] Thalia: Homepage.
http://www.thalia.de/shop/tha_homestartseite/show?jsessionid=fdc-f7dqvedb391.tc2, 21.11.2008.
- [Th99] Theis, H.-J.: Handels-Marketing. Analyse- und Planungskonzepte für den Einzelhandel. Deutscher Fachverl., Frankfurt/Main, 1999.

Automotive Software und Service Engineering (ASSE) – Eine Exploration von Herausforderungen und Trends aus Sicht von Branchenexperten

Andreas Zauner, Holger Hoffmann, Jan Marco Leimeister, Helmut Krcmar
Lehrstuhl für Wirtschaftsinformatik
Technische Universität München
Boltzmannstraße 3
85748 Garching bei München
zauner@in.tum.de; holger.hoffmann@in.tum.de
leimeister@in.tum.de; krcmar@in.tum.de

Abstract: *Automotive Software und Services (ASS) sind zentrale Innovationsfaktoren der Automobilindustrie und gewinnen zunehmend an Bedeutung. In diesem Paper beschreiben wir die aktuellen Herausforderungen bei ihrer Erstellung sowie Entwicklungstrends in der Domäne, jeweils aus Sicht der wesentlichen Anspruchsgruppen in der Industrie auf Basis von 21 Experteninterviews. Dazu geben wir einen Überblick über bestehende Automotive Software und beschreiben unser methodisches Vorgehen bei der Datenerhebung und Datenauswertung. Anschließend werden aktuelle Herausforderungen und Trends für die Zukunft anhand der gewonnenen Daten identifiziert. Als Fazit leiten wir vier zentrale Architekturbausteine als Grundlage für die Erstellung von Automotive Software und Services ab.*

1. Einführung und Motivation

Mobile softwarebasierte Mehrwertapplikationen im Automobil, sogenannte Automotive Software und Services [RKR07] gewinnen als Differenzierungsmerkmal bei einer „jungen Generation“ [So08] von Kunden aktuell immer mehr an Bedeutung. Neben der Eigenschaft der Automotive Software und Services als ein wesentliches Differenzierungsmerkmal, sollen sie es den OEMs auch erlauben den Kunden länger zu binden [Me04; RKR07]. Ursache dafür ist, dass 80% der Innovationen rund um das Automobil zukünftig softwarebasiert sein werden und sich die OEMs nur noch hier wirklich unterscheiden können [Me04]. Auch der Anteil von Software im Auto, der 2003 noch einen Umfang von 60MB hatte [Sa03], wird bis 2010 einen Umfang von 1 Gigabyte haben [Pr07]. Die Automobilhersteller befinden sich in einem Wandlungsprozess von reinen Produzenten, die den letzten Kundenkontakt bei Abholung des Fahrzeugs hatten, hin zu Anbietern holistischer Lösungen bei denen der Kunde über die komplette Lebenszeit seines Fahrzeuges mit unterschiedlichen Dienstleistungen versorgt wird [Me04]. In diesem Paper werden ausschließlich Automotive Software und Services betrachtet, die sich im tertiären Aufgabenbereich des Fahrzeuges [Bu03; Eh02] befinden und durch Fahrerassistenzsysteme teilweise in den sekundären Aufgabenbereich hineinreichen. Themen der embedded Software, beispielsweise Steuergeräte, werden explizit nicht betrachtet.

Drahtlose Breitbandtechnologien wie UMTS und WiMAX und breit verfügbare WLAN Netzwerke ermöglichen neue mobile Anwendungen und wecken so die Wünsche nach und Erwartungen an mobilen Anwendungen. Neben mobilen Endgeräten, wie Mobiltelefonen, rückt dabei nun das Auto in den Mittelpunkt. Ein Westeuropäer ist durchschnittlich täglich über 80 Minuten mobil, den Großteil dieser Zeit verbringt er in seinem Fahrzeug [SB06]. Fest verbaute Navigationslösungen mit GPS, die mit der Lokalisierung einen ortsunabhängigen Dienst bereitstellen, gehören in vielen Fahrzeugen zum Standard und werden mit Infotainmentanwendungen erweitert [Eh02]. Außerdem werden die Anwendung der Verkehrstelematik¹ [NM78] eingesetzt, wie beispielsweise der emergencyCall (eCall) bei Unfällen sowie Anwendungen, die über WLAN Daten mit anderen Fahrzeugen oder der Infrastruktur austauschen (Car-2-Car bzw. Car-2-Infrastructure Kommunikation). Zukünftig sollen Fahrzeuge im Sinne eines „Connected Life“ [So08] jederzeit und ortsunabhängig mit dem Internet verbunden sein und sämtliche webbasierten Dienste unterstützen, von der Theaterreservierung bis hin zur Bereitstellung der aktuellen (Sport-)Nachrichten. Automotive Software und Services sind *die* Innovationsfaktoren für alle Anspruchsgruppen in der Automobilbranche: „80% aller Innovation werden aus Software bestehen.“ [Me04].

Auf den folgenden Seiten geben wir nach einer Einführung in grundlegende Begrifflichkeiten und Zusammenhänge einen Überblick über Trends und zukünftige Entwicklungen im Bereich Automotive Software und Services. Hierfür wird zunächst auf das empirische Untersuchungsdesign eingegangen, die Datensammlung beschrieben und anschließend die Ergebnisse kondensiert dargestellt. Auf der Basis der Ergebnisse wird abschließend eine Rahmenarchitektur mit Architekturbausteinen aus den Zukunftstrends erarbeitet, die auch als Handlungsempfehlung für Wissenschaft und Industrie dienen soll.

2. Grundlagen und Struktur der Expertenbefragung

Ziel der Befragung ist es, den aktuellen Stand der Automotive Software und Services in Wissenschaft und Praxis zu identifizieren, Herausforderungen und Hindernisse zu ermitteln und Trends und Strategien für die Zukunft zu erkennen. Die Expertenbefragung wurde bewusst als Methode gewählt, da keine Meinungen durch eine Diskussion beeinflusst werden sollten, wie es bei der Delphi-Methode der Fall ist. In den folgenden Abschnitten stellen wir die Grundlagen unserer Expertenbefragung vor. Das Design und die Durchführung der Expertenbefragung basiert auf dem Modell der Phasen der Datenerhebung von Nieschlag [NDH02], ein angesehenes Modell in der deutschen Sozialwissenschaft. Das Modell wurde in dieser Studie so angepasst und angewandt, damit es die Forschungsfragen und Forschungsziele umfassend beantworten kann. In Abb. 1 ist dieses Modell dargestellt.

2.1 Methodische Grundlagen der Expertenbefragung

Die Expertenbefragung ist eine Methode der empirischen Sozialforschung, um Wissen über ein bestimmtes Themenfeld von einem „key informant“ [Ru05] zu erlangen.

¹ „Telematik“ bezeichnet die Kombinationen von Telekommunikation und Informatik

Hierbei wird erwartet, dass der „key informant“ auch Informationen über das Wissen und die Bestrebungen anderer Player in seinem Aufgabenbereich hat, diese einordnen und bewerten kann. Das Interview wird als ein *semi-strukturiertes Leitfadengespräch* [GL04] geführt. Dies bedeutet, dass nur wenige Fragen vor Anfang des Interviews vorbereitet sind und der Großteil der Fragen als Reaktion auf die Antworten des Interviewpartners formuliert wird. Diese Art des Interviews ermöglicht eine sehr hohe Flexibilität, da weder „Frageformulierung noch die Reihenfolge der Fragen verbindlich [sind].“ [GL04]. Außerdem werden die Experten mit Aussagen anderer Experten konfrontiert und werden somit in eine Richtung gedrängt, die sie dann im Kontext ihres Unternehmens weiter ausführen. Ziel des Leitfadengesprächs ist es, die Gedanken und Handlungen der verschiedenen Ansprechpartner zu identifizieren und sich einen Eindruck über deren Strategien und deren Bestrebungen für die Zukunft zu machen.

2.2 Aufbau und Zielsetzung des Gesprächsleitfadens

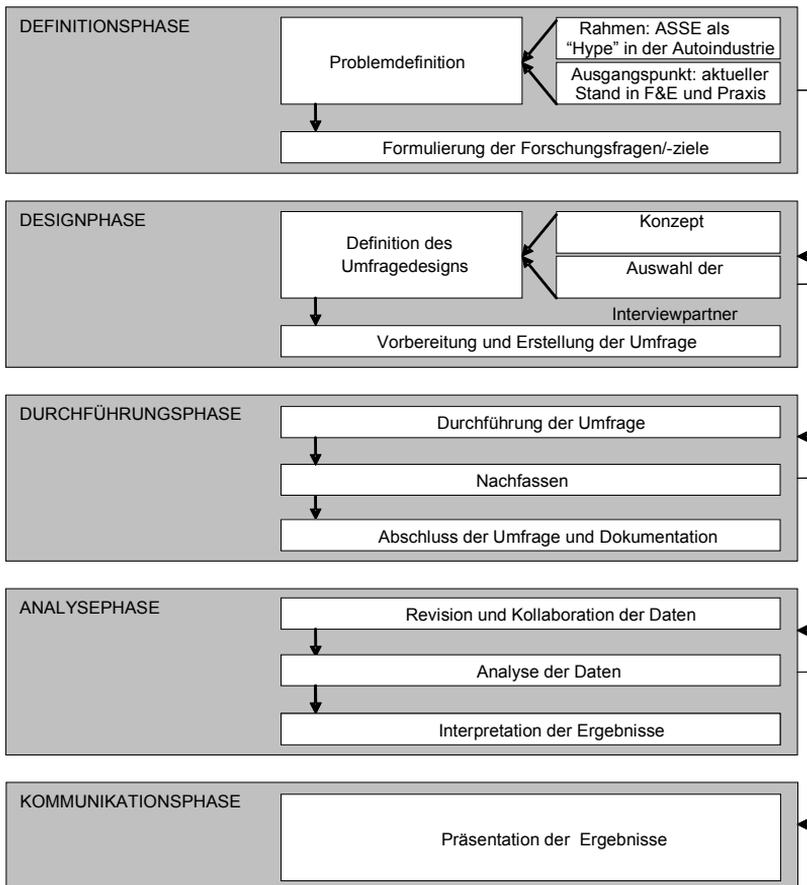


Abbildung 1: Umfragedurchführung und Prozess

Quelle: in Anlehnung an [LKK07; NDH02]

Durch ausgiebige Literaturrecherche und Internetrecherche wurden zunächst wichtige Wissensgrundlagen zu Automotive Software und Service Engineering geschaffen. Aufbauend darauf wurden die verschiedenen Gruppen auf Basis von Veröffentlichungen (Papern, Presseberichten etc.) und öffentlich zugänglichen Informationen (Bewertungen von anderen Experten, eigene Bewertung der Unternehmen) miteinander verglichen und erste Unterschiede in deren Strategien und Vorgehensweisen identifiziert. Diese Erkenntnisse wurden mit wissenschaftlicher Literatur [Bu03; Me04; RKR07; RMF02] verglichen, sodass ein Gesprächsleitfaden erstellt werden konnte, der die aus der Theorie generierten Thesen in der Befragung validiert. Bewusst wurde auf eine sehr allgemeine Fragestellung geachtet, um Aussagen über die Bedeutung der, zum Teil bereits vorhandenen, Ansatzpunkte in der Praxis zu erlangen. Die Expertenbefragung soll als Mittel zur Evaluation der in der Literatur dokumentierten Ergebnisse dienen. Neben einleitenden Fragen, über die Position und die Aufgabengebiete des Interviewpartners, sowie einer kurzen Erläuterung des Themas, wird der wesentliche Teil des Interviews in drei Themenblöcke gegliedert.

Der erste Themenblock beschäftigt sich mit Fragestellungen zur *Definition und Kategorisierung* der Automotive Software und Services im jeweiligen Arbeitgeber. Im zweiten Themenblock sollen Erkenntnisse über die *Konkurrenz* und die *Wettbewerber* der Befragten gewonnen werden, indem zunächst auf eigenen Stärken und Schwächen in der Entwicklung der Automotive Services fokussiert wurde und anschließend eine Einschätzung über die Wettbewerber abgegeben werden sollte. Zuletzt beschäftigt sich der dritte Themenblock mit der *Zukunft* der Automotive Software und Services. Hierzu wurden die Experten nach den relevanten/wichtigen/entscheidenden Chancen der (mobilen) Mehrwert-Applikationen in zukünftigen Diensten und in zukünftigen Kundenwünschen in 3, 5 und 10 Jahren gefragt. Nach der abschließenden Frage, ob etwas Wichtiges vergessen wurde, wurde das Interview nach dankenden Worten abgeschlossen.

2.3 Auswahl, Anschreiben und Rückmeldungen der Experten

Das Ziel der Expertenbefragung ist es, ein möglichst umfassendes Bild über die Automotive Software und Services zu gewinnen. Dazu müssen verschiedenste Ansprechpartner befragt werden. Diese Ansprechpartner werden bezüglich ihrer Rollen in der Industrie unterschieden: OEMs, Zulieferer, 3rd Party Organisationen (wie Telekommunikationsanbieter) und Verbände bzw. Interessensgemeinschaften (wie ADAC, VDI, VDA). Außerdem wird noch eine Gruppe Berater, die eine übergeordnete Sichtweise hat, in die Befragung integriert. Basierend auf der Kategorisierung in verschiedene Anspruchsgruppen wurden passende Ansprechpartner gesucht. Entsprechende Kontaktinformationen wurden von Unternehmenswebseiten, auf wissenschaftlichen Veröffentlichungen, Tagungsunterlagen etc. gesammelt und um bestehende Kontakte aus früheren Forschungsprojekten ergänzt. Ziel der Befragung in der ersten Ausbaustufe ist die Erfassung des deutschen Marktes, daher wurden keine Experten außerhalb Deutschlands erfasst. An die so abgeleiteten 60 Email Adressen wurden persönliche Einladungen verschickt. Die Anzahl der Zusagen von 20 Experten (33%) ist außergewöhnlich hoch. Normalerweise ist die Rücklaufquote bei einem solchen Vorgehen nicht höher als 13% [Ha98; Po99], wer eine Befragung „samt

freundlichen Anschreiben ohne weitere Maßnahmen verschickt, wird selten Rücklaufquoten über 20% erzielen.“ [Di95]. Die Steigerung der Quote um, je nach Quelle, 50% bis 150%, ist für uns ein deutliches Indiz für die Relevanz des Themas Automotive Software und Services in der Industrie. Die Zusagen der Experten verteilen sich gleichmäßig auf die Gruppen OEM, Tier-1 Zulieferer, 3rd Party, wie Telekommunikations- und Softwareunternehmen, Verbände und Berater.

2.4 Durchführung und Auswertung der Interviews

Die Interviews wurden im Juli 2008 telefonisch durchgeführt. Die Gespräche wurden, nach Zustimmung des Interviewpartners, mit Hilfe des Programms Total Recorder im Mp3-Format aufgezeichnet und anschließend transkribiert. Die Interviewpartner wurden dazu aufgefordert die Fragen offen und mit ihrer persönlichen Meinung zu beantworten. Das Datenanalyseverfahren basiert auf der Strukturierung der gleich gewichteten Expertenmeinungen, d.h. der Abgrenzung der jeweiligen Sichtweisen und Meinungsbildern und Gruppierung der gemeinsamen, invarianten Komponenten zu Themengebieten. Um aussagekräftige Meinungsbilder identifizieren zu können, wurden Stichworte (Items) anhand eines Code Books [NDH02] identifiziert und gruppiert. Anschließend wurden die so identifizierten Stichwort-Cluster näher beschrieben. Um eine objektive Auswahl und Gruppierung von Items zu gewährleisten (Intercoder Reliability) wurde dieser Schritt unabhängig von zwei Personen durchgeführt, abschließend wurden deren Ergebnisse konsolidiert [Ru05]. Für das Maß der Intercoder Reliability wurde mit Hilfe eines statistischen Maßes, Cohens Kappa, der Wert 0,81 berechnet [Lo02]. Ein derart hoher Wert ist laut [Ba00] absolut zufriedenstellend, er wird von Greve/Wentura [GW97] als ausgezeichnet und von Landis/Koch [LK77] als fast vollkommene Übereinstimmung deklariert. Damit können die im Rahmen dieser Befragung ermittelten Ergebnisse als objektive Abbildung der Meinungsbilder bezeichnet werden.

3. Resultate der Expertenbefragung

3.1 Herausforderungen/Obstacles

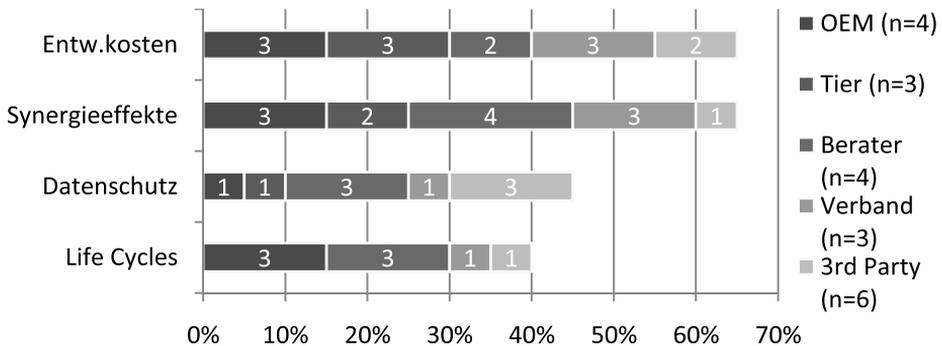


Abbildung 2: Herausforderungen – Häufigkeit der Nennungen gruppiert nach Anspruchsgruppen

Die Herausforderungen in Automotive Software und Services an die Automobilindustrie sind mannigfaltig. Die meist genannten Herausforderungen (Abb. 2) sind das Angleichen von Entwicklungskosten (68%), das Fehlen von Synergieeffekten (63%), der Datenschutz (44%) und die unterschiedlichen Life Cycles (38%).

Das Niveau von Kundennutzen und Entwicklungskosten bei Automotive Software und Services muss angeglichen werden. Bei Automotive Software und Services fallen sehr hohe Entwicklungskosten an. Ein neuer Dienst kostet die Unternehmen 60-70 Mio. Euro an Entwicklungsgeldern [Ta08]. Diese Summe ist eine hohe Investition für einen OEM oder Tier 1 Zulieferer und es muss erst geschafft werden, im Anschluss die 2.500-3.500 Euro teuren festeingebauten Geräte an den Kunden zu verkaufen [Ta08]. Die OEMs haben hier in Vergangenheit sehr viel Geld verloren, weil sie zwar neue Systeme entwickelt haben, diese jedoch nur in Fahrzeugen der oberen Mittelklasse und Oberklasse verkaufen konnten [Tc08]. Auch hier muss unterschieden werden, was der Kunde eigentlich will, um hohe Verkaufszahlen zu erzielen und somit die Entwicklungskosten zu decken. Ferner will kein OEM als erster neue Dienste entwickeln, da diese Dienste nur 3-5-Jahre geschützt werden können und dann die Konkurrenz dieselben Dienste ohne eigene Entwicklung(-kosten) anbieten kann [Oa08]. Die Entwicklungskosten können angeglichen werden, indem die Akzeptanz der Kunden [Tb08], diese neuen Services zu nutzen, steigt. Es muss zwischen Mehrwert und Erwartungshaltung des Kunden unterschieden werden. Während ein ESP im Auto keinen Mehrwert darstellt, ist ein Nachsichtassistent ein Aufpreis pflichtiges System, wobei der Kunde selbst und persönlich entscheidet ob er dieses System will oder nicht [FS03; Fu01; VC08; Wr05]. Die Unternehmen gehen davon aus, dass der Kunde nicht dazu bereit ist für solche Dienste Geld auszugeben [BeB08; Ta08; Tb08; Tc08]. Hier herrscht auch eine Diskrepanz zwischen dem, was der Kunde möchte und dem, was der Hersteller für einen vernünftigen Preis liefern kann [Oa08]. Der Hersteller steht vor dem Problem, „dass das gebaut werden muss, was der Kunde will, aber andererseits muss er Dinge anbieten, wovon der Kunde noch gar nicht weiß, dass es sie gibt und er sie braucht.“ [VC08]. Der Kunde sieht es als selbstverständlich an, dass Automotive Software und Services im Auto kostenlos angeboten werden, so wie es bei Radiogeräten jetzt schon ist [Oa08]. Falls der Kunde den Mehrwert dieser Dienste erkennt wird er auch bereit sein, dafür zu zahlen.

Eine weitere Herausforderungen ist die Datensicherheit und der Datenschutz im „Connected Car“ [So08; Tc08]. Vor allem im Bereich Car-2-Car und Car-2-Infrastructure „stellt sich dann die Frage, wer darf mit wem, welche Daten austauschen.“ [VB08]. „Diese Technologie darf kein verlängerter Arm des Gesetzes werden.“ [VB08]. Hier muss strikt darauf geachtet werden, dass Informationen über das Fahrverhalten, über die Geschwindigkeit etc. beim Fahrer bleiben und ausschließlich den Verkehrsteilnehmern zugänglich gemacht werden, die diese auch benötigen und auch berechtigt sind, diese zu verarbeiten. Wobei dieses Datenschutzproblem im europäischen Raum noch langer Diskussionen bedarf, ist es im asiatischen Markt schon gelöst. Dies ist jedoch ein gesellschaftliches Thema, wie mit Daten umgegangen wird [VC08]. „Das gute Grundkonzept [von C2X] könnte durch ganz einfache psychologische Aspekte ausgehebelt werden.“ [BeC08].

Ein großes Problem der OEMs und der Zulieferer sind die unterschiedlichen Product Life Cycles [Ha04; Mo06]. Es stehen sich die alteingesessenen Automobilhersteller mit Entwicklungszyklen von 7 Jahren und die junge Branche der Dienstentwickler mit einem Entwicklungszyklus von 6 Monaten gegenüber [BeC08; VB08]. Daraus ist offensichtlich, dass die traditionellen Hersteller kaum noch mit Unternehmen wie Telcos oder Drittanbieter im Navigationssektor mithalten können. Es werden technische Entwicklungen, die gut durchdacht, funktionstüchtig und qualitativ hochwertig sind von der weiteren Entwicklung einfach überrollt [BeC08], weil sie einfach zu lange Entwicklungszeiten haben. Die Automobilhersteller sind auch hier zu unflexibel, schotten sich sogar teilweise vom Markt ab und weigern sich Nachrüstlösungen funktionstüchtig ins Fahrzeug zu integrieren [BeC08; Oa08], obwohl sie besser als die eigenen sind.

Ein letztes Hindernis ist das Fehlen von Synergieeffekten durch firmenübergreifende Teams. Aus Sicht vieler Interviewpartner fehlt den OEMs im Gebiet der Automotive Software und Services die Kompetenz, die Effizienz und die Struktur, um innovative Automotive Software und Services gewinnbringend zu entwickeln [BeA08; VC08]. Hier hat es auch einen Paradigmenwechsel in der Automobilindustrie gegeben [Tb08]. In Vergangenheit lag das ganze Wissen über das Auto bei den OEMs und diese haben sich dann den Tier 1 Zulieferern geöffnet, wenn Hilfe von Nöten war. In der Gegenwart jedoch ist es anders, das Wissen über die Automotive Software und Services liegt zum Großteil bei den Tier 1 Zulieferern, den Telekommunikationsunternehmen und weiteren Drittanbietern, wie beispielsweise Softwareunternehmen und die OEMs sind auf diese Expertise angewiesen. Außerdem sind „viele Entwicklungen noch nicht zu Ende gedacht, auch vom Geschäftsmodell her und viele OEMs turnen noch im Experimentierstadium herum.“ [BeA08]. Viele OEMs haben noch nicht den Sinn von Softwareentwicklung verstanden. „Man behandelt Software genauso wie Blechteile.“ [BeA08]. „Diese Größenordnung an Softwareentwicklung [passt] in einen normalen Fahrzeugablauf schon lange nicht mehr hinein.“ [Ob08]. Die OEMs verfolgen hier seit vielen Jahren ein „Geschäftsmodell (...), an dem sie aber festhalten (...), das aber für neue Dienste nicht erfolgreich sein wird.“ [3rdA08]. Es werden zu viele Dinge vom OEM selbst gemacht, sie wollen ihren Marktanteil abschöpfen und sehen sich nicht als einen Teil des ganzen Konstrukts [3rdA08; BeA08]. „Die perfekten Lösungen, die die OEMs suchen gibt es nicht!“ [3rdA08].

3.2 Trends

In der Automobilindustrie gibt es eine Vielzahl von Trends bezüglich des Automotive Software & Service Engineerings. Die vier meist genannten Trends (**Fehler! Verweisquelle konnte nicht gefunden werden.**³) sind eine Automotive Service Plattform (70%), das Entwickeln neuer Vorgehensmodelle (65%), das Finden von Normen und Standards (60%) und die Killerapplikationen Sprit sparen durch Software (50%).

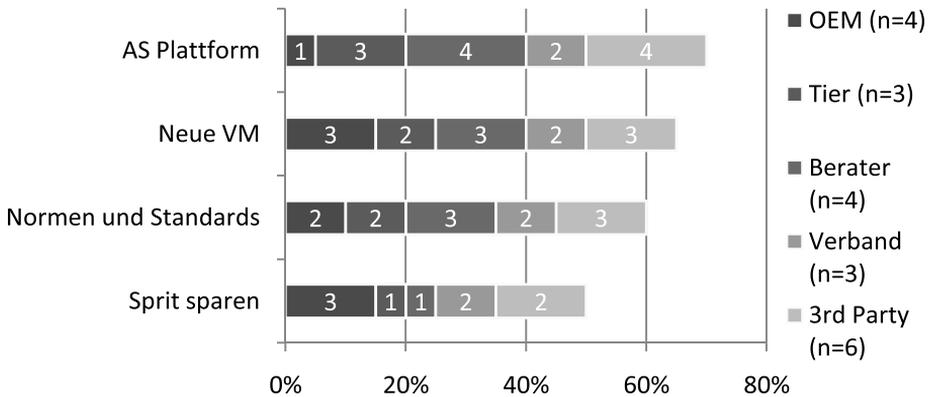


Abbildung 3: Entwicklungstrends – Häufigkeit der Nennungen gruppiert nach Anspruchsgruppen

Mit Hilfe einer einheitlichen Automotive Service Plattform könnten viele der Herausforderungen gemeistert werden. Jedoch herrscht in der Plattformfrage große Uneinigkeit unter den Experten. Drei mögliche Lösungen wurden genannt, die jede selbst und auch alleinstehend ein Schlüssel zum Erfolg sein könnte. Diese genannten Plattformen sind erstens ein telematisch basiertes PKW-Mautsystem [BeC08; Sp08; Ta08; Tb08; VA08], zweitens der digitale Rundfunk mit dem TPEG-Format [3rdB08; VB08] und drittens der eCall über UMTS [3rdA08; Sp08; St08; VB08]. Eine telematisch basiertes PKW-Maut System, würde zwangsweise dazu führen, dass in jedem neuen Fahrzeuge ein Modul zur Erfassung der Mautdaten integriert ist, wie es jetzt schon bei LKWs Standard ist. Problemlos können dann weitere Dienste auf diese Basis aufgesetzt werden und somit schnell einem breiten Markt zur Verfügung gestellt werden. [Ta08]. Dies wäre auch technisch ohne weiteres zu realisieren [Sp08]. Jedoch muss diese Thematik kritisch betrachtet werden und erst hinterfragt werden, ob die Einführung eine PKW-Maut in Deutschland politisch gewollt ist. Hierbei gehen die Meinungen sehr auseinander: Wobei die Zulieferer hier ihr Potential für die Zukunft sehen und dadurch auf hohe Einnahmen hoffen [Ta08; Tb08], stehen vor allem die Verbände und Verbraucherschützer dieser Entwicklung kritisch gegenüber [VA08; VB08]. Hier stellen sich viele politische Fragen, wer welche Daten, wie und für was nutzen darf, da „ein Mautsystem politisch ganz andere Ziele und Zwecke verfolgt, als die Kommunikation [wie C2X] untereinander.“ [VA08]. Dennoch kann ein Mautsystem eine Lösung sein, wenn es offene und standardisierte Schnittstellen für alle hat [3rdA08; BeC08]. Der digitale Rundfunk kann ein zweiter Weg sein. Jedoch ist es zur Zeit noch so, dass die Geräte, die derzeit im Auto sind (speziell die Radiogeräte) noch nicht mit dem neuen Format der Transport Protocol Experts Group (TPEG) kompatibel sind, sondern noch immer das Format Traffic Message Channel (TMC) als Grundlage verwenden. Durch TPEG können viel mehr Inhalte als mit TMC transportiert werden und auch eine Kommunikation zur Infrastruktur ist möglich [3rdB08]. Das Problem hierbei ist, dass zwar eine Plattform durch TPEG existiert [VB08], die auch technisch völlig ausgereift ist, jedoch keine Verbreitung am Markt findet. Die dritte Möglichkeit ist der eCall. Diese Plattform ist bei jedem Autofahrer, der ein Mobiles Device besitzt, schon automatisch im Auto präsent und darauf können neben dem Notruf auch weitere Dienste, basierend auf

UTMS, aufgesetzt werden. Auch über die GPS Lokalisierung können Mehrwertdienste realisiert werden [St08; Ta08]. Der eCall ist besonders für Telekommunikationsanbieter von Interesse wobei diese Plattform von den OEMs nicht favorisiert wird. Diese drei Möglichkeiten stehen zurzeit nebeneinander und werden unter den Experten diskutiert, wobei eine Relevanz von einer Automotive Service Plattform im Auto nicht zur Debatte steht, weil die Experten hier die Zukunft sehen. Eine sehr wahrscheinliche Lösung ist die mautbasierte Plattform, da hier schon viel Infrastruktur geschaffen wurde und das System auch eine sehr hohe Stabilität und Sicherheit bietet. Die Verfügbarkeit ist auch schon auf allen deutschen Autobahnen zu 100% abgedeckt und im zweiten Schritt kann diese Infrastruktur auch auf die Bundesstraßen und die Städte erweitert werden, wie es das japanische VICS-System in Tokio bereits ermöglicht [Vi08].

Ein weiterer Trend ist die Entwicklung von Vorgehensmodellen zur Gestaltung von Automotive Software und Services und die Entwicklung von Werkzeugen für die Vorgehensmodelle. Hier werden drei Themengebiete genannt. Zunächst müssen Konzepte des Softwareengineering in der Entwicklung von Automotive Software und Service Engineering berücksichtigt werden und Werkzeuge für die Entwicklung definiert werden [BeA08; BeC08; Oa08; Oc08]. Die Schwierigkeit besteht darin, dass ein Maschinenbauingenieur kein Softwareentwickler ist und sich alteingesessene Ingenieure weigern, neue Wege zu gehen. Das zweite Themengebiet ist der Paradigmenwechseln in der Automobilindustrie weg von der Produktorientierung „Auto“ hin zu einer ausgeprägten Kundenorientierung [BeC08; Ob08; Oc08; Ta08; VB08; VC08]. Es muss darauf eingegangen werden, was der Kunde wirklich wünscht und auch auf das, was er sich leisten kann. Die Automotive Software und Services dürfen nicht mehr als Produkt verkauft werden, sondern es müssen der Dienst und der Mehrwert für den Kunden im Vordergrund stehen. Es darf keine gesellschaftliche Diskussion darüber entstehen, wieso sich der eine, der es sich leisten kann, ein sichereres Auto bekommt, als der andere. Diese Sicherheitsdienste müssen jedem Kunden kostengünstig angeboten werden können und nur in anderen Mehrwertdiensten kann differenziert werden [3rdC08]. Das dritte Themengebiet beschäftigt sich mit der Integration der verschiedenen Beteiligten. Bei Automotive Software und Service Engineering müssen viele verschiedene Menschen zusammengebracht werden um ein erfolgreiches, neues System zu entwickeln [BeA08; BeC08; Ta08; VB08; VC08]. „Es muss über den Tellerrand hinaus geschaut werden und hier müssen Fahrzeugtechniker, Betriebswirte, Maschinenbauer und Softwaretechniker auch von Universitäten zusammenarbeiten.“[BeA08]. Die Kundenorientierung und die Integration von allen Beteiligten sind zum Teil schon realisiert, wobei sich die Einbindung von Software Engineering Konzepten noch sehr im Anfangsstadium befindet und noch viel Arbeit bedarf.

Ein wesentlicher Trend ist auch die Findung von Normen und Standards [3rdA08; 3rdB08; 3rdC08; 3rdD08; BeA08; BeB08; BeC08; BeD08; Ob08; Ta08; Tb08; VA08; VB08]. Es müssen einheitliche Architekturen gefunden werden, um zusammen diese Automotive Software und Services zu gestalten. Mit AUTOSAR gab es einen Versuch, der aber nur mehr oder weniger erfolgreich war [BeA08; Tb08]. Es besteht die Vision, den Entwicklern eine „einheitliche API zur Verfügung zustellen, mit der sich interaktive mobile Services im Auto [...] umsetzen lassen.“ [HLK07]. Es muss eine für alle

akzeptable Lösung gefunden werden, die dann die zukünftige Entwicklung besser, schneller und effizienter gestaltet.

„Spritsparen durch Software.“ Dieser Trend ist in Zeiten hoher Energie- und Ressourcenpreise absolut entscheidend [3rdA08; 3rdC08; BeB08; Oa08; Ob08; Tb08; VA08; VB08]. Dieser Trends ist nicht nur für die OEMs und Zulieferer von großer Bedeutung, sondern die Kunden werden nach diesem Merkmal beim Autokauf differenzieren. Applikationen, die zum einen den Rohstoffverbrauch verringern und gleichzeitig den CO2-Ausstoß auf das Minimum begrenzen werden als *die* Killerapplikationen für die Automotive Software und Services in naher Zukunft bezeichnet. Der Kunde kann hier für sich persönlich einen sehr hohen Mehrwert feststellen. Dadurch wird er auch gewillt sein, beim Autokauf am Anfang mehr für Software im Auto zu bezahlen. In diesen Trend spielen Themen wie autonomes Fahren, dynamische und ökologische Routenführung, sowie Fahrerassistenzsysteme, wie Start-Stopp-Funktion an der Ampel, ACC in der Kolonne, Überholassistent etc., eine entscheidende Rolle. In diesem Gebiet wird die Zukunft der Software im Automobil liegen, da das ganze Thema von alternativen Antrieben für Fahrzeuge noch lange nicht Realität sein wird [Oa08; Ob08]. Dieser Trend wird sehr stark von den Verbraucherschützern und Automobil Verbänden forciert und auch die OEMs selbst stellen fest, dass neue innovative Systeme, die einen hohen Energieverbrauch darstellen, auf dem Markt nicht den gewünschten Absatz finden.

In Abb. 4 werden die Zusammenhänge zwischen den Herausforderungen und den Trends dargestellt. Ein „x“ bedeutet, dass der jeweilige Trend eine Lösung für die Herausforderung sein kann. Es handelt sich hier nicht um eine Einordnung nach der Literatur, sondern es handelt sich hier um die Aussagen und Überzeugungen der befragten Experten, in welchem Bereich sie Ansatzpunkte zur Beseitigung der Hindernisse sehn. Die Trends werden gemäß ihrem gegenwärtigen Realisierungsgrad in der Praxis angeordnet. Die Herausforderungen sind angeordnet nach der Anzahl der möglichen beschriebenen Lösungsdimensionen durch die verschiedenen Trends.

		Realisierungsgrad			
		1	2	3	4
Lösungsdimensionen	I	x	x	x	x
	II		x	x	x
	III	x	x		x
	IV			x	x

Trends
 1: „Killerapplikation“ Benzin sparen
 2: Entwicklungstools & Vorgehensmodelle
 3: Automotive Services Plattform
 4: Normen und Standards

Herausforderungen
 I: Entwicklungskosten & Kundennutzen
 II: Life Cycles Auto vs. Software/Service
 III: Datensicherheit im vernetzten Fahrzeug
 IV: Synergieeffekte in der Entwicklung

Abbildung 4: Herausforderungen und deren Lösungsdimensionen durch die Trends

4. Implikationen und Ausblick

Das Thema Automotive und Service Engineering wird ein wesentliches Thema der Zukunft sein, an dem alle Anspruchsgruppen der Automobilindustrie, beginnend bei den OEMs, Zulieferer, über Telekommunikationsunternehmen bis hin zum Kunden, großes Interesse haben werden. Dies beweist die sehr hohe Response Rate der Expertenbefragung. Außerdem wollen wir die sehr offenen Interviewpartner erwähnen, die uns durch ihre sehr aufgeschlossene Art und sehr ehrliche, auch persönliche, Meinung erst die Grundlage für diese Studie geschaffen haben. Bei der von uns durchgeführten Studie handelt es sich um eine sehr auf Deutschland und den deutschen Markt fixierte Sicht. In der Zukunft würden wir diese Sicht gerne ausdehnen auf den internationalen Markt und dabei vor allem auch die Position wichtiger Vorreiter im Bereich der Automotive Services, z.B. in Asien (genauer: Japan), mit einbeziehen.

Die mit den unterschiedlichen Anspruchsgruppen geführten Interviews haben eindrücklich vor Augen geführt, dass die in einigen Studien gezeigte langfristige Bedeutung von mobilen Mehrwertapplikationen im Fahrzeug nicht übertrieben ist. Aktuell besitzen jedoch weder Automobilhersteller, noch deren Zulieferer oder externe Anbieter, die nötige Kompetenz, um die Gestaltung und Einführung von Automotive Services, zielgerichtet zu betreiben. Eine Kernkompetenz, die hierfür notwendig ist, stellt die Auswahl neuer Ideen, Konzepte und Technologien für Dienste im Bereich „Automotive“ dar. Ziel dieser Studie war es explizit die subjektiven Meinungsbilder hochrangiger Experten aus den verschiedenen Anspruchsgruppen der Automobilindustrie wiederzugeben und zu analysieren.

Aus unserer Studie können folgende Bausteine für das Automotive Software und Service Engineering abgeleitet werden (Abb. 5). Der Kooperationsgrad, der bei der Entwicklung von Nöten ist, wird umso größer, desto tiefer man in Richtung Normen und Standards gelangt. Außerdem ist die Sichtbarkeit für den Endkunden, den Autofahrer, erst bei den Killerapplikationen wirklich gegeben.



Abbildung 5: Bausteine für das Automotive Service und Software Engineering

Zielsetzung unserer weiteren Forschung ist es deshalb, gemeinsam mit einem Automobilhersteller ein werkzeuggestütztes Rahmenkonzept zum systematischen Prototyping von Automotive Services zu konzipieren und zu entwickeln. Das Rahmenkonzept selbst beschreibt die Schwerpunkte des Prototypings mit ihren jeweiligen Anspruchsgruppen. Das Werkzeug stellt Ingenieuren eine Möglichkeit zur Verfügung, interaktive mobile Anwendungen und Services im Auto prototypisch umzusetzen, um diese für das Rahmenkonzept im Fahrzeug pilotieren und evaluieren zu können. Sind diese Fragen erst beantwortet eröffnen sich viele zahlreiche Möglichkeiten innovative, zukunftsweisende Automotive Services in großem Stil und ohne hohe initiale Kosten umzusetzen und mit Kunden zu pilotieren.

Interviewverzeichnis

- [3rdA08] Experteninterview, Mitarbeiter der Firma 1, 2008.
- [3rdB08] Experteninterview, Mitarbeiter der Firma 2, 2008.
- [3rdC08] Experteninterview, Mitarbeiter der Firma 3, 2008.
- [3rdD08] Experteninterview, Mitarbeiter der Firma 4, 2008.
- [BeA08] Experteninterview, Mitarbeiter des Beratungsunternehmens 1, 2008.
- [BeB08] Experteninterview, Mitarbeiter des Beratungsunternehmens 2, 2008.
- [BeC08] Experteninterview, Mitarbeiter des Beratungsunternehmens 3, 2008.
- [BeD08] Experteninterview, Mitarbeiter des Beratungsunternehmens 4, 2008.
- [Oa08] Experteninterview, Mitarbeiter des Automobilhersteller 1, 2008.
- [Ob08] Experteninterview, Mitarbeiter des Automobilherstellers 2, 2008.
- [Oc08] Experteninterview, Mitarbeiter des Automobilherstellers 3, 2008.
- [Ta08] Experteninterview, Mitarbeiter der Zulieferfirma 1, 2008.
- [Tb08] Experteninterview, Mitarbeiter der Zulieferfirma 2, 2008.
- [Tc08] Experteninterview, Mitarbeiter der Zulieferfirma 3, 2008.
- [VA08] Experteninterview, Mitarbeiter des Industrieverbandes 1, 2008.
- [VB08] Experteninterview, Mitarbeiter des Industrieverbandes 2, 2008.
- [VC08] Experteninterview, Mitarbeiter des Industrieverbandes 3, 2008.

Literaturverzeichnis

- [Ba00] Bakeman, R. (2000). Behavioral observation and coding. In: Reis, T.; Judge, E.M. (Hrsg.), *Handbook of Research Methods in Social and Personality Psychology*. New York: Cambridge University Press.
- [Bu03] Bubb, H. (2003). Fahrerassistenz primär ein Beitrag zum Komfort oder für die Sicherheit? In: VDI (Hrsg.), *Der Fahrer im 21. Jahrhundert*, 25-33. Düsseldorf: VDI-Verlag.
- [Di95] Dieckmann, A. (1995). *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*, Hamburg.
- [Eh02] Ehmer, M. (2002). Mobile Dienste im Auto - Die Perspektive für Automobilhersteller? In: Reichwald, R. (Hrsg.), *Mobile Kommunikation*, 459-472. Wiesbaden: Gabler Verlag.
- [FS03] Frost&Sullivan (2003). *Customer Attitudes*
- [Fu01] Fuhr, A. (2001). Die Telematik ist tod - es lebe die rollende Schnittstelle. In: *Proceedings der Jahrestagung Telematik*, Bonn.
- [GL04] Gläser, J.; Laudel, G. (2004). *Experteninterviews*, 1 Aufl., Wiesbaden: UTB Verlag.
- [GW97] Greve, W.; Wentura, D. (1997). *Wissenschaftliche Beobachtung: Eine Einführung*, Weinheim: PVU/Beltz.
- [Ha04] Hartmann, J. (2004). Wo viel Licht ist, ist starker Schatten - Softwareentwicklung in der Automobilindustrie. *Automatisierungstechnik*, 52(8), 353-358.
- [Ha98] Harkness, J.; Mohler, P.; Schneid, M.; Bernhard, C. (1998). Inventives in Two German Mail Surveys 1996/79 & 1997. *Zuma-Nachrichten Spezial*, 4, pp. 201-218.
- [HLK07] Hoffmann, H.; Leimeister, J.M.; Krcmar, H. (2007). Anforderungen an Werkzeuge zum Prototyping von Automotive Services. In: *Proceedings der Multikonferenz Wirtschaftsinformatik, München - Garching*.
- [LK77] Landis, J.R.; Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- [LKK07] Leimeister, J.M.; Knebel, U.; Krcmar, H. (2007). RFID as enabler for the boundless real-time organisation: empirical insights from Germany. *IJNVO*, 4(1), 45-64.
- [Lo02] Lombard, M.; Snyder-Dutch, J.; Bracken, C.C. (2002). Content analysis in mass communication: assessment and reporting of intercoder reliability. *Human Communication Research*, 28, 597-604.
- [Me04] Mercer (2004). *Future Automotive Industry Structure (FAST) 2015*, Frankfurt a. M.: Henrich Druck + Medien.

- [Mo06] Mohan, L.R. (2006). Driving down the Fast Lane: Increasing Automotive Opportunities the EMS Provider Way <http://www.frost.com/prod/servlet/market-insight.pag?docid=67150588>, zugegriffen am: 19.07.
- [NDH02] Nieschlag, R.; Dichtl, E.; Hörschgen, H. (2002). Marketing, 19. Aufl., Berlin: Duncker&Humboldt.
- [NM78] Nora, S.; Minc, A. (1978). Die Informatisierung der Gesellschaft, 1. Aufl., Frankfurt am Main: Campus Verlag.
- [Po99] Porst, R. (1999). Thematik oder Incentives? . Zuma-Nachrichten, 45, 72-87.
- [Pr07] Pretschner, A.; Broy, M.; Krüger, H.; Stauner, T. (2007). Software Engineering für Automotive Systems - A Roadmap. In: Proceedings der International Conference on Software Engineering, Minneapolis.
- [RKR07] Reichwald, R.; Krcmar, H.; Reindl, S. (Hrsg.). (2007). Mobile Dienste im Auto der Zukunft. Lohmar - Köln: EUL Verlag.
- [RMF02] Reichwald, R.; Meier, R.; Fremuth, N. (2002). Die mobile Ökonomie - Definitionen und Spezifika. In: Reichwald, R. (Hrsg.), Mobile Kommunikation Wiesbaden: Gabler Verlag.
- [Ru05] Rubin, H.J.; Rubin, I.S. (2005). Qualitative Interviewing: The Art of Hearing Data.
- [Sa03] Saad, A. (2003). Prototyping bei der BMW Car IT GmbH. Java Spektrum(2), 49-53.
- [SB06] Statistisches Bundesamt (2006). Verkehr in Deutschland 2006, Wiesbaden: SFG Fachverlage.
- [So08] Stolte, P. (2008). Das Auto im Connected Life. In: Proceedings der Das vernetzte Automobil, München.
- [Sp08] Springer, J. (2008). Mauttechnologie: Plattform für Dienste im Fahrzeug. In: Proceedings der Das vernetzte Automobil, München.
- [St08] Steiger, P.-E. (2008). 112 and eCall: The LifeService platform. In: Proceedings der Das vernetzte Automobil, München.
- [Vi08] VICS-Center (2008). Vehicle Information and Communication System (VICS). Tokio.
- [Wr05] Werder, H. (2005). Verkehrstelematik als Element der Verkehrspolitik. In: Proceedings der, Olten.

Near Field Communication Use in Retail Stores: Effects on the Customer Shopping Process

Thomas Wiechert, Andreas Schaller, Frederic Thiesse

Institute of Technology Management
University of St. Gallen
Dufourstrasse 40a
9000 St. Gallen (Switzerland)
thomas.wiechert@unisg.ch
andreas.schaller@motorola.com
frederic.thiesse@unisg.ch

Abstract: This paper explores NFC based applications adequate for the use in retail stores and analyses the influence that these could have on the prevailing customer shopping process. It emphasizes the fact that NFC could provide shoppers with more benefits beyond faster payments. On the one hand, NFC based rebate coupons and loyalty cards could further accelerate the check-out process. On the other hand, NFC devices could be used to support customers with more information on available products. This could also reduce the customers' need for store personnel assistance. The paper concludes that NFC would not fundamentally change the customer shopping process but merely support it. This research was partially funded by the European Union through the FP6 project StoLPaN.

1 Introduction

Near Field Communication (NFC) is an emerging technology which combines contactless smart card technology with the convenience of mobile devices. The integration of NFC hardware e.g. into a mobile phone, enables it to emulate contactless smart cards as well as to read from and write onto such cards. The ability to act as a smart card makes it possible for NFC compatible phones to be used as replacement for physical contactless cards. The majority of the trials that were conducted with NFC technology focused on payment and ticketing applications. This means that the trial participants were either enabled to use NFC enabled mobile phones to pay at point of sales or to use them as electronic tickets for public transportation. This application focus is reflected by the coverage in trade publications: NFC is, as a general rule, described as a technology which enables the owners of compatible devices to use them for payment and ticketing services. There are, however, other applications which can be implemented based on this new technology. Against this background, this paper explores the NFC based applications adequate for the use in retail stores and analyses the influence that these applications could have on the prevailing customer shopping process.

2 Retail Applications for NFC

In order to accomplish a first assessment of the possible uses of NFC in stationary retail, we conducted an analysis of available industry publications pertaining to the technology. Publications from the GSM Association (GSMA) which represents 750 mobile network operators, the European association for standardizing information and communication systems (Ecma), the NFC Forum, and Innovision, a company designing NFC chips were considered [EC05,IN06,NF06,GS07]. The NFC applications most mentioned in these industry publications were: mobile payment, mobile ticketing, the transfer of data from one device to another, the easy pairing of devices, such as mobile phones and Bluetooth headsets, and the download of information, such as trailers and ring tones from so called smart posters. All in all, 14 different applications for NFC were named¹ and described, of which five were identified as deployable for the use in retail stores: Payment, the download of information, loyalty applications, rebate coupons, and product information.

In order to visualize how these five applications would change the way people shop, flowcharts of prevailing customer shopping processes were generated based on site surveys and discussions with store employees. Subsequently, the flowcharts were amended by the different possible NFC applications in order to identify the changes resulting from technology implementation. In order to verify the authenticity and realism of the processes, semi-structured interviews were conducted with nine European retailers for feedback in their regards. Five of the retailers were interviewed in their own offices in Switzerland and Germany, three were interviewed by phone, and the ninth was contacted by e-mail and also responded through the same channel.

While the process flowcharts were generated for apparel and footwear stores, book stores, gas stations, pharmacies, and supermarkets, the latter will be taken as example. The reason for the choice of the supermarket process lies in the fact that most NFC applications would mainly help to accelerate the checkout process. As the interviews have shown, the perception of the retail industry is, that solutions which can speed up the checkout process seem to yield the highest potential when implemented in supermarkets.

3 The current customer shopping process

The customer shopping process in supermarkets can be summarized as follows: The customer enters the supermarket, searches for the products he intends to buy or strolls through the store. When a product is located, he checks if it fulfils his requirements and makes a decision on whether or not to buy it. Optionally, the customer may ask the store personnel for assistance in order to locate a product or to receive more detailed information on a product he is uncertain of. This part of the shopping process is repeated until the customer makes the decision not to look for any more products. He then adjourns to the check-out area, where he might have to wait in line before placing his

¹ mobile payment, information download, contactless loyalty cards, electronic rebate coupons, product information, data transfer, easy device pairing, mobile ticketing, physical access control, logical access control, health care file storage, car ignition key storage, field force solution, support of children and elders

shopping on a counter or a conveyor belt. A cashier then scans the barcodes of all products and informs him about the total amount that is to be paid. Optionally, the customer can hand over coupons in order to claim rebates or his identification card for the retailer's loyalty program in order to collect reward points. The shopping process ends with the customer paying for his purchases by means of cash or electronic payment.

This shopping process does not feature many technological resources that help customers or improve their shopping experience. The only exceptions are the barcode scanners, which improve the check out process by accelerating it and by eliminating the need for manually capturing products, and the payment terminals which enable customers to pay for their purchases without the need to carry large amounts of cash. Nevertheless, 95% of the retailers consider waiting lines at their check-outs as the most serious problem to solve in order to better service their customers [CM02]. NFC technology could contribute to the alleviation of this problem. As the nine retailers have stated, this problem perception is especially shared by the operators of supermarkets and drug stores, which are very interested in accelerating the check out process. The other retailers such as apparel and footwear store operator perceive waiting lines at the check-out as less of a problem.

During the conducted interviews, different explanations were given concerning this problem: Shoppers usually don't go to supermarkets for pleasure or to treat themselves, but to fulfil basic needs of acquiring food, beverages, and other household necessities. Also, many customers do not go to supermarkets on their day off, but squeeze the chore of shopping between leaving work and going home. Therefore, most customers wish to spend as little time as possible in a supermarket. Due to the large number of products that the average customer buys in a supermarket when compared to other stores and owing to the requirement to have each of these products scanned by a cashier, the check-out process in a supermarket takes up a larger amount of time than in other stores.

4 The NFC enabled customer shopping process

In section 2, five NFC applications were identified as deployable for the use in retail stores. These applications can be split into two groups: Those that support the shoppers while they are on the store floor on the one hand, and those that could become part of the check out process.

The download of information and the procurement of product information constitute two different applications of a single solution. Both applications could support shoppers on the store floor. While the download of information generally describes the access of data such as rebate coupons and links to web based content triggered by reading an NFC tag embedded in a smart poster, the procurement of product information describes accessing data after tapping such a tag embedded in a product packaging or attached to the shelf holding the product. Both applications could be used to help shoppers to find out more about the product that retailers have to offer and to reduce their need for assistance by store personnel. Store personnel could also use these solutions in order to procure information on products and thereby better serve their customers.

The implementation of NFC based payment, loyalty applications and coupons on the other hand could give retailers a means to accelerate parts of the check-out process. As was stated before, the checkout process seems to yield the largest potential for an improvement of the entire customer shopping process [CM02]. NFC devices could hold the customers' payment cards, loyalty cards, and rebate coupons at the same time. Holding one NFC device up to a contactless reader could replace having to get two plastic cards and several coupons out of a wallet or purse.

As these descriptions show, the NFC based applications would not fundamentally change the customer shopping process, but merely support it on the store floor and in the check out area. They also illustrate, that the majority of the promoted NFC applications are focused on supporting the check-out share of the customer shopping process.

5 Feedback from retailers

While retailers that operate supermarkets and drug stores showed most interest in those NFC applications that could accelerate the check-out process, the operators of apparel stores and other specialty stores were more interested in providing customers with information on the sales floor. As mentioned before, supermarkets have a higher likelihood of customers being forced to wait in line at the check-out than other retail stores. They also have a large number of customers that want to leave the store as soon as they have gathered the products they require. On the other hand, solutions that would require item-level tagging of products with NFC tags, such as the retrieval of product information, sparked less interest on the part of supermarket operators. This reluctance was justified with the low average prices of supermarket products and the low margins of food retailers, which makes the tagging of products to expensive in the near future. In contrast, providing customers with product information or the possibility to download coupons by means of smart posters or shelf tags seemed to constitute an interesting possibility. The embedded NFC tags could be used by a large number customers over a long time period and do not have to be replaced when a product unit is sold.

Retailers operating department stores on the other hand were less interested in accelerating their check out processes, because, in their opinion, waiting lines constitute less of a problem in their stores than they do in supermarkets. Also, the check-out process is a less important part of the shopping process at a department store than it is in a supermarket. This is in part due to the smaller average number of products bought by customers which results in shorter scanning processes. Another reason is the fact that many customers do not just satisfy specific needs in department stores, but stroll through the stores and enjoy the shopping experience. This means that a smaller share is in a hurry and wishes to leave the store as soon as possible. While NFC based payment, loyalty programs, and coupons were not rejected by these retailers, their adoption is not a priority. Department store retailers were on the other hand very interested in NFC solutions that could provide shoppers with information on the shop floor. The main interest consisted in possibilities that could reduce the customers' need to consult store personnel. Discussed applications included the provisioning of general information about products or their availability in the front store and the back store upon touching a NFC

tag on a shelf or embedded in a smart poster. The possibility of tagging individual products was also discussed. Due to the higher average price of products in department stores, the tagging of individual products with NFC tags constitutes a more realistic possibility to this part of the retail industry than it does to the operators of supermarkets.

6 Conclusion

NFC is a technology that enables mobile devices to emulate contactless smart cards, read from or write to compliant RFID tags, and to communicate with each other in a peer-to-peer mode. These abilities make it possible for such devices to be used as payment cards, hold electronic tickets, and easily download data from tags and networks without the need for manual interaction such as typing in URLs. The NFC applications appropriate for the implementation in stationary retail most mentioned in industry publications are mobile payment, loyalty applications, electronic coupons, the download of information from smart posters, and the procurement of product information.

With the exception of barcode scanners and conventional contact based payment terminals, the customers' shopping process in most retail outlets does not include the use of any technology solutions to improve the customers' shopping experience or to accelerate the check-out process. NFC based payment, loyalty applications and coupons could support and accelerate the check out process. The download of information from tags attached to smart posters, shelves, or individual products on the other hand could make the time shoppers spend on the store floor more convenient. Interviews with nine European retailers indicate that supermarket and drug store operators indicate a greater interest in solutions that can accelerate the checkout process, while department store operators are more interested in improving the shopping experience on the store floor.

This paper has shown that the implementation of NFC technology in the current supermarket environment would not fundamentally change the customer shopping process, but merely support it. However, the combination of NFC technology with other current retail innovations such as mobile check-outs and self-scanning concepts could have a stronger influence on the way in which people shop. Such concepts could ultimately render regular check-outs unnecessary.

References

- [CM02] Chu, J.; Morrison, G.: Enhancing the customer shopping experience: 2002 IBM/NRF 'Store of the Future. IBM / National Retail Foundation, 2002.
- [EC05] Near Field Communication. Ecma International, 2005.
- [GS07] Mobile NFC Services. GSM Association, 2007.
- [IN06] Near Field Communication in the real World: Turning the NFC Promise into profitable everyday Applications. Innovision, 2006.
- [NF06] Near Field Communication and the NFC Forum: The Keys to Truly Interoperable Communications. NFC-Forum, 2006.

RFID-Systeme für Messen und Konferenzen: Untersuchung der Einflussgrößen auf die Besucherakzeptanz

Ulrich Bretschneider, Jan Marco Leimeister, Helmut Krcmar

Lehrstuhl für Wirtschaftsinformatik
Technische Universität München
Boltzmannstr. 3; 85748 Garching b. München
ulrich.bretschneider@in.tum.de
leimeister@in.tum.de
krcmar@in.tum.de

1 Einleitung

Radiofrequenzidentifikation (RFID) ist gegenwärtig ein häufig diskutiertes Thema, sowohl in der Praxis als auch in der Wissenschaft [vgl. bspw. Übersichten bei Le09 und LKK07]. Ein neues und bis dato in der wissenschaftlichen Forschung noch kaum berücksichtigtes Einsatzgebiet für RFID-Systeme ergibt sich für den Bereich der Veranstaltungsorganisation, insbesondere für Messen und Konferenzen. Die RFID-Technologie wird in diesem Umfeld wegen ihrer Eigenschaft der automatischen Identifikation (Auto-ID) diskutiert. Werden Namensschilder oder Eintrittskarten mit RFID-Transpondern bestückt, können anhand dieser die Besucher identifiziert werden. So sind beispielsweise Besucherstromanalysen, Authentifizierungen oder Zutrittskontrollen effektiver durchführbar. Zudem kann durch den RFID-Transponder eine erhöhte Fälschungssicherheit der Eintrittskarten realisiert werden.

Das Ziel des vorliegenden Beitrages ist zweigeteilt: So gilt es zum einen relevante Faktoren, die die Akzeptanz der Benutzer solcher RFID-Namensschilder und -Eintrittskarten, also die Besucher von Messen und Konferenzen, beeinflussen könnten, zu identifizieren. Zum anderen werden die identifizierten Einflussfaktoren bei Messe- und Konferenzbesuchern im Rahmen quantitativer Akzeptanzstudien empirisch abgefragt. Die Ergebnisse aus diesen Befragungen gewähren erste Anhaltspunkte zur Einschätzung der Benutzerakzeptanz.

2 Einflussgrößen auf die Benutzerakzeptanz

Die Einflussgröße wahrgenommene einfache Benutzbarkeit kann als ein grundlegender Einflussfaktor auf die Akzeptanz von Informationstechnologien angesehen werden. Dies wurde von Davis [Da89] in seinem Technologieakzeptanzmodell (TAM) nachgewiesen. Davis/Bagozzi [DB89] definieren die wahrgenommene einfache Benutzbarkeit als „...the degree to which the prospective user expects the target system to be free of ef-

fort.” Ebenso wie die wahrgenommene einfache Benutzbarkeit wies Davis auch den wahrgenommenen Nutzen als einen grundlegenden Einflussfaktor auf die Nutzerakzeptanz nach [Da89]. Den wahrgenommenen Nutzen definieren Davis/Bagozzi [DB89] als „the prospective user’s subjective probability that using a specific application system will increase his or her job performance within an organizational context.“ Diese Dimensionen wurden von zahlreichen Autoren wegen ihres allgemeingültigen Charakters für die Entwicklung verschiedener Akzeptanzmodelle bzw. die Durchführung verschiedener Akzeptanzuntersuchungen übernommen. Hossain/Prybutok [HP08] nutzten diese Faktoren erstmals im Kontext der RFID-Technologie. Sie stellten ein aus dem TAM abgeleitetes Akzeptanzmodell für die RFID-Technologie auf und verwenden darin auch die Faktoren **wahrgenommene einfache Benutzbarkeit** und **wahrgenommener Nutzen**.

Auch für den vorliegenden Untersuchungsgegenstand spielen die beiden Faktoren eine Rolle. Dabei ergibt sich für den Besucher einer Messe oder einer Konferenz bei Verwendung der mit RFID-Tags bestückten Eintrittskarten bzw. Namensschilder insofern ein Nutzen, als dass beispielsweise dem Besucher einer Konferenz via Bildschirm angezeigt werden kann, wie viele Besucher sich während eines Vortrages in einem Konferenzsaal befinden. Desweiteren könnte der Besucher einer Messe oder einer Konferenz persönlich auf einem im Eingangsbereich platzierten Bildschirm begrüßt werden, einem Messebesucher personalisierte Informationsunterlagen zugänglich gemacht werden oder Kontaktdaten elektronisch an Dritte übermittelt werden. Den Aufwand, den Messe- oder Konferenzbesucher zur Verwendung des oben aufgezählten Nutzens in Kauf nehmen müssen, ist als die wahrgenommene einfache Benutzbarkeit zu interpretieren. Die einfache Benutzbarkeit manifestiert sich in dem Umstand, dass die Besucher, um bestimmte Mehrwerte zu aktivieren, seinen RFID-Tag an ein Lesegerät halten müssen.

Eine weitere zu berücksichtigende Einflussgröße, die zwar bislang wenig Eingang in die wissenschaftliche Literatur gefunden hat, dafür aber in der breiten Öffentlichkeit im Zuge der RFID-Diskussion debattiert wird, ist die Echtheitszertifizierung, die durch RFID-Technologie ermöglicht werden kann. Beispielsweise wird in diesem Kontext in der Praxis derzeit erprobt, winzige RFID-Tags in Geldscheine zu integrieren. In den USA werden seit Oktober 2006 nur noch Reisepässe ausgegeben, die einen RFID-Tag tragen, und der Pharmahersteller Pfizer nutzt die RFID-Technologie zur Echtheitszertifizierung, indem er Medikamentenpackungen mit RFID-Transpondern ausstattet. Auch für den Bereich der Organisation von Messen und Konferenzen spielt die Echtheitszertifizierung eine Rolle (vgl. oben). Da die RFID-Technologie zur **Fälschungssicherheit** von Eintrittskarten etc. beiträgt und diese letztendlich auch im Interesse von Besuchern ist, soll dieser Aspekt als Einflussgröße auf die Benutzerakzeptanz für die zu Grunde liegende Untersuchung berücksichtigt werden.

Wenn RFID-Tags in der Messe- und Konferenzorganisation eingesetzt werden, muss auch die Privatsphäre berücksichtigt werden. Die Privatsphäre der Benutzer übt insofern einen Einfluss auf die Akzeptanz aus, als dass die Bewegungen und Aufenthaltsorte von Messe- und Konferenzbesuchern mit Hilfe der RFID-Tags aufgezeichnet werden können (vgl. oben). Dieses **Tracking** wird in der Literatur aus der Perspektive der Benutzer als möglicher Eingriff in die Privatsphäre diskutiert [Jo04; OSK05; SZ04].

Ein weiterer Faktor, der einen Einfluss auf die Benutzerakzeptanz ausüben könnte, ist der Aspekt der **Datensicherheit**. So wird das von Nutzern unbemerkte Auslesen der persönlichen Nutzerdaten auf dem RFID-Tag in der Literatur als ernstzunehmendes Problem beschrieben [GS04; OSK05; Sm05; SZ04]. Übertragen auf den Untersuchungsgegenstand könnte ein AngstszENARIO so aussehen, dass beispielsweise die persönlichen Daten der Messebesucher von Ausstellern unbemerkt ausgelesen werden, um diese für ungefragte Werbezwecke zu benutzen. Aus diesem Grund soll der Faktor Datensicherheit auch für die vorliegende Untersuchung herangezogen werden.

Außerdem könnten **gesundheitliche Bedenken** der Besucher einen relevanten Einflussfaktor darstellen. Dieser Aspekt ist in der öffentlichen Wahrnehmung wegen der möglichen Einwirkung der elektromagnetischen Felder auf den menschlichen Körper häufig diskutiert. Aus diesem Grund soll er auch in der vorliegenden Studie Berücksichtigung finden.

3 Evaluierung von Einflussfaktoren im Rahmen einer quantitativen Studie

3.1 Untersuchungsdesign

Die oben identifizierten Einflussfaktoren wurden im Rahmen einer quantitativen Befragung abgefragt. Die Zielgruppe der Befragung setzte sich zum einen aus den Besuchern der internationalen Fachkonferenz „Organic Electronics Conference and Exhibition“, die im September 2007 in Frankfurt stattfand, und zum anderen aus den Besuchern der internationalen Fachmesse „MEDIA-TECH Expo“, die im Mai 2008 in Frankfurt stattfand, zusammen.

Für die Erhebung im Rahmen der beiden Events wurde ein identischer Fragebogen in englischer Sprache entworfen. Die identifizierten Einflussfaktoren wurden in Aussagen übersetzt, die die Befragten anhand einer fünfstufigen Rating-Skala nach ihrem subjektiven Empfinden bewerten sollten („trifft voll zu“ = 5 bis „trifft gar nicht zu“ = 1).

Im Rahmen eines Pre-Testes mit 8 Experten zum Thema RFID aus dem Wissenschafts- und Unternehmensumfeld wurde der ursprüngliche Fragebogen überprüft und weiterentwickelt. Die endgültige Version des Fragebogens wurde Besuchern der OEC und der MEDIA-TECH Expo vorgelegt. Da die Auswahl der Befragungsteilnehmer zufällig erfolgte, war die Stichprobe selbstselektierend. Da zudem keine Grundgesamtheit ausgemacht werden konnte, kann nicht sichergestellt werden, dass die Stichprobe repräsentativ ist. Insgesamt nahmen 416 Personen an der Befragung teil, von denen nach Datenbereinigungsmaßnahmen (ungültige Fragebögen wurden entfernt) 387 vollständig auswertbare Antwortsätze für die Auswertung zur Verfügung standen.

3.2 Darstellung und Interpretation der empirischen Ergebnisse

Der arithmetische Mittelwert des Einflussfaktors **wahrgenommener Nutzen** zeigt eine mittlere Ausprägung (Mittelwert (M) = 3,619; Standardabweichung (SD) = 1,123). Demnach bewertet zwar eine knappe Mehrheit der Befragten die Serviceleistungen, die sich mit Hilfe der RFID-bestückten Eintrittskarten für Messe- und Konferenzbesucher gene-

rieren lassen, für sich als einen Mehrwert, doch sehen auch viele hierin keinen echten Mehrwert. Da diese Leistungen weder eine echte Zeit- noch Kostenersparnis für den Besucher einbringen, interpretieren die Skeptiker diese wohl eher als Spielereien, auf die man im Zweifelsfall auch verzichten könnte.

Die **wahrgenommene einfache Bedienung**, die sich in dem Umstand für den Besucher manifestiert, die mit RFID-Tags ausgestatteten Eintrittskarten oder Namensschilder an Lesegeräte zu halten, um einen Besucherservice nutzen zu können, wird von den Befragten im Durchschnitt als mittelmäßig bewertet ($M=3,241$; $SD=1,293$). Dieses Ergebnis ist als logische Konsequenz des Ergebnisses zur Einflussgröße „wahrgenommener Nutzen“ zu werten: Zwar wird der vom Besucher zu leistende Aufwand, um eine Serviceleistung nutzen zu können, nicht wirklich als störend empfunden, aber angesichts des dabei zu erwartenden geringen Mehrwerts der Serviceleistungen (vgl. oben) auch nicht als lohnenswert. Dabei ist davon auszugehen, dass die Befragten rationale Entscheidungen fällen: Wäre der Nutzen für die Befragten ein höherer, würden die Befragten den Aufwand wohl eher in Kauf nehmen und damit entsprechend positiver bewerten.

Ebenso messen die Befragten der Tatsache, dass die RFID-Tags in den Eintrittskarten zur **Fälschungssicherheit** beitragen, eine mittlere bis hohe Bedeutung bei ($M=3,868$; $SD=1,058$). Aus diesem Ergebnis kann abgelesen werden, dass die Mehrheit der Befragten die Fälschungssicherheit, die in erster Linie einen Mehrwert für den Veranstalter bedeutet, auch für sich selbst als einen Vorteil sehen.

Der Möglichkeit, dass die Veranstalter von Messen und Konferenzen ein **Tracking** der Besucher mit Hilfe der RFID-bestückten Eintrittskarten realisieren könnten, wird von den Befragten im Durchschnitt eine mittlere Bedeutung beigemessen ($M=2,897$; $SD=1,344$). Dieses Ergebnis zeigt, dass einem nicht unbedeutenden Teil der Befragten ein mögliches Tracking nichts auszumachen scheint. Damit wird das Ergebnis von Hossain/Prybutok [HP08] widerspiegelt, die in ihrer Untersuchung zu Benutzerakzeptanz von RFID-Systemen in unterschiedlichen Anwendungskontexten exakt denselben Sachverhalt abfragten und dabei zu der Erkenntnis gelangten, dass die Möglichkeit des Personentracking keinen signifikanten Einfluss auf die Akzeptanz von RFID-Systemen ausübt. Eine mögliche Erklärung für diese Wahrnehmung mag darin liegen, wie RFID-Systeme im Allgemeinen und im Speziellen im vorliegenden Untersuchungskontext genutzt werden. So operieren solche Systeme meist im Verborgenen, außerhalb der Sichtweite des Benutzers und somit außerhalb des Bewusstseins des Benutzers.

Ebenso überraschend wie das Ergebnis zur Einflussgröße „Tracking“ ist das Ergebnis zum Faktor **Datensicherheit**. So schätzen die Befragten die Gefahr, dass die auf dem RFID-Tag gespeicherten persönlichen Daten unbefugt ausgelesen werden könnten, im Durchschnitt eher als unproblematisch ein ($M=2,406$; $SD=1,261$). Damit steht das Ergebnis in einem Kontrast zu den Ergebnissen von Hossain/Prybutok [HP08], die in ihrer Untersuchung denselben Faktor abfragten. Als Erklärung hierfür ist wohl der Neuheitscharakter der RFID-Technologie in diesem Anwendungskontext heranzuziehen. So dürfte den meisten der Befragten die Problematik der Datensicherheit zu diesem Zeitpunkt noch nicht ausreichend bewusst gewesen sein. **Gesundheitliche Risiken**, die aus der

Funktechnologie der RFID-Tags resultieren könnten, befürchten die Befragten im Durchschnitt nicht (M=1,887; SD=1,059).

4 Fazit und weiterer Forschungsbedarf

Die RFID-Technologie stellt für die Durchführung von Messen und Konferenzen ein vielversprechendes Anwendungsfeld dar, das in der wissenschaftlichen Literatur bislang keine Berücksichtigung gefunden hat. Der vorliegende Beitrag hat die Einflussgrößen, die potenziell auf die Besucherakzeptanz wirken, herausgearbeitet und diese im Rahmen einer Befragung von Besuchern einer Messe und einer Konferenz bewerten lassen. Die Ergebnisse erheben nicht den Anspruch, eine Akzeptanzmessung als solches zu sein, da die Studie zu einem Zeitpunkt durchgeführt wurde, zu dem diese Technologie noch nicht als etabliert bezeichnet werden konnte und der RFID-Einsatz in diesem Bereich in der Wahrnehmung der Benutzer noch nicht ausreichend verankert war. So kann sie nur als eine Vorstudie für eine Akzeptanzmessung dienen, aus der aber erste Tendenzen abgelesen werden können. Daraus ergibt sich der zukünftige Forschungsbedarf nach einer nochmaligen, zu einem späteren Zeitpunkt durchzuführenden Akzeptanzmessung.

Literaturverzeichnis

- [Da89] Davis, F. D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. In: MIS Quarterly 13 (1989), S. 318-339.
- [DB89] Davis, F. D.; Bagozzi, R. P.: User acceptance of computer technology: A comparison of two theoretical models. In: Management Science 35 (1989), S. 982-1003.
- [GS04] Günther, O. S.; Spiekermann, S.: RFID vs. Privatsphäre – ein Widerspruch? In: Wirtschaftsinformatik 46 (2004), S. 245-246.
- [HP08] Hossain, M. M.; Prybutok, V. R.: Consumer Acceptance of RFID Technology: An Exploratory Study. In: IEEE Transactions on Engineering Management 55 (2008), S. 316-328.
- [Jo04] Jones, P.; Clarke-Hill, C.; Hillier, D.; Shears, P.; Comfort, D.: Radio frequency identification in retailing and privacy and public policy issues. In: Management Research News 27 (2004), S. 46-56.
- [Le09] Leimeister, S.; Leimeister J. M.; Knebel, U.; Krcmar, H.: A cross-national comparison of perceived strategic importance of RFID for CIOs in Germany and Italy. In: International Journal of Information Management 29 (2009).
- [LKK07] Leimeister, J. M.; Knebel, U.; Krcmar, H.: RFID as Enabler for the Boundless real-time Organisation: Empirical Insights from Germany. In: International Journal of Networking and Virtual Organisation 4 (2007), S. 45-64.
- [OSK05] Ohkubo, M.; Suzuki, K.; Kinoshita, S.: RFID privacy issues and technical challenges. In: Communication of the ACM 48 (2005), S. 66-71.
- [Sm05] Smith, A.: Exploring radio frequency identification technology and its impacts on business systems. In: Inf. Manage. Comp. Security 13 (2005), S. 16-28.
- [SZ04] Spiekermann, S.; Ziekow, H.: Technische Analyse RFID-bezogener Angstsszenarien. Institut für Wirtschaftsinformatik, Humboldt-Universität zu Berlin, Berlin, 2004.

GI-Edition Lecture Notes in Informatics

- P-1 Gregor Engels, Andreas Oberweis, Albert Zündorf (Hrsg.): Modellierung 2001.
- P-2 Mikhail Godlevsky, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications, ISTA'2001.
- P-3 Ana M. Moreno, Reind P. van de Riet (Hrsg.): Applications of Natural Language to Information Systems, NLDB'2001.
- P-4 H. Wörn, J. Mühlhing, C. Vahl, H.-P. Meinzer (Hrsg.): Rechner- und sensorgestützte Chirurgie; Workshop des SFB 414.
- P-5 Andy Schürr (Hg.): OMER – Object-Oriented Modeling of Embedded Real-Time Systems.
- P-6 Hans-Jürgen Appelrath, Rolf Beyer, Uwe Marquardt, Heinrich C. Mayr, Claudia Steinberger (Hrsg.): Unternehmen Hochschule, UH'2001.
- P-7 Andy Evans, Robert France, Ana Moreira, Bernhard Rumpe (Hrsg.): Practical UML-Based Rigorous Development Methods – Countering or Integrating the extremists, pUML'2001.
- P-8 Reinhard Keil-Slawik, Johannes Magenheimer (Hrsg.): Informatikunterricht und Medienbildung, INFOS'2001.
- P-9 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Innovative Anwendungen in Kommunikationsnetzen, 15. DFN Arbeitstagung.
- P-10 Mirjam Minor, Steffen Staab (Hrsg.): 1st German Workshop on Experience Management: Sharing Experiences about the Sharing Experience.
- P-11 Michael Weber, Frank Kargl (Hrsg.): Mobile Ad-Hoc Netzwerke, WMAN 2002.
- P-12 Martin Glinz, Günther Müller-Luschnat (Hrsg.): Modellierung 2002.
- P-13 Jan von Knop, Peter Schirmbacher and Viljan Mahni_ (Hrsg.): The Changing Universities – The Role of Technology.
- P-14 Robert Tolksdorf, Rainer Eckstein (Hrsg.): XML-Technologien für das Semantic Web – XSW 2002.
- P-15 Hans-Bernd Bludau, Andreas Koop (Hrsg.): Mobile Computing in Medicine.
- P-16 J. Felix Hampe, Gerhard Schwabe (Hrsg.): Mobile and Collaborative Business 2002.
- P-17 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Zukunft der Netze – Die Verletzbarkeit meistern, 16. DFN Arbeitstagung.
- P-18 Elmar J. Sinz, Markus Plaha (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2002.
- P-19 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3.Okt. 2002 in Dortmund.
- P-20 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3.Okt. 2002 in Dortmund (Ergänzungsband).
- P-21 Jörg Desel, Mathias Weske (Hrsg.): Promise 2002: Prozessorientierte Methoden und Werkzeuge für die Entwicklung von Informationssystemen.
- P-22 Sigrid Schubert, Johannes Magenheimer, Peter Hubwieser, Torsten Brinda (Hrsg.): Forschungsbeiträge zur “Didaktik der Informatik” – Theorie, Praxis, Evaluation.
- P-23 Thorsten Spitta, Jens Borchers, Harry M. Sneed (Hrsg.): Software Management 2002 – Fortschritt durch Beständigkeit
- P-24 Rainer Eckstein, Robert Tolksdorf (Hrsg.): XMIDX 2003 – XML-Technologien für Middleware – Middleware für XML-Anwendungen
- P-25 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Commerce – Anwendungen und Perspektiven – 3. Workshop Mobile Commerce, Universität Augsburg, 04.02.2003
- P-26 Gerhard Weikum, Harald Schöning, Erhard Rahm (Hrsg.): BTW 2003: Datenbanksysteme für Business, Technologie und Web
- P-27 Michael Kroll, Hans-Gerd Lipinski, Kay Melzer (Hrsg.): Mobiles Computing in der Medizin
- P-28 Ulrich Reimer, Andreas Abecker, Steffen Staab, Gerd Stumme (Hrsg.): WM 2003: Professionelles Wissensmanagement – Erfahrungen und Visionen
- P-29 Antje Düsterhöft, Bernhard Thalheim (Eds.): NLDB'2003: Natural Language Processing and Information Systems
- P-30 Mikhail Godlevsky, Stephen Liddle, Heinrich C. Mayr (Eds.): Information Systems Technology and its Applications
- P-31 Arslan Brömmme, Christoph Busch (Eds.): BIOSIG 2003: Biometric and Electronic Signatures

- P-32 Peter Hubwieser (Hrsg.): Informatische Fachkonzepte im Unterricht – INFOS 2003
- P-33 Andreas Geyer-Schulz, Alfred Taudes (Hrsg.): Informationswirtschaft: Ein Sektor mit Zukunft
- P-34 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenberg, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 1)
- P-35 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenberg, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 2)
- P-36 Rüdiger Grimm, Hubert B. Keller, Kai Rannenberg (Hrsg.): Informatik 2003 – Mit Sicherheit Informatik
- P-37 Arndt Bode, Jörg Desel, Sabine Rathmayer, Martin Wessner (Hrsg.): DeLFI 2003: e-Learning Fachtagung Informatik
- P-38 E.J. Sinz, M. Plaha, P. Neckel (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2003
- P-39 Jens Nedon, Sandra Frings, Oliver Göbel (Hrsg.): IT-Incident Management & IT-Forensics – IMF 2003
- P-40 Michael Rebstock (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2004
- P-41 Uwe Brinkschulte, Jürgen Becker, Dietmar Fey, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle, Thomas Runkler (Edts.): ARCS 2004 – Organic and Pervasive Computing
- P-42 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Economy – Transaktionen und Prozesse, Anwendungen und Dienste
- P-43 Birgitta König-Ries, Michael Klein, Philipp Obreiter (Hrsg.): Persistence, Scalability, Transactions – Database Mechanisms for Mobile Applications
- P-44 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): Security, E-Learning, E-Services
- P-45 Bernhard Rumpe, Wolfgang Hesse (Hrsg.): Modellierung 2004
- P-46 Ulrich Flegel, Michael Meier (Hrsg.): Detection of Intrusions of Malware & Vulnerability Assessment
- P-47 Alexander Prosser, Robert Krimmer (Hrsg.): Electronic Voting in Europe – Technology, Law, Politics and Society
- P-48 Anatoly Doroshenko, Terry Halpin, Stephen W. Liddle, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications
- P-49 G. Schiefer, P. Wagner, M. Morgenstern, U. Rickert (Hrsg.): Integration und Datensicherheit – Anforderungen, Konflikte und Perspektiven
- P-50 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 1) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm
- P-51 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 2) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm
- P-52 Gregor Engels, Silke Seehusen (Hrsg.): DELFI 2004 – Tagungsband der 2. e-Learning Fachtagung Informatik
- P-53 Robert Giegerich, Jens Stoye (Hrsg.): German Conference on Bioinformatics – GCB 2004
- P-54 Jens Borchers, Ralf Kneuper (Hrsg.): Softwaremanagement 2004 – Outsourcing und Integration
- P-55 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): E-Science und Grid Ad-hoc-Netze Medienintegration
- P-56 Fernand Feltz, Andreas Oberweis, Benoit Otjacques (Hrsg.): EMISA 2004 – Informationssysteme im E-Business und E-Government
- P-57 Klaus Turowski (Hrsg.): Architekturen, Komponenten, Anwendungen
- P-58 Sami Beydeda, Volker Gruhn, Johannes Mayer, Ralf Reussner, Franz Schweiggert (Hrsg.): Testing of Component-Based Systems and Software Quality
- P-59 J. Felix Hampe, Franz Lehner, Key Pousttchi, Kai Ranneberg, Klaus Turowski (Hrsg.): Mobile Business – Processes, Platforms, Payments
- P-60 Steffen Friedrich (Hrsg.): Unterrichtskonzepte für informatische Bildung
- P-61 Paul Müller, Reinhard Gotzhein, Jens B. Schmitt (Hrsg.): Kommunikation in verteilten Systemen
- P-62 Federrath, Hannes (Hrsg.): „Sicherheit 2005“ – Sicherheit – Schutz und Zuverlässigkeit
- P-63 Roland Kaschek, Heinrich C. Mayr, Stephen Liddle (Hrsg.): Information Systems – Technology and its Applications

- P-64 Peter Liggesmeyer, Klaus Pohl, Michael Goedicke (Hrsg.): Software Engineering 2005
- P-65 Gottfried Vossen, Frank Leymann, Peter Lockemann, Wolfried Stucky (Hrsg.): Datenbanksysteme in Business, Technologie und Web
- P-66 Jörg M. Haake, Ulrike Lucke, Djamshid Tavangarian (Hrsg.): DeLFI 2005: 3. deutsche e-Learning Fachtagung Informatik
- P-67 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 1)
- P-68 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 2)
- P-69 Robert Hirschfeld, Ryszard Kowalczyk, Andreas Polze, Matthias Weske (Hrsg.): NODE 2005, GSEM 2005
- P-70 Klaus Turowski, Johannes-Maria Zaha (Hrsg.): Component-oriented Enterprise Application (COAE 2005)
- P-71 Andrew Torda, Stefan Kurz, Matthias Rarey (Hrsg.): German Conference on Bioinformatics 2005
- P-72 Klaus P. Jantke, Klaus-Peter Fähnrich, Wolfgang S. Wittig (Hrsg.): Marktplatz Internet: Von e-Learning bis e-Payment
- P-73 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): "Heute schon das Morgen sehen"
- P-74 Christopher Wolf, Stefan Lucks, Po-Wah Yau (Hrsg.): WEWoRC 2005 – Western European Workshop on Research in Cryptology
- P-75 Jörg Desel, Ulrich Frank (Hrsg.): Enterprise Modelling and Information Systems Architecture
- P-76 Thomas Kirste, Birgitta König-Riess, Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Informationssysteme – Potentiale, Hindernisse, Einsatz
- P-77 Jana Dittmann (Hrsg.): SICHERHEIT 2006
- P-78 K.-O. Wenkel, P. Wagner, M. Morgens-tern, K. Luzi, P. Eisermann (Hrsg.): Land- und Ernährungswirtschaft im Wandel
- P-79 Bettina Biel, Matthias Book, Volker Gruhn (Hrsg.): Softwareengineering 2006
- P-80 Mareike Schoop, Christian Huemer, Michael Rebstock, Martin Bichler (Hrsg.): Service-Oriented Electronic Commerce
- P-81 Wolfgang Karl, Jürgen Becker, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle (Hrsg.): ARCS '06
- P-82 Heinrich C. Mayr, Ruth Brey (Hrsg.): Modellierung 2006
- P-83 Daniel Huson, Oliver Kohlbacher, Andrei Lupas, Kay Nieselt and Andreas Zell (eds.): German Conference on Bioinformatics
- P-84 Dimitris Karagiannis, Heinrich C. Mayr, (Hrsg.): Information Systems Technology and its Applications
- P-85 Witold Abramowicz, Heinrich C. Mayr, (Hrsg.): Business Information Systems
- P-86 Robert Krimmer (Ed.): Electronic Voting 2006
- P-87 Max Mühlhäuser, Guido Röbling, Ralf Steinmetz (Hrsg.): DELFI 2006: 4. e-Learning Fachtagung Informatik
- P-88 Robert Hirschfeld, Andreas Polze, Ryszard Kowalczyk (Hrsg.): NODE 2006, GSEM 2006
- P-90 Joachim Schelp, Robert Winter, Ulrich Frank, Bodo Rieger, Klaus Turowski (Hrsg.): Integration, Informationslogistik und Architektur
- P-91 Henrik Stormer, Andreas Meier, Michael Schumacher (Eds.): European Conference on eHealth 2006
- P-92 Fernand Feltz, Benoît Otjacques, Andreas Oberweis, Nicolas Poussing (Eds.): AIM 2006
- P-93 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 1
- P-94 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 2
- P-95 Matthias Weske, Markus Nüttgens (Eds.): EMISA 2005: Methoden, Konzepte und Technologien für die Entwicklung von dienstbasierten Informationssystemen
- P-96 Saartje Brockmans, Jürgen Jung, York Sure (Eds.): Meta-Modelling and Ontologies
- P-97 Oliver Göbel, Dirk Schadt, Sandra Frings, Hardo Hase, Detlef Günther, Jens Nedon (Eds.): IT-Incident Mangament & IT-Forensics – IMF 2006

- P-98 Hans Brandt-Pook, Werner Simonsmeier und Thorsten Spitta (Hrsg.): Beratung in der Softwareentwicklung – Modelle, Methoden, Best Practices
- P-99 Andreas Schwill, Carsten Schulte, Marco Thomas (Hrsg.): Didaktik der Informatik
- P-100 Peter Forbrig, Günter Siegel, Markus Schneider (Hrsg.): HDI 2006: Hochschuldidaktik der Informatik
- P-101 Stefan Böttinger, Ludwig Theuvsen, Susanne Rank, Marlies Morgenstern (Hrsg.): Agrarinformatik im Spannungsfeld zwischen Regionalisierung und globalen Wertschöpfungsketten
- P-102 Otto Spaniol (Eds.): Mobile Services and Personalized Environments
- P-103 Alfons Kemper, Harald Schöning, Thomas Rose, Matthias Jarke, Thomas Seidl, Christoph Quix, Christoph Brochhaus (Hrsg.): Datenbanksysteme in Business, Technologie und Web (BTW 2007)
- P-104 Birgitta König-Ries, Franz Lehner, Rainer Malaka, Can Türker (Hrsg.) MMS 2007: Mobilität und mobile Informationssysteme
- P-105 Wolf-Gideon Bleek, Jörg Raasch, Heinz Züllighoven (Hrsg.) Software Engineering 2007
- P-106 Wolf-Gideon Bleek, Henning Schwentner, Heinz Züllighoven (Hrsg.) Software Engineering 2007 – Beiträge zu den Workshops
- P-107 Heinrich C. Mayr, Dimitris Karagiannis (eds.) Information Systems Technology and its Applications
- P-108 Arslan Brömme, Christoph Busch, Detlef Hühnlein (eds.) BIOSIG 2007: Biometrics and Electronic Signatures
- P-109 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 1
- P-110 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 2
- P-111 Christian Eibl, Johannes Magenheimer, Sigrid Schubert, Martin Wessner (Hrsg.) DeLFI 2007: 5. e-Learning Fachtagung Informatik
- P-112 Sigrid Schubert (Hrsg.) Didaktik der Informatik in Theorie und Praxis
- P-113 Sören Auer, Christian Bizer, Claudia Müller, Anna V. Zhdanova (Eds.) The Social Semantic Web 2007 Proceedings of the 1st Conference on Social Semantic Web (CSSW)
- P-114 Sandra Frings, Oliver Göbel, Detlef Günther, Harido G. Hase, Jens Nedon, Dirk Schadt, Arslan Brömme (Eds.) IMF2007 IT-incident management & IT-forensics Proceedings of the 3rd International Conference on IT-Incident Management & IT-Forensics
- P-115 Claudia Falter, Alexander Schliep, Joachim Selbig, Martin Vingron and Dirk Walther (Eds.) German conference on bioinformatics GCB 2007
- P-116 Witold Abramowicz, Leszek Maciszek (Eds.) Business Process and Services Computing 1st International Working Conference on Business Process and Services Computing BPSC 2007
- P-117 Ryszard Kowalczyk (Ed.) Grid service engineering and management The 4th International Conference on Grid Service Engineering and Management GSEM 2007
- P-118 Andreas Hein, Wilfried Thoben, Hans-Jürgen Appelrath, Peter Jensch (Eds.) European Conference on ehealth 2007
- P-119 Manfred Reichert, Stefan Strecker, Klaus Turowski (Eds.) Enterprise Modelling and Information Systems Architectures Concepts and Applications
- P-120 Adam Pawlak, Kurt Sandkuhl, Wojciech Cholewa, Leandro Soares Indrusiak (Eds.) Coordination of Collaborative Engineering - State of the Art and Future Challenges
- P-121 Korbinian Herrmann, Bernd Bruegge (Hrsg.) Software Engineering 2008 Fachtagung des GI-Fachbereichs Softwaretechnik
- P-122 Walid Maalej, Bernd Bruegge (Hrsg.) Software Engineering 2008 - Workshopband Fachtagung des GI-Fachbereichs Softwaretechnik

- P-123 Michael H. Breitner, Martin Breunig, Elgar Fleisch, Ley Pousttchi, Klaus Turowski (Hrsg.)
Mobile und Ubiquitäre Informationssysteme – Technologien, Prozesse, Marktfähigkeit
Proceedings zur 3. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2008)
- P-124 Wolfgang E. Nagel, Rolf Hoffmann, Andreas Koch (Eds.)
9th Workshop on Parallel Systems and Algorithms (PASA)
Workshop of the GI/ITG Special Interest Groups PARS and PARVA
- P-125 Rolf A.E. Müller, Hans-H. Sundermeier, Ludwig Theuvsen, Stephanie Schütze, Marlies Morgenstern (Hrsg.)
Unternehmens-IT: Führungsinstrument oder Verwaltungsbürde
Referate der 28. GIL Jahrestagung
- P-126 Rainer Gimmich, Uwe Kaiser, Jochen Quante, Andreas Winter (Hrsg.)
10th Workshop Software Reengineering (WSR 2008)
- P-127 Thomas Kühne, Wolfgang Reising, Friedrich Steimann (Hrsg.)
Modellierung 2008
- P-128 Ammar Alkassar, Jörg Siekmann (Hrsg.)
Sicherheit 2008
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 4. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
2.-4. April 2008
Saarbrücken, Germany
- P-129 Wolfgang Hesse, Andreas Oberweis (Eds.)
Sigsand-Europe 2008
Proceedings of the Third AIS SIGSAND European Symposium on Analysis, Design, Use and Societal Impact of Information Systems
- P-130 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
1. DFN-Forum Kommunikationstechnologien Beiträge der Fachtagung
- P-131 Robert Krimmer, Rüdiger Grimm (Eds.)
3rd International Conference on Electronic Voting 2008
Co-organized by Council of Europe, Gesellschaft für Informatik and E-Voting.CC
- P-132 Silke Seehusen, Ulrike Lucke, Stefan Fischer (Hrsg.)
DeLFI 2008:
Die 6. e-Learning Fachtagung Informatik
- P-133 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)
INFORMATIK 2008
Beherrschbare Systeme – dank Informatik Band 1
- P-134 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)
INFORMATIK 2008
Beherrschbare Systeme – dank Informatik Band 2
- P-135 Torsten Brinda, Michael Fothe, Peter Hubwieser, Kirsten Schlüter (Hrsg.)
Didaktik der Informatik – Aktuelle Forschungsergebnisse
- P-136 Andreas Beyer, Michael Schroeder (Eds.)
German Conference on Bioinformatics GCB 2008
- P-137 Arslan Brömme, Christoph Busch, Detlef Hühnlein (Eds.)
BIOSIG 2008: Biometrics and Electronic Signatures
- P-138 Barbara Dinter, Robert Winter, Peter Chamoni, Norbert Gronau, Klaus Turowski (Hrsg.)
Synergien durch Integration und Informationslogistik
Proceedings zur DW2008
- P-139 Georg Herzwurm, Martin Mikusz (Hrsg.)
Industrialisierung des Software-Managements
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschaftsinformatik
- P-140 Oliver Göbel, Sandra Frings, Detlef Günther, Jens Nedon, Dirk Schadt (Eds.)
IMF 2008 - IT Incident Management & IT Forensics
- P-141 Peter Loos, Markus Nüttgens, Klaus Turowski, Dirk Werth (Hrsg.)
Modellierung betrieblicher Informationssysteme (MobIS 2008)
Modellierung zwischen SOA und Compliance Management
- P-142 R. Bill, P. Korduan, L. Theuvsen, M. Morgenstern (Hrsg.)
Anforderungen an die Agrarinformatik durch Globalisierung und Klimaveränderung
- P-143 Peter Liggesmeyer, Gregor Engels, Jürgen Münch, Jörg Dörr, Norman Riegel (Hrsg.)
Software Engineering 2009
Fachtagung des GI-Fachbereichs Softwaretechnik

- P-144 Johann-Christoph Freytag, Thomas Ruf,
Wolfgang Lehner, Gottfried Vossen
(Hrsg.)
Datenbanksysteme in Business,
Technologie und Web (BTW)
- P-145 Knut Hinkelmann, Holger Wache (Eds.)
WM2009: 5th Conference on Professional
Knowledge Management
- P-146 Markus Bick, Martin Breunig,
Hagen Höpfner (Hrsg.)
Mobile und Ubiquitäre
Informationssysteme – Entwicklung,
Implementierung und Anwendung
4. Konferenz Mobile und Ubiquitäre
Informationssysteme (MMS 2009)

The titles can be purchased at:

Köllen Druck + Verlag GmbH

Ernst-Robert-Curtius-Str. 14 · D-53117 Bonn

Fax: +49 (0)228/9898222

E-Mail: druckverlag@koellen.de