

Proceedings of the Seminar

Research Trends in Media Informatics

EDITORS

Florian Schaub, Stefan Dietzel, Boto Bako, Bastian Könings,
Elmar Schoch, Björn Wiedersheim, and Michael Weber

12–13 February, 2010

CONTENTS

Foreword	3
Ubiquitous Computing	
Intelligent Environments	5
<i>Timo Müller</i>	
Ambient Assisted Living	13
<i>Nadhem Kachroudi</i>	
Next Generation Location-based Services	21
<i>Fabian Sulzbacher</i>	
Privacy in Video Surveillance	29
<i>Raimar Wagner</i>	
Devices and Games	
Ubiquitous Devices: History and Research Trends	35
<i>Hendrik Burmann</i>	
Market-ready Ubiquitous Systems	43
<i>Sven Reichel</i>	
Pervasive Gaming	51
<i>Rainer Möhle</i>	
Alternate Reality Games	59
<i>Ferdinand Deger</i>	
Networking Trends	
Bio-inspired Networking	67
<i>Sebastian Schimmel</i>	
Whitespace Networking	75
<i>Johannes Merx</i>	
Cloud Computing	83
<i>Benjamin Erb</i>	

The Future Vehicle <i>Sung-Eun Kang</i>	91
Social Interaction	
The Crowdsourcing Phenomenon <i>Stefan Kaufmann</i>	99
Peer to Peer Social Networks <i>David Langer</i>	105
Mobile Social Applications <i>Harry Brandenburg</i>	113
Social Computing: An Overview <i>Manuel Deil</i>	121

Foreword

Media Informatics is a broad field spanning many research directions and topics. The Internet, Web 2.0, Social Networks, Ubiquitous Computing, Computer Graphics, Usability, Social Interactions, HCI, and Privacy are just a few examples. The goal of the seminar *Research Trends in Media Informatics* was to motivate students to delve deeper into the vast and diverse research fields of Media Informatics, going beyond topics usually covered in lectures. The topics were selected with an emphasis on highlighting current and future research directions and challenges that are focus of active research not only at the Institute of Media Informatics at Ulm University but all over the world. The presented papers included in the proceedings cover the topics ubiquitous computing, social interaction, networking trends, and devices and games. Often the authors take very different vantage points on these themes.

The seminar started in the winter semester 2009/2010 also as an experiment to apply the process of submitting a technical paper to an academic conference to a student seminar. This way, students could become acquainted with the steps in the academic publication cycle. Therefore, participants first prepared their papers on a selected topic. The submissions were then peer reviewed in terms of content, academic quality, and presentation by at least two other seminar participants and one of the editors. The resulting 16 papers are all of high quality and are printed in these proceedings. We are especially content that almost all authors chose to follow our suggestion to prepare their papers in English. At the seminar's mini conference in February 2010, each author presented his paper, often followed by interesting discussions and debates.

We would like to thank all authors for their active participation. Without their engagement the seminar would not have been possible in its current form and quality. Moreover, we would also like to thank our two keynote speakers, Michael Weber and Frank Kargl, for their inspiring insights on the future directions of media informatics research.

Ulm, January 2010

*Florian Schaub, Stefan Dietzel, Boto Bako, Bastian Könings,
Elmar Schoch, and Björn Wiedersheim*

Intelligent Environments

Timo Müller – {firstname.lastname}@uni-ulm.de

Abstract—In intelligent environments, many different technologies come together, obeying several concepts. Lots of sensors measure data and then transmit it to a processing unit which is then making decisions based on previous and present information received. Prediction and privacy are two keywords in intelligent environments, as this shall support the user in his every day life. Large amounts of personal data are being produced and have to be managed while many different devices communicate directly with each other as well as indirectly via internet. This paper briefly presents different methods and approaches to meet the demands which are claimed in smart environments.

I. INTRODUCTION

THE alarm rings at 7.16 am although Paul did not set up anything explicitly. But as he has a meeting at 10.00 am and needs 66 minutes in average from getting up, taking a shower, having breakfast and travelling to his office in the city. His intelligent environment decided to wake him up this early because the traffic information mentions a traffic jam because of an accident on the highway. After Paul left, the heating turns off in order to save energy. Half an hour before Paul comes back home, it is turned on again in order to have exactly 21.3°C at his arrival, just as Paul likes it.

This is a simple and imaginably scenario of an intelligent environment which is able to predict its inhabitant's behaviour, receive the latest relevant news from an objective source and finally acts according to the users preferences. In this paper, intelligent environments will be presented and the most important elements will be discussed. It consists of Technologies (IV) and Concepts (V), followed by a brief overview of Existing Projects (VI). The technology section is separated into sensors, networking, interaction and middleware. Within the concepts section, context awareness, adaptability, decision-making, prediction and security, privacy and trust are discussed.

The paper starts with a general motivation, followed by a definition section where the expression “Intelligent Environment” will be clarified.

II. MOTIVATION

The reasons for the effort that researchers put in the development of intelligent environments are quite obvious. An intelligent environment, private home or public workplace, is a quite old vision as it shall support the user in everyday life. Not only regarding comfort but also concerning efficiency, an environment which “thinks” ahead means a substantial improvement. With automated control of heating, lighting, air conditioning and energy consumption in general, lots of energy and by that money can be saved and nature protected.

III. DEFINITION

As there are many interpretations of the expression “intelligent” in the technical context, it is necessary to define “Intelligent Environment” for this paper. Although Monekosso, Remagnino and Kuno[15] prefer defining by giving examples, they mention that an intelligent environment “is a living or working space that interacts in a natural way and adapts to the occupant”. Here, adapting means that the system should not only react to the different persons using a room and actually present in it, but also consider influences from the outside like seasonal changes. Furthermore, learning is a part of adaptability.

Additionally, Das and Cook[7] define an “Intelligent Environment” “as an environment which is able to autonomously acquire and apply knowledge about the environment and adapt to its inhabitants’ preferences and requirements in order to improve their experience”. The definition is focused on the users’ experience, which may vary among different users. A typical example is the resource optimization, as everyone uses different things at different times, this means high adaption. All the different existing definitions have one thing in common: The basis of an intelligent environment are communicating sensors for the input, effectors as output and a middleware connecting in- and output to the system.

IV. TECHNOLOGIES

As the environment should always be aware of the inhabitants’ position in order to react in line with their intentions, there is a need of a system recognizing these intentions. But not only the person’s localization is of interest; also the surrounding conditions must be known to keep the environment comfortable. Information about temperature, brightness, time, humidity and many more must be measured. In order to react on changes or adjust the conditions, if they do not fit with the inhabitant’s preference, the information must be communicated and processed. As a result, the system can make decisions about how to improve the situation.

A. Sensors

Humans experience their environment with the known five senses. In the same way, computer systems also need special sensors to perceive the surroundings and to determine whether it is comfortable or not, although a more abstract view is necessary. Depending on the tasks the intelligent environment should fulfill, different kinds of sensors have to be integrated. In the following only a few of these tasks will be analysed in detail.

- Locating the inhabitants can be done in different ways. Some of the already implemented techniques and projects are shown in table IV-A.

The most obvious one might be the installation of cameras which recognize the individuals and observe or track them. At least, this can tell if a person is in a specific room or not. If more precise information is required, a multiple camera system computing the position on a map - which even can be created by such a camera system - is possible. Vision provides further properties other approaches do not have. Without further devices, the relative and absolute orientation are detectable, enabling better prediction of what a person is likely to do.[2][8] Another approach of locating the inhabitant is the so called "Smart Floor"[17] developed at the Georgia Institute of Technology and installed in the Aware Home Project (see section VI). The footsteps are recognized and tracked, so that identification and localization becomes possible without a camera.

Using a supersonic system can allow tracking wheelchairs, as long as there is a badge installed. In a nursing scenario as described by Hori[10], the persons are expected to be close to their wheelchair, especially in cases where the persons are handicapped.

- Regulation of temperature depending on the inhabitant's preferences means, that the rooms can be heated or cooled, just like using air conditioning in cars. Of course, one or more chips sensing the temperature are essential.
- Regulation of Brightness is useful in many scenarios, in which automated light control is necessary. For example a person likely to watch a movie at noon and therefore needs a darker room. Or in another example, the need of a toilet visit during night with having difficulties in finding the light switch. In the second case, a sensor registering movements might be necessary as well. More situations can be found easily, in which automated regulation of light makes sense.
- Detecting one persons' condition is especially in remote-care use-cases important. Infrared sensors could give information about a person's body temperature and provide information, whether he or she has a temperature. The already mentioned sensing floor could tell about the weight of individuals and provide additional information about the person's health status. Regarding a hospital, devices to measure pulse, blood pressure and other vital signs could be connected to each other and an emergency system and trigger an alarm in case of irregularities. Each of these examples needs more information of other sensors (movement detector) or devices (TV) to lead to a proper (re)action. This applies to all sensors, as they only gather data. Taking this into account, those receptors are usable for fine tuning of programmed settings or a reason for specific actions. However, it is necessary to communicate the information to something that actually is able to act.

B. Networking

The wireless communication between the several different devices is an essential part of intelligent environments. Every device, which is capable of wireless communication, can connect to another one. This enables wireless sensor networks, like they are used in smart environments. There are mainly two different types of networks:[3]

- The Wireless Local Area Network(WLAN) is the most powerful technique concerning wireless communication as it has a range of up to 500 meters and allows high data rates. To grant this, the devices consume a high amount of energy compared to devices with lower data rate and range. Within the WLAN technology there are different standards which provide different bandwidths depending on the encoding and modulation that is used. Anyway, the complex and extensive topic WLAN will not be discussed here furtherly, as it would extend the papers amount.
- The so-called Wireless Personal Area Network(WPAN) is the counterpart to the wide-ranging WLAN. Its range is much smaller and the power consumption is due to that quite low. This enables dynamic data sharing with devices being close to each other with low data rates. A representative for WPANs is Bluetooth. Bluetooth allows different ranges: 1 meter, 10 meter and 100 meter (approximately). In addition to Bluetooth there is ZigBee which is even more power saving and offers ranges up to 75 meters.
- Besides wireless communication, wired communication via LAN or PowerLine are available as well. In general, they provide higher data rates. As PowerLine uses the cables used for power supply, no additional cables have to be implemented in a building, but special equipment at the ends of the connection is needed. Interferences caused by the normal energy usage and the switch-on and switch-off peaks have to be considered. Because of the dependency on the cables, wired communication is inflexible and static.

These different possibilities of wired and wireless communication allow to choose the range and data rate needed and only to spend as much energy as necessary and, as a result, provide the basement for a flexible ubiquitous computing in smart environments.

C. Interaction

A further basic feature of intelligent environments is control. Combined with the possibilities given by communication technologies, even remote control is possible. Already simple techniques like time switches to automatically switch on or off plugged-in devices enable increasing of comfort. Everything can be turned on and off like a television - a great chance especially for persons with disabilities.[6]

Although this already is remote control, the person still has to be at least around to use the switch in his hands. One step

Table I
LOCATION DETERMINATION TECHNOLOGIES AND USE OF SYMBOLIC/GEOMETRIC COORDINATES[1]

Product/Research Prototype	Primary Goal	Underlying Physical Technology	Techniques Employed	Location Representation
GPS	Outdoor tracking	RF	Triangulation	Geometric
Active Badge	Indoor tracking	Infrared	Vicinity-based reporting	Symbolic
Active Bats	Follow-me indoor computing	Ultrasonic	Paging	Geometric
Cricket	Indoor location tracking	RF and ultrasonic	Location updates	Geometric/Symbolic
RADAR	Indoor location tracking	802.11 WLAN	Triangulation, location updates	Geometric
3D-iD	Indoor location tracking	Active RFID	Triangulation, paging	Geometric
LANDMARC	Indoor location tracking	Active RFID	Multilateration, minimum distance estimation	Geometric
Smart floor	Indoor user tracking	Foot pressure	Location updates	Geometric

further to fully distributed settings and configurations is allowed by using the internet. For example, a nurse care service could be able to observe the inpatient with installed cameras and control the environment with its desktop application.

Of course, direct control might be desirable as well. It might provide a better feedback and is more robust. An already wide-spread control device is the touch screen. It can often be found as a light control and enables intuitive input and feedback. Compared to a normal computer monitor, it is not limited to switching the light on or off but it can offer lots of different services, depending on the systems' implementation.

However, not only controlling the environment should be regarded here. Interaction in an intelligent environment should proceed subconsciously, passively and naturally. The persons interact by behaving normally and the system detects intentions. Hence, input could occur through the recognition of speech. Many applications already deal with the problem to process spoken dialogues. Still, this interaction would happen in an explicit manner if the system is not analysing every discussion made between the occupants.

Implicit interaction could be done by recognizing gestures. Comparable to speech recognition, the recognition of gestures is not a new idea. However, gestures are mainly detected on touch screens or desktop applications by tracking the finger or mouse. The recognition of natural gestures and even mimic could be realised with a camera system, that is already used for locating the user. analysing the movements of the person in context to time, place, heuristics and history, the intention could be made up. The prediction of the occupant's undertaking will be discussed in more detail in section V-B.

In general, another possibility of interaction and control is to set up schedules, for example to let turn the light on during holidays in the evening pretending somebody being at home. Linking sensors, networking, interaction and programs together and enabling integration, reliability and scalability requires software, which supports the system - widely known as "middleware".

D. Middleware

The topic middleware itself could fill books, so that just the very basics are mentioned here.

According to G. M. Youngblood[19], middleware "is connectivity software that joins applications through communication mechanisms creating transparency, scalability, and interoperability" which is conceptually arranged "between the software applications, it assists and the platform it is based

on". Thus, the middleware has to be regarded separately and independently from the application layer, as it interacts in a level of higher abstraction.

The middleware's task is to process the information received by the sensors to an application, which then decides, whether a reaction should occur or not. For that, some decision making artificial intelligence software is needed. Concepts of that will be discussed later.

Hence, middleware can be seen as an interface which provides easier access to complex applications. In general, middleware facilitates the communication between client/server or within a peer-to-peer network. The difficulty regarding an intelligent environment is the variety of devices and services in addition to demands concerning real time and data flow.

A specific example, which allows a better understanding of middleware in intelligent environment is the EasyLiving project in section VI-E.

V. CONCEPTS

Now knowing the conditions and being able to change them theoretically and practically, the system has to know how the persons' being "cared of" like to have their home or working place. First of all that implicates never-ending learning, being aware of the situation and not only of the physical conditions and even being able to predict which action the person is likely to undertake.

A. Context Awareness

To be able to behave properly, that means to predict the user's intentions, support him in his action and record the situation in order to learn, a system actually must know the situation or be aware of it. Context-Awareness has very much to do with locating - which was already basically discussed in section IV-A and interpreting interactions that already happened.

A smart environment can run specific defined programs as already mentioned before. For example turning on the heater in the morning, leading to a comfortable temperature at wake-up time, is such a program. This is increasing comfort for the inhabitant, as long as the inhabitant is at home. Otherwise it would be a waste of energy. Thus, an improvement of the program management becomes necessary. This improvement is gained by the use of ruled-based decisions, where actions are accomplished only if a set of conditions is met [11]. Thereby, simple actions can be performed. An early version of the

iSpace-Project (see section VI), for example, implemented that the light is switched on when someone comes into the room. Now, the light is switched on any time somebody enters the room, independent of the amount of daylight coming into the room.

It is obvious, that “one of the main problems of ruled-based systems is, that they [...] require large numbers of [...] rules”[11], which implies, that they need manual construction by the inhabitant or a programmer and by that are only beneficial for small problems. Intelligent behaviour takes more than just knowing a person’s location and applying some rules.

B. Adaptability, Decision-Making and Prediction

As the preferences, even simple ones like lighting and heating, differ between the potential users, it is impossible to match them in advance. Because of that, the techniques that already have been discussed do not provide optimal control and the technology of machine learning is necessary. Learning here means, that the machine - or better the software - has to be trained by the user who finally has to deal with it.

For being able to avoid models and trainers, “reinforcement learning” seems to be an appropriate way of learning. It is “a learning framework in which action policies are learned from interactions with the environment and a scalar reward function”[11] and relies on feedback giving information about the current performance. The reward function - which Mozer[16] derives from the energy cost and the so called “discomfort cost”, given by the manual interactions of the user - leads to an optimal system, as the target (saving energy) and the comfort are regarded and the perfect compromise is aspired.

Decision-making and prediction are quite close to each other and the expressions are sometimes used interchangeably. While decision-making might be understood as a reaction on a situation happening right now based on some information, prediction is rather guessing or better knowing, what the person will do next and for that preparing the environment.

An example for decision-making might be that the light is turned on in the morning, when someone enters the bathroom. Prediction however would recognize the alarm clock ringing and switching on the bathroom light before the person is about to enter the bathroom.

Again, to predict someone’s behaviour, recording his acting history is necessary and there is always still an uncertainty if he really will do the same again. Hence, the demands on a prediction algorithm are of course a minimal number of prediction errors, ability of learning and computational speed for real time usage.[4]

Learning to predict the inhabitant’s intentions is possible, as humans have certain patterns in their daily routine, which do not change randomly and frequently. Of course not all the intentions can be predicted, but some actions like stepping into the bathroom, taking a shower and drinking some coffee occur every day and are prepareable. In multi-inhabitant homes, predicting a single person’s behaviour is much more challenging because the “relevant contexts of multiple inhabitants [...] are

often inherently correlated and thus interdependent with each other”, as S.K. Das[6] mentions. She also proves that multi-person-tracking is an NP-hard problem.

For that reason, many projects only research the single-inhabitant case so far.

C. Security, Privacy and Trust

In an intelligent environment, special privacy demands apply. This is a consequence to the fact, that within intelligent environment lots of communication, especially wireless, occurs. In many cases - like the ordering refrigerator, remote or distributed control or nurse care services and so on - even communication exceeding the borders of the house through the internet takes place.

This involves secure communication and trust, just like in “normal” wireless networks, too. But it is also possible, that critical information might be transmitted, requested or obtained by cracking the system. Critical information refers to those who allow for conclusions on peoples’ habits like the amount of energy or water consumed or even data stored for prediction reasons.

According to Nixon et al. [18], privacy topics in intelligent environments are different than in general cases:

- Ubiquity
The infrastructure will be everywhere, consequently affecting every aspect of live
- Invisibility
The infrastructure will be cognitively or physically invisible to the users; the users will have no idea of when or where they are using the computer
- Sensing
Input to the ever-present invisible computer will consist of everything the user does or says, rather than everything the user types
- Memory amplification
Every aspect of these interactions, no matter how personal, has the potential to be stored, queried, and replayed

In the following, the three major aspects announced in the sections’ headline, security, privacy and trust will be described more closely, without digging to deep in the material, as this paper only wants to provide an overview.

Security

General techniques of the security aspect also apply in intelligent environments. These are encryption, decryption and authentication.

As the principals¹ of an intelligent environment are of a dynamic and decentralized nature, standard authentication with Public Key Infrastructure (PKI) relying on certification authority has limitations due to a fixed hierarchy. The “Pretty Good Privacy” project (PGP)² approach is one of a “web of trust”, where users sign the key certificates of other users. This approach is more generic than the first mentioned hierarchal

¹Principal is the term used in security to signify the entities (people, agents, devices, etc.) of interest[18]

²<http://www.pgpi.org/>

approach, but due to that more difficult to manage. Also PGP is not a solution for the security problems in intelligent environments, as “trusting a key is not the same as trusting a key’s owner”³.

In addition to general security aspects in intelligent environments, the physical nature of wireless communication offers possibilities for attacks. This is especially the case as protocols neglect to implement security features or when they are not resistant enough, which can be noticed in the Bluetooth technology[12].

Furthermore, security matters for the devices that have to be treated. Mobile devices join the environment, which has no information about the devices’ previous domain. Even a given device which has been absent could have been modified and thus be harmful. In case a device is getting stolen, not only the device itself but even its identity gets stolen. Regarding this, the integrity of devices can be questioned. With respect to sensors, it is not only necessary to control and grant the correctness of the data transmitted. It is also required to prevent masking the data a sensor actually measures.

Privacy

A lot of personal data is collected and stored in an intelligent environment, so that the risk of misuse by malicious and curious users is immense. Thus the danger of scenarios like Big Brother of George Orwell’s “1984” come into mind. Seven principles concerning privacy have been made by Langenheinrich[13]:

- Openness and transparency
There should be no secret record keeping.
This includes both the publication of the existence of such collections as well as their contents
- Individual participation
The subject of a record should be able to see and correct the record
- Collection limits
Data collection should be proportional and not excessive compared to the purpose of the collection
- Data quality
Data should be relevant to the purposes for which they are collected and should be kept up to date
- Use limits
Data should only be used for their specific purpose by authorized personnel
- Appropriate security
Adequate security safeguards should be put in place, according to the sensitivity of the data collected
- Accountability
Record keepers must be accountable for compliance with the other principles

The Platform for Privacy Preferences Project⁴ allows the user to define his own privacy preferences in the internet, as long as the browser supports it. There is also further software that alleviates privacy management under development, which will not be focused on in this study. Anyway, these approaches

have not been successfully implemented in intelligent environments. One reason seems to be the bidirectional relationship between the different users and difficulties in balancing all the privacy requirements with the whole functionality of the environment.[18]. Again according to Langenheinrich, there are six guidelines for a design of a ubiquitous computing system like an intelligent environment:

- Notice
The user has to be well informed so that he can make the privacy decisions. The so-called “awareness infrastructure” is a baseline technology for smart environments’[18]
- Choice and Consent
Choice and Consent are main aspects in individual control. The user should have the choice, whether data is stored or not and if the user interacts, his agreement is necessary.
- Anonymity and Pseudonymity
On the one hand, by anonymity, explicit consent seeking can be avoided. On the other hand, the aim of personalized interactions is suspended
- Proximity and Locality
Data should only be provided to those parties, that match specific rules which in case are limited distances from the source
- Adequate Security
Encryption has to be used. Due to the different capabilities of devices, the use of encryption should be “adequate” and proportional
- Access and Recourse
This guideline refers to collection limit, use limits and accountability and is rather of a process or purpose guideline than a technology guideline.

These guidelines have been proposed concerning possible accidental invasion in everyday use and not intended malicious attacks.

Trust

It was already mentioned before that integration of devices which have been in an unknown environment before - and in the same way the integration of new users - carries a certain risk. In order to profit from the whole range an intelligent environment offers, it is essential to provide secure autonomous decision-making and thus to automatically determine whether that device can be trusted or not. “Trust” is defined by Blaze et al.[14] as “a unified approach to specifying and interpreting security policies, credentials and relationships that allow direct authorization of security-critical actions”. As the composition of devices within an smart environment is flexible and changes dynamically, hard-coded approaches are unsuitable. Thus it is necessary, to make up a trust-management-system which is able to handle the flexibility. The use of credentials however, implicates that somekind of fixed security infrastructure exists. This is not in accordance with the idea of a flexible and dynamic intelligent environment.

³Phil Zimmermann, creator of PGP, quoted in [18]

⁴<http://www.w3.org/P3P/>

Nixon et al.[18] discuss some ideas of how to meet the demands. However, they state, that the “existing trust management systems are not sufficient because they do not address fine-grained, dynamic trust evolution”. This is the reason why these approaches are not discussed here.

VI. EXISTING PROJECTS

As mentioned in the introduction, defining an intelligent environment can potentially be done easier by giving specific examples. Within the last sections of this paper, samples for elements of intelligent systems with concrete scenarios have been interspersed.

Following below, a few projects presented by Monekosso et al.[15] which do research on smart home environments will be introduced briefly.

A. iSpace / iDorm

This project at Essex University (UK)⁵ rebuilt a typical dormitory room and implemented sensors measuring temperature, humidity, motion and brightness. A video camera enables monitoring the room. These possibilities for input already allow a satisfactory amount of output. Therefore, embedded effectors are air circulators, fan heaters, motorized vertical window blinds, automated window openers and light dimmers. An active door lock only provides access to the right person being identified.

Meanwhile the project’s successor⁶, iSpace2, has been established. The new room is much larger and its technical equipment has been modernized. An example is a special wall that hides the heterogeneous network infrastructure from views.

B. Adaptive Home

While iSpace is only inhabited by a single person, the Adaptive Home⁷ tries to develop itself observing the multiple inhabitants. 75 sensors deliver information and a complex control system, based on a neural network that uses reinforcement learning as mentioned in section V-B, which allows to predict the inhabitants’ behaviour to effectively decide how to act in a desired manner.

C. MavHome

At the University of Texas at Arlington and the Washington State University⁸, iSpace and the Adaptive Home are taken a step further. MavHome collects health data of its inhabitants and analyses them long and short termed. Very much of what was told about prediction and decision-making in section V-B has been developed there. Thus, in a simulated evaluation on MavHome, the system was fed with simple artificial data in special periodically repeated constellations during a 30-day trial. After cleaning the results from noise, the filtered accuracy reached 100% [5]. It is not told, what kind of data has been

⁵<http://iieg.essex.ac.uk/idorm.htm>

⁶<http://iieg.essex.ac.uk/idorm2/>

⁷<http://www.cs.colorado.edu/~mozer/house/>

⁸<http://ailab.wsu.edu/mavhome/>

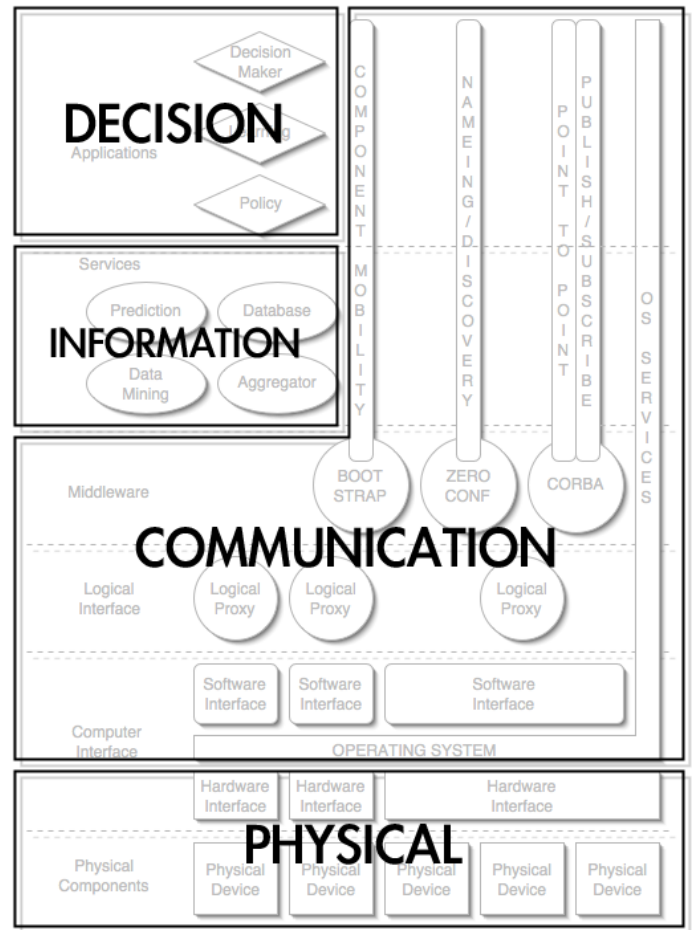


Figure 1. MavHome Abstract Architecture [9]

erased. In real scenarios, this accuracy can not be reached. Anyway, other features of MavHome are the refrigerator, which knows its content and orders new food, one robot which cleans the floor while one cuts the grass and the TV that automatically records shows which the inhabitants might be interested in. The architecture is built up on open source software and designed modularly over a monolithic system. It consists of four layers, shown in Figure 1, physical layer (sensors, actuators, network), communication layer accessible by all other layers (middleware), information layer (knowledge gathering and storing) and decision layer (learning).

D. Aware Home

Since 1998, the Aware Home Project⁹ demonstrates a smart environment, which is focused on health monitoring. As a consequence, several methods of displaying health information have been developed. Aware Home also provides easy sharing of media content. It is embedded in a building with an approximate size of about 500 m². In section IV-A it has already been mentioned, that there is a SmartFloor installed in the Aware Home.

⁹<http://awarehome.imtc.gatech.edu/>

E. Easy Living

In contrast to the projects above, Easy Living[2] at Microsoft Research is not an independent smart environment which implements middleware but moreover an approach of developing a generic middleware which is reusable in different environments. As a result, it can be seen as “a software toolkit to construct intelligent environments” [15]. Different concepts are analysed and improvements are realised. A special communication approach, an own geometric model and a special person tracking system relying on visual data have been developed.

VII. CONCLUSION AND CRITICAL VIEW

In this paper, the techniques and concepts of an intelligent environment have been presented. The current state of research shows, that we are able to build powerful networks of sensors and effectors in order to facilitate the inhabitants’ actions. In single-inhabitant cases, good prediction of basic and periodically repeated actions can be obtained. With more people to be observed and supported, the current algorithms and approaches seem to reach their limits quickly, whereas improvements have been made recently.

However, it is obvious that there is definitely still a lack of security and privacy in such systems. The deficits have to be eliminated before smart environments can be applied. The potential risk of malicious invasion in a network of many wireless devices is still given, especially as the system processes lots of critical personal data in order to predict and act properly.

Last but not least, the peoples’ acceptance of a “thinking” environment that knows them perfectly well and communicates invisibly and independently, is still doubtful. As the environment needs a system to trust the devices, the user needs to trust the system, where the security problems again are brought into account negatively.

If it is possible to grant security and privacy and to guarantee, that a user is not limited in his activities but supported by an intelligent environment, this is the future in modern housing.

REFERENCES

- [1] S. A. Misra, “Location estimation (determination and prediction) techniques in smart environments,” in *Smart Environments: Technology, Protocols and Applications*. John Wiley and Sons, 2005, pp. 214–228.
- [2] B. Brumitt, B. Meyers, J. Krumm, A. Kern, and S. Shafer, “EasyLiving: technologies for intelligent environments,” in *Handheld and Ubiquitous Computing*, 2000, pp. 97–119.
- [3] M. Conti, “Wireless communications and pervasive technology,” in *Smart Environments - Technologies, Protocols and Applications*, 2005, pp. 63–99.
- [4] D. Cook, “Prediction algorithms for smart environments,” in *Smart Environments: Technology, Protocols and Applications*. John Wiley and Sons, 2005, pp. 175–192.
- [5] D. Cook, M. Youngblood, E. Heierman, K. Gopalratnam, S. Rao, A. Litvin, and F. Khawaja, “MavHome: an agent-based smart home,” in *Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on*, 2003, pp. 521–524.
- [6] S. Das and D. Cook, “Designing and modeling smart environments,” in *World of Wireless, Mobile and Multimedia Networks, 2006. WoWMoM 2006. International Symposium on a*, 2006, pp. 5 pp.–494.
- [7] S. Das and D. C. et al., “The role of prediction algorithms in the mavhome smart home architecture,” in *IEEE Wireless Communications*, 2002, pp. pp. 77–84.
- [8] R. Elias and A. Elnahas, “An accurate indoor localization technique using image matching,” in *Intelligent Environments, 2007. IE 07. 3rd IET International Conference on*, 2007, pp. 376–382.
- [9] L. H. G.M. Youngblood and D. Cook, “The mavhome architecture.” University of Texas at Arlington, 2004.
- [10] T. Hori, Y. Nishida, and S. Murakami, “Pervasive sensor system for evidence-based nursing care support,” in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, 2006, pp. 1680–1685.
- [11] M. Huber, “Automated decision making,” in *Smart Environments: Technology, Protocols and Applications*. John Wiley and Sons, 2005, pp. 229–247.
- [12] M. Jakobson and S. Wetzel, “Security weaknesses in bluetooth,” in *Proceedings of the RSA Cryptographer’s Track (RSA CT ’01)*. Springer-Verlag, 2005, pp. 176–191.
- [13] M. Langenheinrich, “Privacy by design—principles of privacy aware ubiquitous systems,” in *UBICOMP 2001*. LNCS 2201, 2005, pp. 273–291.
- [14] J. M. Blaze and J. Lacy, “Decentralized trust management,” in *Proceedings of the 1996 IEEE Symposium on Security and Privacy*, 1996, pp. 164–173.
- [15] Monekosso, Remagnino, and Kuno, *Intelligent Environments - Methods, Algorithms and Applications*. London: Springer-Verlag, 2009.
- [16] M. Mozer, “Lessons from an adaptive home,” in *Smart Environments: Technology, Protocols and Applications*. John Wiley and Sons, 2005, pp. 273–294.
- [17] R. Orr and G. Abowd, “The smart floor: A mechanism for natural user identification and tracking,” in *Proc. of ACM Conf. on Human Factors in Computing Systems*, 2000.
- [18] C. P.A. Nixon, W. Wagealla and S. Terzis, “Security, privacy and trust issues in smart environments,” in *Smart Environments: Technology, Protocols and Applications*. John Wiley and Sons, 2005, pp. 249–270.
- [19] G. Youngblood, “Middleware,” in *Smart Environments: Technology, Protocols and Applications*. John Wiley and Sons, 2004, pp. 101–127.

Ambient Assisted Living

Nadhem Kachroudi – {firstname.lastname}@uni-ulm.de

Abstract—Laut demographischen Statistiken wird der Anteil älterer Menschen in der Bevölkerung weiter zunehmen. Diese unterstützungsbedürftigen Personen bevorzugen ein längeres selbstständiges Leben in ihren eigenen Wohnungen gegenüber der dauerhaften Unterbringung in Pflegeheimen. Mit dem Einsatz von neuen innovativen Technologien aus dem Forschungsgebiet Ambient Assisted Living, wollen Elektroingenieure mithilfe von Wissenschaftlern aus anderen Branchen ihnen diesen Wunsch ermöglichen, sowie weitere wirtschaftliche und soziale Ziele erreichen. Einzelne Lösungen gibt es bereits, doch die eigentliche Herausforderung liegt darin diese in eine intelligente vernetzte Infrastruktur mit standardisierten Schnittstellen zu überführen.

I. MOTIVATION UND ZIELSETZUNG

DANK der besseren Lebensqualität und dem medizinischen Fortschritt ist die Lebenserwartung länger geworden. Die Geburtenrate ist aber in den letzten Jahren weiter gesunken und die Bevölkerung wird nach statistischen Angaben immer älter [5]. Da ältere Menschen für bestimmte Krankheiten (Alzheimer, Demenz, ...) anfälliger sind, sind Krankenkassen, Krankenhäuser und Angehörige zunehmend überfordert, und können den Betroffenen die nötige Unterstützung und Pflege immer weniger gewährleisten.

Ambient Assisted Living (AAL) ist ein interdisziplinäres Forschungsgebiet, das sowohl von diesem soziodemographischen Wandel als auch dem Wunsch älterer Personen ihre eigene Wohnungen nicht Richtung Pflege- und Altenheime zu verlassen vorangetrieben wurde. Ein wesentliches Ziel ist es, mittels einer unaufdringlichen, im Hintergrund laufenden intelligenten Infrastruktur, pflegebedürftigen Menschen ein längeres, sicheres und selbstbestimmtes Leben im eigenen Wohnumfeld zu ermöglichen. Unter AAL werden „Konzepte, Produkte und Dienstleistungen verstanden, die neue Technologien und soziales Umfeld miteinander verbinden und verbessern mit dem Ziel, die Lebensqualität für Menschen in allen Lebensabschnitten zu erhöhen“ [6]. Laut dem Bundesministerium für Forschung und Bildung (BMBF) könnte man AAL am besten mit „Altersgerechte Assistenzsysteme für ein gesundes und unabhängiges Leben“ übersetzen [1].

Im nächsten Kapitel werden die vier Entwicklungsbereiche des AAL näher beschrieben. Das dritte Kapitel befasst sich mit den Anforderungen an die Assistenzsysteme, zum Beispiel Skalierbarkeit, Erweiterbarkeit und Vernetzung sowie Anforderungen an die verschiedenen Akteure wie beispielsweise Akzeptanz, Vertrautheit und Kommunikation. Im vierten Kapitel werden relevante Forschungsgebiete vorgestellt und im fünften Kapitel ein Überblick über einige vom BMBF geförderte AAL-Projekte gegeben. Abschließend wird im letzten Kapitel ein fertiges Produkt näher betrachtet.

II. AAL-ENTWICKLUNGSBEREICHE

Laut dem BMBF kann man zwischen vier Entwicklungsbereichen des Ambient Assisted Living unterscheiden, wobei eine klare Trennung dieser Bereiche wegen der ineinander übergehenden Konzepte des AAL weder möglich, noch gewollt ist. Diese sind „Gesundheit und HomeCare“, „Sicherheit und Privatsphäre“, „Versorgung und Haushalt“ und „Soziales Umfeld“ [6] [1].



Abbildung 1. AAL-Entwicklungsbereiche [1]

A. Gesundheit und HomeCare

In diesem Bereich werden Anwendungen zur Fernüberwachung, -diagnose und -rehabilitation sowie Präventionstechniken entwickelt. Hierbei sollen intelligente Assistenzsysteme den bedürftigen Menschen die Möglichkeit geben, zwischen den eigenen vier Wänden in einer unterstützten Umgebung selbstbestimmt und unabhängig zu leben, statt von einer direkten ständigen Betreuung durch Angehörige und Freunde oder einer stationären Unterkunft in Pflegeheimen abhängig zu sein. Neben dem geringeren Bedarf an externer Unterstützung, werden die Pflege- und Gesundheitskosten sowohl für den Patienten als auch die Gesellschaft dadurch reduziert [6] [1].

Verschiedene Dienstleistungen, zum Beispiel eine Fortbewegungsunterstützung, eine Erinnerungsfunktion an Medikamenteneinnahme oder ein mentales Training können in diesen Systemen integriert werden. Mithilfe der mit Sensorik ausgestatteten Umgebung und miteinander verknüpften Mess- und

Kommunikationsgeräten werden wichtige Informationen des Patienten wie Vital-, Präsenz- und Bewegungsdaten überwacht. Damit sollen Erkrankungen oder eine Verschlechterung des Gesundheitszustands frühzeitig verhindert werden, sowie gegebenenfalls Notfälle beispielsweise ein Sturz erkannt und entsprechende Alarmierungs- und Hilfemaßnahmen rechtzeitig ergriffen werden.

B. Sicherheit und Privatsphäre

Um den Personen mit Unterstützungsbedarf eine sichere Lebensführung bieten zu können, soll die unterstützte Umgebung außerdem mit zuverlässigen Schutzsystemen ausgestattet sein. Hierfür kommen verschiedene „Security und Safety“ Anwendungen zum Einsatz. Die Zugangsberechtigung könnte z.B. über Fingerabdruck oder Gesichtserkennung kontrolliert werden. Alarmfunktionen für Feuer, Wasser und Gas, sowie automatische Notrufe und Bewegungs- bzw. Präsenzmelder sorgen für eine größtmögliche Sicherheit des Patienten. Darüberhinaus ermöglicht die vernetzte technische Infrastruktur der Wohnung und des Wohnumfeldes die Fernüberwachung und -steuerung dieser Anwendungen durch Sicherheitsunternehmen [6] [1]. Zum Schutz der Privatsphäre wird es in einigen AAL-Lösungen auf Kameras und Mikrofone verzichtet [9]. Die Patientendaten sollen außerdem lokal verarbeitet und gespeichert werden und sie werden nur über einen sicheren Kanal übertragen.

C. Versorgung und Haushalt

Zu diesem Bereich gehören Assistenzsysteme und Geräte, die älteren und pflegebedürftigen Menschen helfen können, ihr Haushalt selbständiger und leichter zu führen. Staubsauger- bzw. Reinigungsroboter, die selbsttätig Räume reinigen können, ferngesteuerte oder vollautomatische Klimaanlage oder Automatik-Programme von Haushaltsgeräten gibt es seit Jahren und gehören mittlerweile zum Alltag vieler Personen. Die Forschung und Entwicklung in dieser Kategorie befasst sich heutzutage mehr mit der Vernetzung von bereits existierenden sowie neuen Geräten und Anwendungen zu einer intelligenten, erweiterbaren und benutzerfreundlichen Struktur, die an den Bedürfnissen des Benutzers angepasst werden kann. Stichworte in diesem Zusammenhang sind intelligentes Wohnen oder SmartHouse [6] [1].

D. Soziales Umfeld

Kommunikationssysteme für die Unterhaltung mit Verwandten und Freunden, spezielle Geräte, die die Mobilität des Patienten unterstützen, Anwendungen, die digital gestützte Nachbarschaften ermöglichen oder Einrichtungen für eine aktivere Freizeitgestaltung lassen sich diesem Bereich zuordnen. Das Ziel der Entwicklungen in diesem Bereich ist die Stärkung der sozialen Beziehungen alleinlebender älterer oder in Mobilität eingeschränkter Menschen. Leichtbedienbare, in der häuslichen Infrastruktur vernetzte AAL-Systeme sollen den bedürftigen Personen beim Kennenlernen, bei der Kommunikation und bei der Erhaltung der Kontakte unterstützen um ihre körperliche und seelische Gesundheit zu fördern [6] [1].

III. ANFORDERUNGEN

Für eine erfolgreiche Umsetzung des AAL-Konzeptes müssen sowohl die entwickelten Assistenzsysteme als auch die verschiedenen Nutzer der AAL-Technologien, Dienste und Dienstleistungen folgende Anforderungen erfüllen:

A. Anforderungen an die Assistenzsysteme

1) *Unauffällige Überwachung:* Die Gesundheit des Patienten und dessen Aktivitäten sollen so unauffällig und unaufdringlich wie möglich im Hintergrund überwacht werden. Der Benutzer des AAL-Systems soll seinen Alltag ganz normal führen können, ohne dabei von den integrierten Unterstützungs- und Überwachungseinheiten körperlich oder seelisch eingeschränkt oder gestört zu werden [12].

2) *Abgestuftes Alarmierungskonzept:* Die Assistenzsysteme sollen in der Lage sein, ein abgestuftes Sicherheits- und Alarmierungskonzept zu verfolgen, das sich an die bestimmten Alltags- und gegebenenfalls Notfallsituationen anpassen lässt. Der Benutzer soll beispielsweise bei einer Verschlechterung seines Gesundheitszustandes selber Alarm auslösen können, aber sobald er sich in einem Notfall befindet, zum Beispiel im Falle eines Sturzes oder einer Ohnmacht, sollen automatische Notrufsysteme diese gefährliche Situation rechtzeitig melden. Gefährliche elektronische Haushaltsgeräte, wie Herd oder Bügeleisen sollen z.B. nicht nur die Temperatur mittels entsprechender Signalleuchten anzeigen können, sondern sich auch bei Erkennung von Gefahrensituationen automatisch abschalten lassen. Eingangstür, Fenster, Heizung, Rolladen oder Licht sollen, außer von den gewöhnlichen Stellen, auch vom Bett oder vollautomatisch schließbar bzw. steuerbar sein [12].

3) *Skalierbarkeit und Erweiterbarkeit:* Die profitierenden Zielgruppen von AAL-Produkten variieren je nach Alter, Art der Behinderung oder Erkrankung und nach besonderer Interesse an Pflege, Sicherheit und Komfort. Man unterscheidet z.B. zwischen alten Menschen, in Mobilität eingeschränkten Patienten, an Demenz Erkrankten und chronisch oder akut kranken Personen. Um die Bedürfnisse der verschiedenen Nutzer der AAL-Technologien und -Dienstleistungen befriedigen zu können, müssen die angebotenen Systeme so skalierbar sein, dass sie genau an die Wünsche und Bedürfnisse der Zielperson angepasst werden können.

Außerdem soll die eingebaute intelligente Infrastruktur relativ wartungsfrei und einfach erweiterbar sein, um die zukünftigen Bedürfnisse von Patienten mit zunehmendem Alter oder verändertem Krankungsverlauf weiterhin berücksichtigen und behandeln zu können.

Die Realisierung einer skalierbaren und erweiterbaren unterstützten Umgebung sollte mit der Standardisierung der Schnittstellen zwischen den verschiedenen eingesetzten Einrichtungen vereinfacht werden [6] [12].

4) *Vernetzung und Verknüpfung der Systeme:* Ambient Assisted Living Lösungen bestehen nicht aus einzelnen Geräten, die ihre Aufgaben unabhängig voneinander erledigen, sondern aus intelligenten, miteinander verknüpften Systemen, die zusammen arbeiten und miteinander interagieren, um Sicherheits-, Bewegungs- oder Vitaldaten austauschen zu können, und um wichtige Entscheidungen je nach Bedürfnissen des Patienten zu treffen.

Kommunikationsanwendungen benutzen beispielsweise die von eingebetteten Sensoren erfassten Daten um eine entsprechende Aktion oder Alarmierung auszulösen. Sicherheitsunternehmen sollen die im Wohnumfeld integrierten Schutzsysteme fernsteuern können und Dienstleistungsanbieter sollen in der Lage sein eine externe Wartung oder Reparatur durchzuführen. Hierbei sollen Datenschutz und Privatsphäre berücksichtigt werden. Dies ist nur mittels einer sicheren und zuverlässigen Vernetzung der AAL-Systeme möglich, welche ihrerseits einheitliche und standardisierte Schnittstellen zwischen den zusammenspielenden Einheiten voraussetzen [4].

B. Anforderungen an die Akteure

1) *Akzeptanz und Vertrauen:* Die Akzeptanz der innovativen Dienstleistungen sowie das Vertrauen auf die im Hintergrund agierenden Technologien ist ein wesentlicher Faktor für den Erfolg des AAL-Konzeptes. Um die Zufriedenheit der Benutzer zu erreichen arbeiten Elektroingenieure, Multimediale Designer, Sozialwissenschaftler und Ärzte zusammen um Produkte zu entwickeln, die den Fähigkeiten, Bedürfnissen und Wünschen der pflegebedürftigen Personen angemessen sind. Auf das Vertrauen an die eingesetzte Techniken können die Hausärzte der Patienten durch Empfehlungen und Betreuung Einfluss nehmen, was hier auch eine wichtige Rolle spielt [6].

2) *Vertrautheit:* Die Benutzer der AAL-Technologien und Dienstleistungen, ob Patienten, Ärzte, Familienangehörige, Freunde oder Nachbarn sollen neben Akzeptanz und Vertrauen auch mit den entwickelten Assistenz- und Schutzsystemen gut umgehen können. Damit diese Systeme auch beim zunehmendem Alter oder dem Eintreten einer zufälligen Verschlechterung der Gesundheit leicht und problemlos bedient werden können, ist für die unterstützungsbedürftige Person außerdem eine frühzeitige Vertrautheit von Vorteil und kann mit Komfortdiensten begonnen werden [12].

3) *Kommunikation:* Die audiovisuelle Unterhaltung mit Familienangehörigen und Freunden sowie einige Telemonitoring-Anwendungen erfordern eine gewisse Bereitschaft von allen teilnehmenden Personen zur aktiven Kommunikation. Die Erfüllung oben beschriebener Anforderungen unter anderem Akzeptanz und Vertrautheit werden für eine einfache Kommunikation mit den angebotenen AAL-Produkten vorausgesetzt.

IV. RELEVANTE FORSCHUNGSGEBIETE

Ambient Assisted Living (AAL) ist ein interdisziplinäres Forschungs- und Entwicklungsgebiet. Medizintechnik, Mikrosystemtechnik (MST) und Informations- und Kommunikationstechnik (IKT) sind nur eine Auswahl unter mehreren Wissenschaftsbereichen, die zur Entwicklung von AAL-Lösungen beitragen. In den folgenden Unterabschnitten werden diese und weitere Technologien im Zusammenhang mit AAL vorgestellt.

A. Mikrosystemtechnik

Durch die Verwendung miniaturisierter MST-Produkte wird die Entwicklung einer intelligenten Umgebung, die unauffällig auf die Bedürfnisse des Patienten reagiert, vorangetrieben. So können Mikrosysteme nicht nur in Wohneinrichtungen,

sondern auch bequem in die Kleidung oder am Körper des unterstützungsbedürftigen Benutzers integriert werden, und ihn unsichtbar und unaufdringlich unterstützen [4].

Ein gutes Beispiel in diesem Bereich ist die Entwicklung von sogenannten intelligenten Kleidungen (Smart Clothes), die mit Sensoren und anderen elektronischen Geräten ausgestattet sind um physiologische Körperwerte (Herzfrequenz, Bewegung, Temperatur, Feuchtigkeit, ...) über den Träger ständig zu erfassen. Abbildung 2 zeigt die Architektur eines Smart Shirts der Firma Sensatex. Das SmartShirt sammelt die Körperparameter des Trägers mittels verschiedener Sensoren, die an ein Netz aus elektro-optischen Fasern angeschlossen sind. Eine Prozesseinheit digitalisiert die erfassten Daten und sendet sie, beispielsweise via Bluetooth, an eine Basisstation, wo sie angezeigt und/oder gespeichert werden können [10] [13].

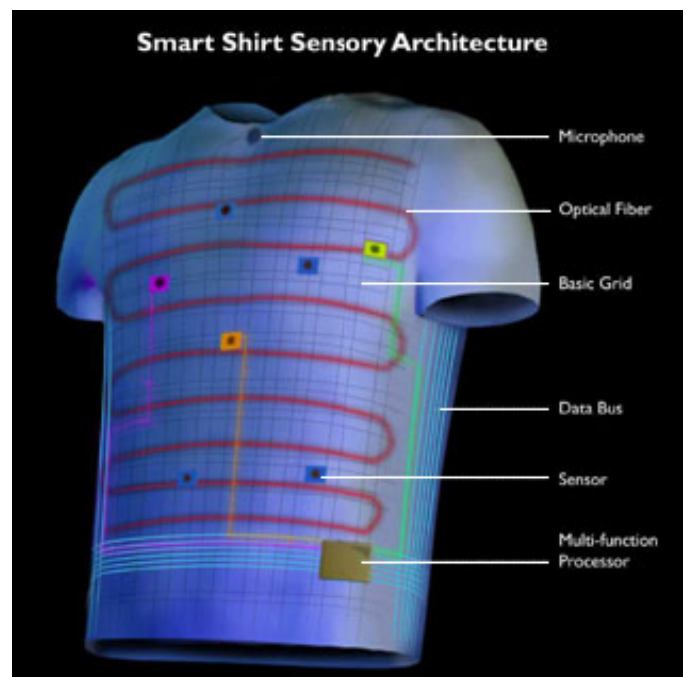


Abbildung 2. Sensatex Smart Shirt [13]

B. Informations- und Kommunikationstechnik

Die Produkte aus dem Bereich Informations- und Kommunikationstechnik stellen zusammen mit denen aus der Mikrosystemtechnik-Branche die Basis für den Aufbau von intelligenten AAL-Infrastrukturen dar, welche hauptsächlich aus vernetzten und miteinander verknüpften Geräten, Sensoren, Kommunikations- und Mikrosystemen bestehen.

Die von IKT-Technologien unterstützte Umgebung soll pflegebedürftigen Personen helfen ihre alltäglichen Tätigkeiten sicher und selbständig durchzuführen, sowie ihre Vital- und Umgebungsdaten, die mittels körpernahe oder räumlich verteilte Sensoren erfasst werden, auf benutzerfreundlichen Displays anzuzeigen oder zu speichern, zu verarbeiten und über sichere Kommunikationskanäle zum medizinischen Dienstleister zu übertragen [4]. Das Zusammenspiel von Geräten, Sensoren und Daten wird in der Abbildung 3 gezeigt.

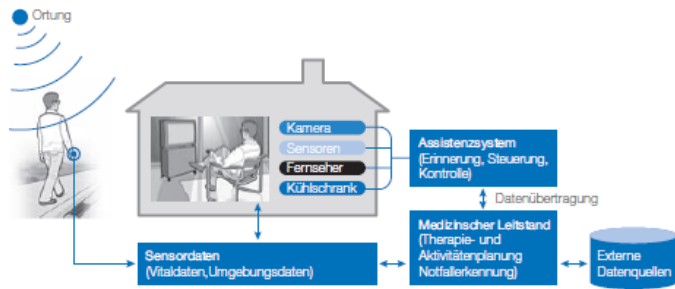


Abbildung 3. Zusammenspiel von Geräten, Sensoren und Daten [4]

C. Intelligentes Wohnen

Bei der Entwicklung von AAL-Systemen werden bereits bekannte Ideen aus dem Forschungsbereich „Intelligentes Wohnen“ oder „Smart Home“ eingesetzt. Beispiele sind Energieeffizienz, Licht- und Heizungssteuerung oder Haushaltsgeräte-Automation und Vernetzung. Natürlich sind nicht alle Anwendungen des Smart Home für das AAL-Konzept interessant, aber die meisten Lösungen, die in der Lage sind dem Patienten Pflege, Sicherheit oder Komfort zu bieten, können in Ambient Assisted Living sinnvoll integriert werden [4].

Ein Beispiel ist das Pilotprojekt der technischen Universität Kaiserslautern in Zusammenarbeit mit der Kaiserslauter BauAG [4]. In diesem Projekt werden zwanzig Wohnungen mit den nötigen intelligenten Gesundheits-, Sicherheits- und Unterhaltungssystemen ausgestattet, welche von einem auf Tablet-PC basierenden Assistenten ferngesteuert werden können. Seine freundliche Benutzeroberfläche, die über ein Touchscreen intuitiv bedient werden kann, ist in Abbildung 4 abgebildet.



Abbildung 4. Persönlicher Assistent für unterstütztes Leben (PAUL) [8]

Dieser Assistent wird persönlicher Assistent für unterstütztes Leben (PAUL) genannt und dient als Steuerzentrale für alle Komfort- und Sicherheitsfunktionen, zum Beispiel für die Bedienung der Rolläden, das Ansprechen der Türkamera, die Erkennung der Gefahrenstufen und gegebenenfalls die automatische Alarmauslösung. PAUL dient außerdem als Unterhaltungsgerät und bietet verschiedene Unterhaltungs- und Kommunikationsfunktionen wie beispielsweise Internet, Fernseher oder Radio. Zusätzlich speichert der Assistent die verschiedenen Sensordaten und interpretiert sie, um das Akti-

vitätsprofil des Patienten zu erstellen und seine Gesundheit zu überwachen [4].

D. Medizintechnik

Medizintechnische AAL-Produkte stehen besonders im Zusammenhang mit Telemonitoring-Systemen zur Fernüberwachung, -diagnose und -untersuchung. Sensoren am Körper oder an der Kleidung des Patienten erfassen wichtige Werte wie Blutdruck, Gewicht oder Herzfrequenz. Die Messgeräte können beispielsweise mit einer nahen Basisstation drahtlos kommunizieren, welche ihrerseits Daten speichert und zum betreuenden medizinischen Personal weiterleitet. In solchen Anwendungen sollen Datenschutz und Sicherheit der Übertragungskanäle berücksichtigt werden [7]. Das Konzept der Telemonitoring-Systeme wird in Abbildung 5 veranschaulicht.

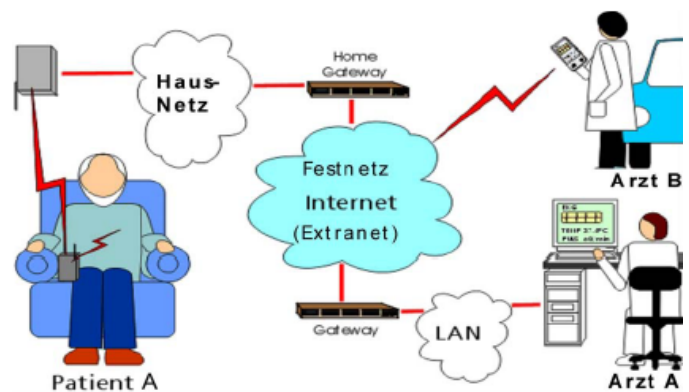


Abbildung 5. Telemonitoring-System [7]

E. Psychologie und Sozialwissenschaften

Ambient Assisted Living Technologien sollen benutzerorientiert entwickelt werden, um die notwendige Akzeptanz und Zufriedenheit der Patienten zu erreichen. Mithilfe von psychologischen und sozialwissenschaftlichen Untersuchungen werden die eingesetzten Systeme auf Gebrauchs- und Alltagstauglichkeit sowie ihrer Anpassung an die Gewohnheiten und Bedürfnisse bestimmter Zielgruppen erprobt [12].

V. ÜBERBLICK ÜBER EINIGE AAL-PROJEKTE

Im folgenden Kapitel werden einige vom Bundesministerium für Bildung und Forschung geförderte Projekte vorgestellt.

A. Projekt „SenseFloor“ (2009-2012)

In diesem Projekt wird ein sensibler mitdenkender Bodenbelag entwickelt. Diese Intelligenz wird so erreicht, indem mehrere Sensoren im Fußboden eingebettet werden. Diese erfassen unmerklich die Position und Bewegungsdaten des Bewohners, und arbeiten mit weiteren Hard- und Software-Komponenten zusammen um dem Benutzer mehr Sicherheit (z.B. Sturzerkennung, Fluchtwegplanung, Abschalten gefährlicher Geräte) und Komfort (z.B. automatisches An-/Ausschalten des Lichtes) zu bieten [2]. Die Funktionsweise

dieses Produktes wird in Abbildung 6 verdeutlicht. Durch Mustererkennung ist es zum Beispiel möglich, einen Sturz zu erkennen. Ein Notruf wird dann automatisch ausgelöst, falls die betroffene Person liegen bleibt. Das letzte Bild zeigt wie das Licht eingeschaltet werden kann, sobald man sein Bett verlässt oder das Zimmer betritt [11].

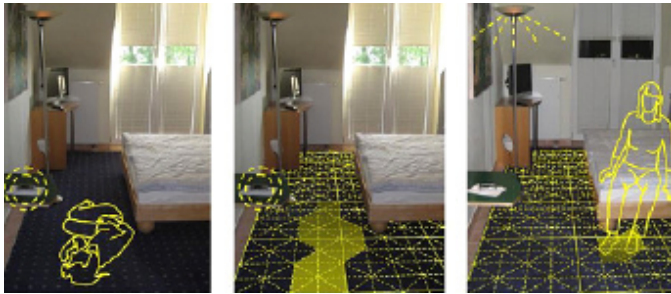


Abbildung 6. (SensFloor) sensibler Fußboden [2]

B. Projekt „AAL@Home“ (2010-2013)

Die Verwendung von hochempfindlichen Ultrabreitband-Sensoren, die aus einem Abstand von bis zu 8 Metern in der Lage sind Vital-, Bewegungs- und Positionsdaten zu erfassen, stellt die wesentliche Innovation in diesem Projekt dar. Individuelle Sensorkarten ermöglichen die Identifizierung der sich im häuslichen Umfeld befindlichen Personen. Die Lebenssituation des Pflegebedürftigen ergibt sich aus den von den innovativen UWB-Sensoren, sowie von anderen intelligenten Haustechniken gesammelten Informationen, welche mithilfe eines verteilten lernfähigen Assistenzsystem interpretiert werden können. Schließlich wird durch die Integration der abgeleiteten Lebenssituation in einer vernetzten Umgebung, bestehend aus Familienangehörigen, Pflegedienst, Hausarzt und Klinik, die Sicherheit und Unabhängigkeit des allein lebenden Menschen gewährleistet. Die Kommunikation mit diesem Versorgungsnetz sowie die Steuerung der verschiedenen technischen Hauseinrichtungen können zusätzlich mit einem nutzerfreundlichen internetfähigen Endgerät unterstützt werden [2].

C. Projekt „NutriWear“ (2007-2010)

Ziel dieses Projektes ist die Entwicklung eines textilintegrierten, intelligenten Systems zum Ernährungs- und Wasserhaushaltsmanagement. Durch eine unauffällige und bequeme Überwachung kann die Unterernährung des Patienten rechtzeitig erkannt und behandelt werden. Abbildung 7 veranschaulicht die Funktionsweise dieses Systems. Die Körperzusammensetzung des Patienten kann mithilfe der Bioimpedanzspektroskopie bestimmt werden. Textile Elektroden, die in der Kleidung eingebettet sind, ermöglichen die Erfassung von Muskel-, Fett- und Wasseranteile im Körper. Elektrische Leiter leiten die Messwerte über eine Schnittstelle zu einer Mikroelektronik, die als Aufgabe hat, diese Werte in digitale Daten umzuwandeln, damit sie schließlich von einer Auswerteeinheit interpretiert werden können, und dementsprechend die Ergebnisse dem Benutzer oder dem medizinischen Dienstleister übermitteln [3].

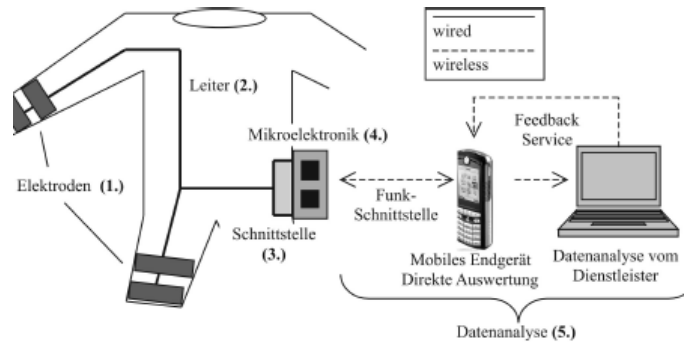


Abbildung 7. (NutriWear) textilintegriertes intelligentes System [3]

D. Projekt „VAMOS“ (2009-2012)

VAMOS steht für „Versorgungseffizienz durch assistive, modulare Technologien in bedarfsorientierten Szenarien“. Je nach Alter, Behinderung oder Erkrankung des Benutzers der Ambient Assisted Living Technologien können deren Bedürfnisse an Pflege, Sicherheit oder Unterhaltung variieren. Diese Tatsache hat zur Folge, dass die Assistenzsysteme die Anforderungen der Skalierbarkeit und Erweiterbarkeit erfüllen müssen, um sich an die entsprechenden Benutzergruppen anpassen zu können. Dafür bietet VAMOS jedem Pflegebedürftigen eine Sammlung von Gesundheits-, Sicherheits- und Unterhaltungsdiensten, die an seinen Bedarf angemessen ist. Die Assistenzfunktionen werden mit einem Fernseher vernetzt und die erfassten Daten des Patienten können zu seinen Familienangehörigen und Ärzten weitergeleitet werden um seinen Gesundheitszustand zu überwachen. Abbildung 8 gibt einen Überblick über dieses bedarfsorientierte System [2].

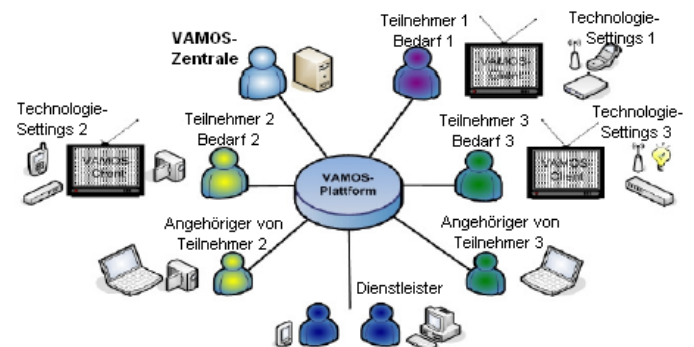


Abbildung 8. VAMOS Projekt [2]

VI. ÜBERBLICK ÜBER EIN FERTIGES PRODUKT

Da die oben betrachteten Projekte noch in Entwicklung sind, wollen wir in diesem Kapitel ein fertiges AAL-Produkt näher anschauen. Es handelt sich um das Videocare-System BETAVISTA der österreichischen Firma Zydacron. In den nächsten Unterkapiteln werden dessen Komponenten und die darunterliegenden Plattform vorgestellt [14].

A. BETAVISTA Komponenten

Die wichtigsten Komponenten des BETAVISTA Videocare Systems sind die Klientbox, die Operator Station Software und die Family Software.

Die Klientbox ist eine Set-Top-Box. Sie dient zur Videokommunikation und auch zur Übertragung der gemessenen Werte des Pflegebedürftigen. Diese Komponente wird in Abbildung 9 gezeigt.



Abbildung 9. BETAVISTA Klientbox [14]

Die Operator Station Software ermöglicht Ärzten ihre Patienten virtuell zu besuchen und mithilfe von Pflegepersonal sogar mehrere Kunden gleichzeitig zu behandeln. Zu dieser Station gehört auch eine Z-Server Software, mit derer Hilfe die Patientendaten und -messwerte an eine zentrale Datenbank automatisch übertragen und dort gespeichert werden können. Abbildung 10 zeigt wie Messdaten auf der Operator Station Benutzeroberfläche gezeichnet werden und wie sie mit dem Patienten besprochen werden können.

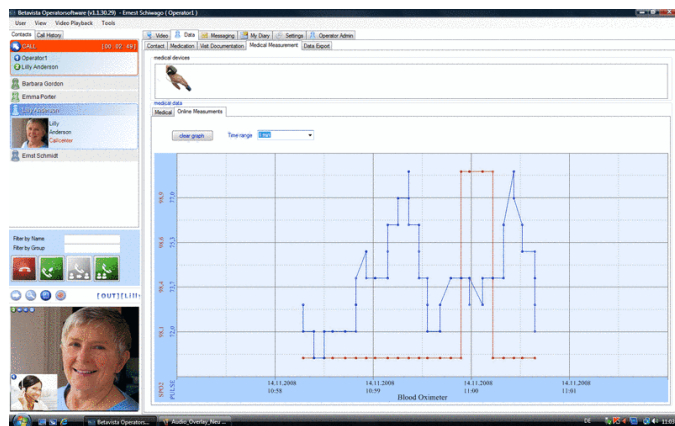


Abbildung 10. BETAVISTA Operator Software [14]

Die Family Software ermöglicht eine videobasierte Kommunikation des Patienten mit seinen Familienangehörigen und seinen Freunden. Außerdem bietet die Software weitere Funktionen, zum Beispiel Textnachrichten oder Mehrpunktkonferenzen.

B. Plattform

Mithilfe der skalierbaren BETAVISTA Service Plattform können neben der videobasierten Kommunikation auch weitere Dienstleistungen im Videocare-System integriert werden, wie

beispielsweise Notrufsysteme, medizinische Messgeräte, Sicherheitssysteme, „Meals on Wheels“ oder Infotainment. Diese Serviceangebote können an die Bedürfnisse der Benutzer angepasst werden. Unter Verwendung der oben beschriebenen Komponenten können die Vitaldaten der Patienten über eine sichere Übertragung zu dem behandelnden Arzt oder Pflegepersonal übermittelt werden. Die Daten werden schließlich in einer zentralen Datenbank gespeichert. Diese erweiterbare Plattform ist auf Abbildung 11 zu sehen.

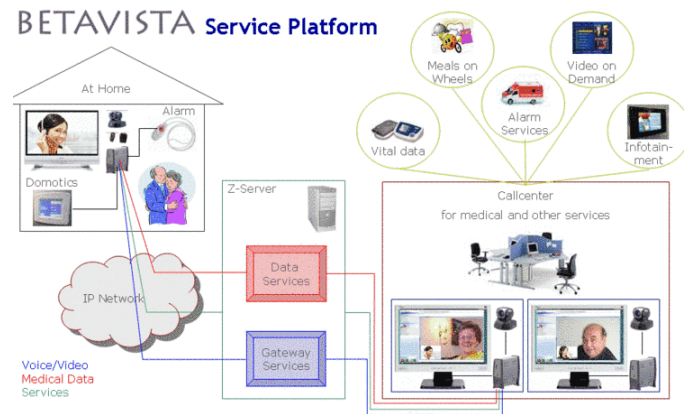


Abbildung 11. BETAVISTA Service Platform [14]

VII. ZUSAMMENFASSUNG

Unter Verwendung von Ambient Assisted Living können ältere Menschen und Pflegebedürftigen in einer von Sensorik und Kommunikationssystemen unterstützten Umgebung ein längeres, unabhängiges und sicheres Leben führen. Doch diese Ziele können nur erreicht werden, falls die Assistenzsysteme skalierbar sind und Anforderungen wie Zuverlässigkeit und Unaufdringlichkeit erfüllt werden können. Die meisten AAL-Projekte befinden sich noch in Entwicklung oder Untersuchung auf Alltagstauglichkeit. Nach Meinungen von Experten müssen noch viele Herausforderungen, zum Beispiel Finanzierungsfragen oder Interoperabilität zwischen den Geräten bewältigt werden [6].

LITERATUR

- [1] BMBF, AAL Broschüre - Altersgerechte Assistenzsysteme für ein gesundes und unabhängiges Leben. Bonn, Berlin: VDI/VDE Innovation + Technik GmbH, 2008.
- [2] BMBF, Assistenzsysteme im Dienste des älteren Menschen - Steckbriefe der ausgewählten Projekte in der BMBF-Fördermaßnahme. VDI/VDE Innovation + Technik GmbH, 2009.
- [3] BMBF, Textilintegriertes, intelligentes System zum Ernährungs- und Wasserhaushaltsmanagement - NutriWear. Bonn, Berlin: VDI/VDE Innovation + Technik GmbH, 2009.
- [4] B. Eberhardt, VDE-Positionspapier Intelligente Assistenz-Systeme im Dienst für eine reife Gesellschaft. Frankfurt am Main, Germany: Verband Der Elektrotechnik Elektronik Informationstechnik e.V., 2008.
- [5] M. Eisenmenger, O. Pötzsch, and B. Sommer, Bevölkerung Deutschlands bis 2050, 11th ed. Wiesbaden: Statistisches Bundesamt, 2006.
- [6] P. Georgieff, Ambient Assisted Living - Marktpotenziale IT-unterstützter Pflege für ein selbstbestimmtes Altern, ser. FAZIT-Schriftenreihe. Stuttgart, Germany: MFG Stiftung Baden-Württemberg, 2008.
- [7] R. Hoffmann and J. Dehm, VDE-Positionspapier Telemonitoring zur Prävention von Diabetes-Erkrankungen. Frankfurt am Main, Germany: Verband Der Elektrotechnik Elektronik Informationstechnik e.V., 2008.

- [8] L. Litz and M. Gross, "Concepts and realization of an assisted living project by extended home automation," *Festival of International Conferences on Caregiving, Disability, Aging and Technology, Toronto 2007*, 2007.
- [9] P. Panek and W. L. Zagler, "Das Living Lab für E-HomeCare und Ambient Assisted Living (AAL) Technologien in Schwechat," *Medical Informatics meets eHealth. Tagungsband der eHealth2008 – Wien, Mai 2008*.
- [10] Sensatex. (2009, dec). [Online]. Available: <http://www.sensatex.com/>
- [11] A. Steinhage and C. Lauterbach, "Sensfloor®: Ein AAL Sensorsystem für Sicherheit, Homecare und Komfort," *Tagungsband 1. Deutscher AAL-Kongress Berlin*, pp. 37–41, 2008.
- [12] TU-Kaiserslautern. (2009, dec) Das Projekt: Generationenübergreifendes Wohnen mit Assisted-Living-Komponenten. [Online]. Available: http://www.eit.uni-kl.de/litz/assisted_living/projekt.html
- [13] S. C. Vargas, *Smart Clothes - Textilien mit Elektronik: Was bietet der Markt der Intelligenten Bekleidung*. Hamburg: Diplomica Verlag GmbH, 2009.
- [14] Zydacron. (2009, dec) Betavista. [Online]. Available: <http://www.zydacron.com/de/das-ist-betavista>

Next Generation Location-Based Services

Fabian Sulzbacher – {firstname.lastname}@uni-ulm.de

Abstract—Adaptive services are able to provide a user with different content, depending on his actual context. Location-based services take a users geographical position into account to thereby deliver tailored, relevant and timely information. GPS is the most common technology, which enables exact pinpointing, and has become increasingly popular in the last years. Nevertheless there are other technologies with up to better accuracy, even indoors, which worth a closer look. Navigation systems can already be found in many cars, whereas next generation location-based services are just about to start their triumph.

I. INTRODUCTION

WHAT are Location-based Services? Location-based services (in the following referred to as LBS) are mobile services which rely on a user's location to deliver context aware functionality.

Between some 5 to 10 years ago, LBS were estimated to become the killer-application in mobile computing. LBS have been described as being, "without a doubt, one of the most exciting developments to emerge from the mobile telecommunications sector." [18] Contrary to expectations, they did not make their commercial breakthrough. Recent LBS-projects ,however, raise the hopes, that the era of LBS is finally about to begin.

ABI research, a marketing research company based in New York, states "Nine of every ten smartphones will contain GPS ICs in 2014, compared with one in three in 2008." [19] This is due to the ongoing demand for feature-rich smartphones and reduced hardware-prices. With GPS becoming ever-more common, LBS will increasingly become popular. This is also due to the success of app stores as a distribution channel.

Indeed, location-based information did not come up recently with the invention of mobile communication-devices. Position specific information is virtually everywhere, ranging from the post-it note on the fridge over a poster in the movie theater, meant to reach a mass audience, to traffic signs, submitting navigational information. But all this information is not tailored to particular needs. For a driver, who is heading to Ulm, for example, traffic-sign for other cities are not relevant [22]

This is where LBS comes into play. The term LBS covers a wide diversity of applications, which all have in common, that they rely on knowledge of a user location to provide customized information or services to a mobile device. This information will generally be transmitted over a mobile network. The tailored information must not only be on the users location, but can also consider a user profile and the surrounding context. Thereby, these services can help to orientate within an unknown environment.

Kirsi Virrantaus, professor in geoinformatics and cartography and head of the laboratory of geoinformation and positioning technology at the Helsinki University of Technology



Figure 1. Modern smart phones often are equipped with GPS-modules [24]

defines LBS as "information services accessible with mobile devices through the mobile network and utilizing the ability to make use of the location of the mobile device." [22] Another definition is given by the Open Geospatial Consortium: "A wireless-IP service that uses geographic information to serve a mobile user. Any application service that exploits the position of a mobile terminal" [22]. In a nutshell: LBS is an information, entertainment, or media service that is making use of geographical position of a mobile device.

II. TECHNICAL BACKGROUND

This section is about the components, which are necessary for a LBS, namely the mobile device itself, the communication network, the service provider, the content provider and positioning technique [22]. The positioning techniques are then discussed in the second part of this chapter.

A. LBS Components

- 1) **Mobile Devices:** The Device must be able to present information in some way. This can happen by means of speech or by simply displaying pictures and text. It must also enable the user to request information from a service provider. A device can be a Mobile Phone, Smartphone or PDA, but also a Laptop, a navigation device for cars or e.g. a n for road pricing.
- 2) **Communication Network:** The Communication Network transmits user-requests to a service provider and then transmits the requested information back to the device.
- 3) **Positioning Techniques:** Location-based services have in common, that they determine the user's position to provide him with tailored Information. The most common technique to pinpoint the user's position is GPS.

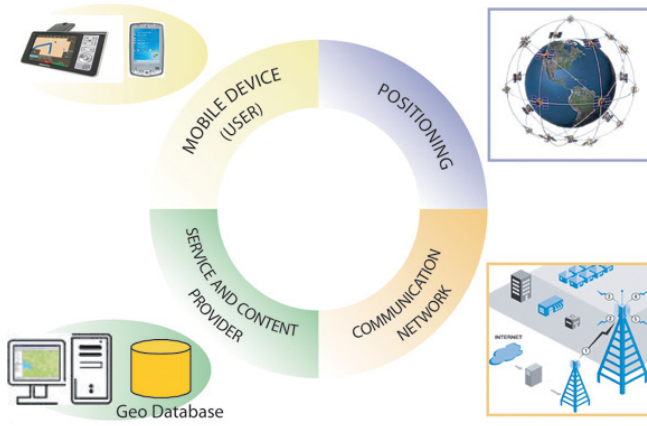


Figure 2. LBS Components[22]

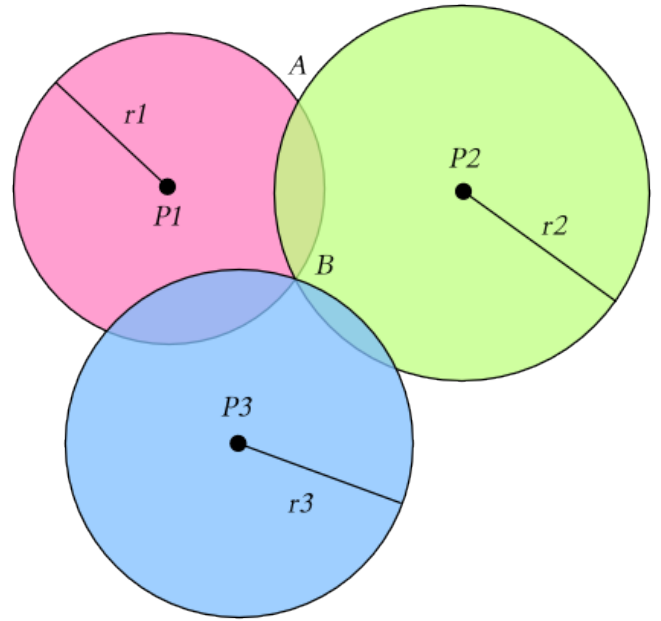
Furthermore the position can be obtained via WLAN- Stations, RFID, Bluetooth or time-based positioning. We will have a closer look on these methods in the following.

- 4) Service Provider: The service provider is responsible for the service request processing. The request will be analyzed in order to provide the user with tailored information with respect to his actual position. Such services can offer the calculation of the position, finding a route, finding a nearby restaurant or friend, finding specific information on objects of user interest and many more.
- 5) Content Provider: Service providers will not actually store and maintain all the data which can be requested by users. Therefore a Content Server provides the service provider with specific information.

B. Positioning Techniques

LBS should work inside and outside buildings with high accuracy, anywhere and anytime. Unfortunately, these requirements cannot be matched by a single technology at the present time. For now, LBS depends on a combination of different technologies. Most of those methods rely on a method, called trilateration, which allows the exact calculation of a position in a plane, when knowing the position of three reference points and their distance. To additionally obtain the altitude of a position, four reference points are necessary.

- 1) GPS relies on a minimum of 24 positioning satellites, circling the earth in an average height of about 20200 kilometers. They are constantly moving with a speed of 3.9 kilometers per second, orbiting the earth twice within less than 24 hours. By constantly emitting uniquely identified radio signals, which can be received by specialized units on earth, they enable exact positioning. The signals contain a timestamp and information about the identity of the satellite and its position. They are used to precisely measure the distance from the receiving instruments to each satellite visible overhead, using trilateration. To compute an accurate position on earth, a minimum of four signals from different

Figure 3. Trilateration: Points P1, P2, P3 and their corresponding distances $r1$, $r2$ and $r3$ are known. Point B (a mobile device) is, where the three circles intersect.[24]

satellites is needed. The signal strength, however, is insufficient to pervade walls. Therefore GPS allows positioning exclusively outdoors. The US Federal Aviation Administration states, that GPS guarantees a horizontal positioning accuracy with an error of less than thirteen meters in 95% of measurements and the combination with augmentation systems make even higher accuracies possible.[23, 1, 7]

- 2) Wireless Local Area Network (WLAN): For cost reasons it would be desirable to use an already existing, widely distributed infrastructure. Therefore, a WLAN positioning system could be implemented with the least effort and cost, compared to Bluetooth or RFID. Although WLAN was never designed for the purpose of positioning, there are two ways to pinpoint a user's location via WLAN. The first uses information about the signal strength (which can be obtained without knowing the password) and translates it into an approximation of the distance from the access point to the mobile device. The position can then be calculated via trilateration, assuming that there are at least three signals from different access points. This methods accuracy is about 4-5 meters. The difficulty with this approach is, to procure the distance accurately, because the propagation of indoor radio signals is very complicated, due to distance and penetration losses. The other way of WLAN positioning is known as location fingerprinting. The key idea behind it, is to map location-dependent parameters of measured radio signals to the area of interest, in order to assign the specific radio signal, received by the mobile device, to a certain reference point on the map. This postulates a training phase in order to generate a fingerprint-database. In densely populated urban areas, WLAN positioning

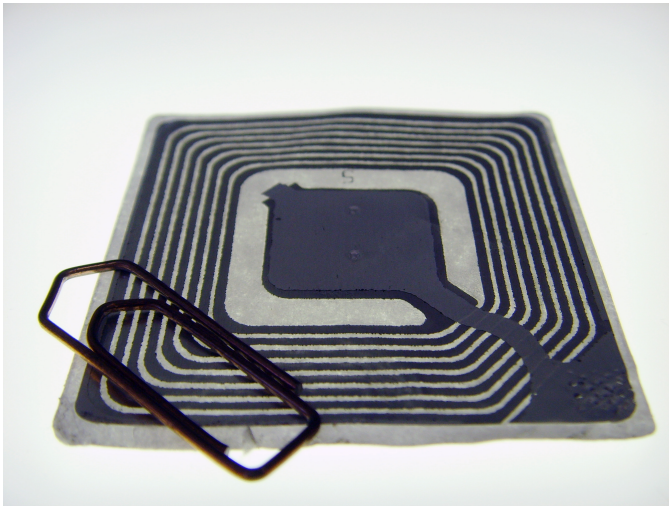


Figure 4. Modern consumer retail RFID tag[24]

can be used inside and outside of buildings, without changing the existing infrastructure.[16]

- 3) **Bluetooth:** Bluetooth is a low cost, low power, short range radio technology, originally developed as a cable replacement to connect devices such as mobile phones, headsets, PDAs and portable computers. The range of the Bluetooth device affects the accuracy of a position, acquired via the Bluetooth system. There is an indicator that displays the received signal strength, but it is not considered to determine distance. However, a received signal indicates, that the device must be within a maximum distance of ten meters, because this is the maximum communication range between two Bluetooth devices. Trilateration can be used in order to improve the accuracy of distance measurement.[14]
- 4) **Radio-frequency identification (RFID):** RFID-tags are electronic chips that usually contain an integrated circuit in order to store and process information and to modulate and demodulate a radio-signal and an antenna to receive to send and receive radio-signals. They are attached to an object, animal or person for purpose of identification and tracking. There are active, passive and battery-assisted passive RFID tags. While active RFID tags can cover up to one hundred meters around them, passive tags may only cover just a few meters. RFID is increasingly gaining importance in various sectors, for their characteristics that allow a great diversity of applications, but also for their low cost of less than 6 cents for each tag. They are being used for E-Tolling on roads or bridges, public transit, product tracking, animal identification, race timing, but also in libraries and museums. Triangulation, in order to enhance positioning accuracy, is theoretically possible, but hardly feasible, because there is no indication of signal-strength and because their range can vary, depending on what they are designed for. Thus RFID-positioning accuracy usually depends on the range, a RFID-tag can cover.[13]
- 5) **Cell of Origin (COO):** COO is a positioning technology

to find a user's cell within a communication network. Therefore the base stations are connected with geographical information. Depending on the number of base stations in the search area, accuracy may differ. In urban areas a location can be determined with an accuracy up to one hundred meters, but in a sparsely populated area it may be a lot less accurate (up to thirty kilometers discrepancy from target). COO is being used by emergency services in many countries, meanwhile they didn't find a wide acceptance in commercial applications.

- 6) **Magnetometers:** Magnetometers are appearing in mobile phones. They can serve like a compass to enhance the positioning by adding the direction, a user is facing.

III. CONTEXT

Context is any information about aspects, considering persons, the place, objects or the situation, which is relevant to the interaction between the user and the application. Adaptive LBS use these information in order to react dynamically to situations. Annu-Maaria Nivala, researcher at the Finnish Geodetic Institute, came up with a classification of context aspects, with specific reference to mobile services. The 9 types of context, shown in the figure, are: Mobile map user: Information about the user, like age, gender, personal preferences, friends and colleagues, to name just some of them.[22]

A. Location

As you can tell, location is the most commonly considered aspect of context. It can either be absolute, e.g. GPS-Coordinates, or relative, e.g. distance from some point of interest or from a friend.

B. Time

Time usually refers to the time of day, but can also refer to longer periods of time, like morning or evening, day of the week, month, season or year. A possible use would be the suggestion of an event, considering the time, for example a temporal exhibition or restaurant for dinner.

C. Orientation

To determine what a user is currently looking at, not only his position, but also his orientation must be taken into account. An application which considers this information might be an electronic tourist-guide, that supports the user with information about the specific building he is looking at.

D. Navigation history

This context-aspect is about keeping track of the user's navigational history, in order to see where he comes from and where he has been before. This might be useful to help the user to find back, once he got lost, but also to build up a profile on user's likes and dislikes, considering the places he visited recently.

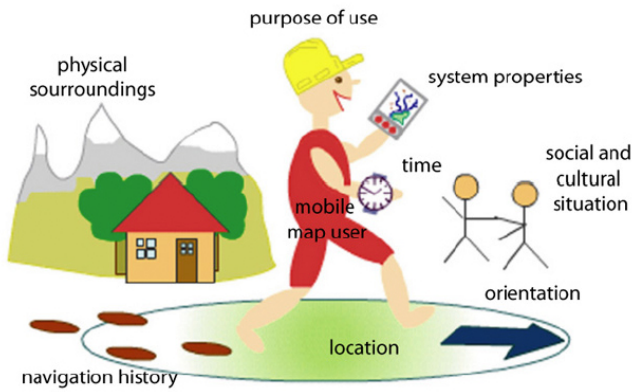


Figure 5. Context categories[22]

E. Purpose of use

This aspect deals with the user's intention when using a service. It is specified by the user's actions, tasks and role. The purpose of use determines the needed information and the way of presentation.

F. Social and cultural situation

The social situation of a user is characterized by its proximity to others, social relationships and collaborative tasks. Applications could tell the user where their friends are, where a popular bar is.

G. Physical Surroundings

This aspect includes everything which is the actual ambience, for example noise level, lighting level, height and so on. This information might be used to better identify the user's surrounding area.

H. System Properties

This aspect of context relates on the employed infrastructure, such as the resolution, memory capacity or computing power of the used device, bandwidth, positioning accuracy and so on. Needless to say, that those parameters must be considered by an application in order to use the available resources efficiently.

IV. CATEGORIES

The definition of LBS as "any information, entertainment, or media service that is making use of geographical position of a mobile device" makes it a broad field with applications in nearly each specific field of computer science. Seven main categories of LBS are identifiable: Advertising Services, Billing Services, Emergency Services, Navigation Services, Information Services, Tracking Services and Games. Nevertheless they can overlap with each other. Eventually Augmented Reality is to be mentioned, not as a separate category of applications, but because it is a technology which is often used in combination with LBS.[22]

A. Advertising Services

Location-based advertising (LBA) is a new form of market communication that sends tailored advertising to a user's mobile device, considering his current position. It is actually quite familiar with context-based Internet-advertising, which is already used, for example by Google. In each search the results also contain a list of google-clients, who pay for their advertising to be displayed. Location-based advertising could work in a similar way. Advertising-clients could provide their potential clients with targeted real-time advertising. E.g. a moviegoer passed a movie theater and received a text message, suggesting a movie.

B. Billing Services

Location-based billing is the ability of a service provider to dynamically charge users of a particular service, depending on their location, when using or accessing the service. Examples of applications are E-tolling on roads in respect to the exact distance that is covered by a truck, measured through a specialized devices on board the vehicle and billed by the end of the month or flexible pricing by mobile network operators, considering the caller's position, which would them allow to compete better with the fixed telephony concurrence. Imaginable would also be the use of RFID tags on tickets for big events, in order to speed up the inflow or to locate remaining persons in case of an evacuation. [9]

C. Emergency Services

One of the most evident and likewise useful applications are emergency services, which reveal the location of a person who is in need of help, but is either unaware its current position or is unable to tell it because of injuries, unconsciousness or other emergency situations. An LBS can detect such a situation and assist the emergency services by providing them with the exact location of the incident and further contextual information, in order to permit them to help quickly and efficiently. In Europe and North-America emergency services are automatically provided by the telephone service provider with the caller's position (E112 and E911). Life-jacket, which are able to send their position, when getting in contact with water exist and are in use. And within the next decade a new System, named eCall, will be introduced, which enables cars to detect involvement in an accident, to advise emergency services and transmit their exact position.

D. Navigation Services

Navigation Services are the best known and most common LBS. They route a user to a specified destination. Those services can either be used outside, using GPS, or inside a building, using one of the earlier mentioned positioning methods. While car navigation devices already made their commercial breakthrough some years ago, other navigation systems are still rare. However, there are many other areas, where they are being used. Electronic guides for fairs, congresses and museums, to mention some.

E. Information Services

Finding the nearest laundry, accessing traffic news, obtaining a local street map, retrieving information about the appearance of certain animals in a national park and so on. These are just some examples of location-based Information Services. The list is literally endless. Location-sensitive information services mostly refer to the digital distribution of information, based on device location. They might be used in electronic tourist guides, in order to enrich the touristic experience to provide the user with specific information about the building, he is looking at. Another interesting field is the upcoming connection of social networks with LBS. User must register and can thereby allow a certain group of people, which they can define, to see where they are at any time of day.

F. Tracking Services

One popular example of Tracking Services refers to the localization postal packages so that customers know where their packets are at any given time. Another application is vehicle tracking, which can be applied to locate and dispatch the ambulance which is next to an accident or a taxi which is next to a calling customer. Other applications allow companies to locate their field personnel (e.g. salespeople and repair engineers) so that they are able to provide their customers with accurate arrival times.

G. Games

Even games are using the positioning technology to provide a totally new gaming experience to the user. Geocaching is the modern way of a treasure-hunt. The participants are supposed to search for a treasure in real world, using a GPS receiver. A Geocache is a waterproof box, prior hidden by a player, usually containing a pencil, a logbook and some sort of trade-item. The exact GPS-coordinates and a description of the surroundings is then posted on a listing site, where other players can see them. Elsewhere positioning is used for games, similar to classic cops and robbers, like "Uncle Roy All Around You" from England, and even location-based social networks like Foursquare include gaming aspects in their concept, rewarding users with special bonuses, when they find a new location in their neighborhood, change location many times in the same night or travel to far away-locations.[3, 8, 6]

H. Augmented Reality

Talking about LBS, it is worth to have a look on a technology called Augmented Reality, which is already used by many applications, especially LBS, and will probably become ever-more popular in the next years. The idea behind it is to pull graphics out of the mobile device or computer display and integrate them into real-world environments, enabling the user to see the real world around him, with computer graphics and sounds overlaid. Thereby real world is not replaced, but enriched. Devices can be head-up display and specialized monitors worn on the head, but also simple cell phones equipped with a camera.



Figure 6. Augmented Reality Visor [2]

V. EXAMPLES

A. Zingo

Zingo is a Taxi service operating in London. An operator can obtain the position of a calling passenger from the phone company, which does the positioning by checking at which mobile phone mast the phone is locked on to (COO). The taxis are equipped with GPS and are sending their position continuously to the control centre. When a client requires a cab, the call will automatically get routed through to the cab driver. The client can thereby tell the exact address and the driver can estimate how long it will take until he arrives. [21, 11]

B. Foursquare

Foursquare is a location-based social network. User can continuously update their position, using text messages or a device specific software and can thereby keep friends informed about their location at any time of day. Furthermore they get awarded with points and badges. Points can be earned for visiting a specific location and badges are gained by visiting a certain location frequently or by visiting a very special location. Users can also choose to keep their Facebook- and Twitter-accounts

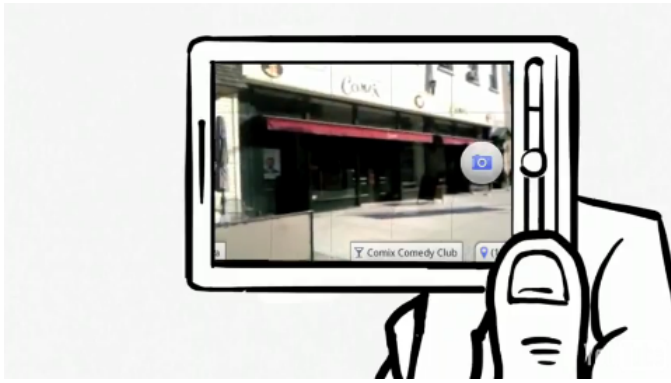


Figure 7. Google goggles, [10]

updated with their locations and achievements. Foursquare is currently used by more than 100.000 users in many cities around the world. [6]

C. Google Goggles

Google Goggles is a visual search software by Google, which contains a functionality that allows the user to just point the camera of his cellphone at any building in a city, to obtain information about it. The software locates the user by GPS and then calculates his line of sight, using the devices built-in magnetometer. Afterwards the image of the camera is compared with the graphical material from the Google-Street-View-project, considering the mentioned position parameters, to tell the user, what he is looking at. [10]

D. Toll Collect

Since January 2005 trucks are obliged to pay a road charge on the German autobahn. Therefore a tolling-system was introduced, which is able to monitor the up to 1.5 million trucks, which travel an estimated 23 billion kilometers per year on the 12,000 kilometers of autobahn. The technology relies on GPS, a website, that allows booking routes in advance, and a "On Board Unit" which keeps track of the position and communicate with specialized control bridges via infrared.[20, 4]

E. Detecting patients suffering from dementia or Alzheimer

Specialized provider, for example LoJack or Quest Guard, offer Tracking Services to localize persons suffering from dementia or Alzheimer. Therefore they get outfitted with a GPS-unit, either in an arm bracelet, ankle bracelet or a shoe. If a, in such a manner equipped, person is missing, the service provider is able to obtain the person's exact position. In some countries this services are not even limited to the mentioned group of people, but to anyone who considers himself to be in risk of being kidnapped for example. The problems, which these human tracking systems imply, will be discussed in the following chapter. [17, 12]

VI. PRIVACY

Technology cannot be good nor evil, but it can be abused to empower those, who intent to use it for purposes, it was never designed for. A tradeoff between the potential hazard and the potential benefit for society must be found. It is the developer's duty to prevent abuse. A person's location must be considered sensitive information. Tracking devices, however, are increasingly becoming popular, bringing the potential for real-time surveillance of individuals to a new level. Geoslavery is a new form of slavery, which society will have to face. This danger has not yet been addressed sufficiently by public officials and is completely ignored by commercial vendors. According to Jerome E. Dobson and Peter F. Fisher, geoslavery can be defined as "a practice in which one entity, the master, coercively or surreptitiously monitors and exerts control over the physical location of another individual, the slave. Inherent in this concept is the potential for a master to routinely control time, location, speed, and direction for each and every movement of the slave or, indeed, of many slaves simultaneously" Human tracking systems, sold commercially without restriction, already enable users to track other fellow humans, though they were not designed for this particular purpose or they are designed to track a certain group of individuals, like alzheimer patients or children. Nevertheless, a system which enable a master to track every step of a slave is feasible and not even hard to implement. Additionally equipped with some sort of transponder, that punishes the slave by shocking, stinging or burning him at the push of a button, the master would gain complete control over the slave, making him equivalent to a human robot. He might tell him where to go, forbid him to enter certain taboo-areas or to meet with certain other individuals. In the western civilization those extreme forms of abuse would hopefully not come to happen, due to social conventions and new laws, addressing this particular thread. However, abuses will not be avoidable and in some countries, where woman already are strictly controlled, freedom of opinion is nonexistent and laborers are exploited, geoslavery might become the principle use of LBS. Needless to say, that the hazard is not only an omnipresent state, keeping its citizens under surveillance, like described in Orwell's *1984*, but also from private persons, e.g. an abusive husband, who wants to keep track of his wife's every step. Currently GPS is offered without any restrictions. In an attempt to respond the thread of geoslavery, GPS signals could be encrypted and licensed, as it is planned for the Galileo GPS system, for commercial reasons, indeed. The future of LBS significantly depends on user acceptance and the user acceptance in turn depends on how specific concerns like location privacy are addressed. Not only the issue of geoslavery, but also some other undesirable consequences of abusive use of LBS, including location-based spam, stalking or inferring personal information, are reasonable concerns. Therefore location privacy is an urgent research issue. Till today no government has established any law, covering restrictions on the use of LBS. [15, 5]

VII. CONCLUSION

Location-based services have an enormous potential, which is not nearly exhausted with car navigation systems. There

are thinkable applications in many specific fields of computer science, especially in combination with other technologies, like augmented reality. However, there still is no satisfying solution for indoor positioning. Nevertheless there are interesting approaches, some of them not even requiring additional infrastructure. Due to the growing smartphone market, falling hardware-prices, the success of app stores and the increasing customer acceptance, the long time predicted breakthrough of LBS, seems to be about to begin, finally. However, the distribution of LBS on a grand scale should be put into effect with responsibility, because, though LBS offers enormous benefits on the one hand side, it carries considerable risks on the other hand side. As with nuclear energy, responsible governments and industries must establish safeguards, to assure that LBS cannot be abused.

REFERENCES

- [1] F. A. Administration. Gps pan report. [Online]. Available: <http://www.nstb.tc.faa.gov/>
- [2] augmentedenvironments.org. [Online]. Available: <http://www.augmentedenvironments.org/>
- [3] blasttheory.co.uk. <http://www.blasttheory.co.uk>. [Online]. Available: http://www.blasttheory.co.uk/bt/work_uncleeroy.html
- [4] T. Collect. [Online]. Available: <http://www.toll-collect.de/>
- [5] J. E. Dobson and P. F. Fisher, "Geoslavery," 2003.
- [6] foursquare.com. [Online]. Available: <http://foursquare.com/>
- [7] Garmin. about gps. [Online]. Available: <http://www8.garmin.com/aboutGPS/>
- [8] geocaching.com. [Online]. Available: <http://www.geocaching.com/>
- [9] Gesellschaft für Informatik. Location-based services. [Online]. Available: <http://www.gi-ev.de/>
- [10] Google. Goggles - use pictures to search the web. [Online]. Available: <http://www.google.com/mobile/goggles/#landmark>
- [11] D. Greek, "High-tech taxi service hits the streets," 2003. [Online]. Available: <http://www.v3.co.uk/vnnet/news/2121771/tech-taxi-service-hits-streets>
- [12] Q. Guard. [Online]. Available: <http://www.questguard.com/>
- [13] J. Hallberg and M. Nilsson, "Positioning with bluetooth, irda and rfid," 2002.
- [14] J. Hallberg, M. Nilsson, and K. Synnes, "Bluetooth positioning."
- [15] L. Kulik, "Privacy for real-time location-based services."
- [16] B. Li, J. Salter, A. G. Dempster, and C. Rizos. Indoor positioning techniques based on wireless lan.
- [17] LoJack. [Online]. Available: <http://www.lojack.com/>
- [18] B. E. Mennecke and T. J. Strader, *Mobile commerce: technology, theory, and applications*, 2003.
- [19] ABI Research. Gps-enabled handsets. [Online]. Available: <http://www.abiresearch.com/>
- [20] roadtraffic technology.com. Lkw-maut electronic toll collection system for heavy goods vehicles, germany. [Online]. Available: <http://www.roadtraffic-technology.com/projects/lkw-maut/>
- [21] P. Rubens. (2003) How to hail a cab with a mobile phone. [Online]. Available: http://news.bbc.co.uk/2/hi/uk_news/2946129.stm
- [22] S. Steiniger, M. Neun, and A. Edwardes, "Foundations of location based services."
- [23] The National Executive Committee for Space-Based Positioning, Navigation, and Timing. [Online]. Available: <http://pnt.gov/>
- [24] wikimedia. [Online]. Available: <http://commons.wikimedia.org>

Privacy in Video Surveillance

Raimar Wagner – {firstname.lastname}@uni-ulm.de

Abstract—In a world where CCTV and video surveillance is part of our daily life privacy its getting more and more an issue. Since cameras are not only standalone TV-like devices anymore, but networked high resolution cameras combined with bleeding edge computer vision techniques we need new approaches for protecting one's privacy in this context. In this paper a brief overview over the development of video surveillance is given. A basic structure for accessing sensitive information is described before three state-of-the-art privacy protecting techniques are presented and rated. This paper closes with a general conclusion.

I. INTRODUCTION

A Couple of years ago, video surveillance was primary a way to secure critical sites and facilities like nuclear plants or military barracks. But since CCTV's were used to fight everyday crime it has reached the average citizen's life. London itself has 10,524 CCTV cameras installed in the city zone, from both police agencies and private owners¹. Focusing on German media, the so called "Kofferbomber" of Cologne and the murder trial at a metro station in Munich were major incidents on how CCTV could not be assumed away anymore. Today, there is hardly any city in the world which lacks video surveillance in public spaces. Almost every shopping mall or supermarket has been equipped with some kind of camera system.

While government and police officials are primary pointing out the security benefits, civil rights groups like the American Civil Liberties Union (ACLU) are raising inconvenient questions. Abuses and privacy concerns are not made up out of thin air. A British study [1] showed that black male men have often been addressed for "no obvious reason". Most countries still lack of a statutory basis when it comes to privacy in this specific field. Combined with privacy insensitive camera operating personell even more privacy concerns are raised.

In Section I a brief overview how video surveillance gained importance in our everyday life will be given. Section II will cover the basics of a layer based policy for accessing video surveillance data. In Section III three different approaches how one's privacy could be protected are presented. This paper closes with Section IV as conclusion.

A. From autark CCTV's to Video Surveillance Networks

Today's CCTV's are usually working as independent devices and providing camera images of low resolution and therefore of bad quality. The video stream is generally manually evaluated and for the most part not even gets completely sighted. In contrast to earlier times when such systems were expensive and only available from special dealers, today such

equipment is on hand even at ordinary do-it-yourself stores. This significant amplification of availability and the decreasing prices have paved the way for a triumphal procession of video surveillance. But not only the video quality and the availability is increasing, today's research on neural networks and pattern recognition leads to scenarios which are usually only known from movies like Minority Report. Having these extreme powerful cameras combined to huge networks of thousands of cameras in which every camera brings in his data makes this technology a very powerful tool for invading one's privacy.

II. A BASIC LAYER BASED ACCESS POLICY TO PRESERVE PRIVACY

Almost all approaches presented later on can be integrated in some kind of layer-based access model which handles privileges and only shows certain information to authorized staff or law enforcement agencies. For every layer in the model it has been defined which data is shown to the user and how raw this data is shown, e.g. ordinary security personell will not be able to see all raw information. To ensure that every user is only able to access information that has been meant for him the underlying architecture is using access-control lists and standard encryption techniques.

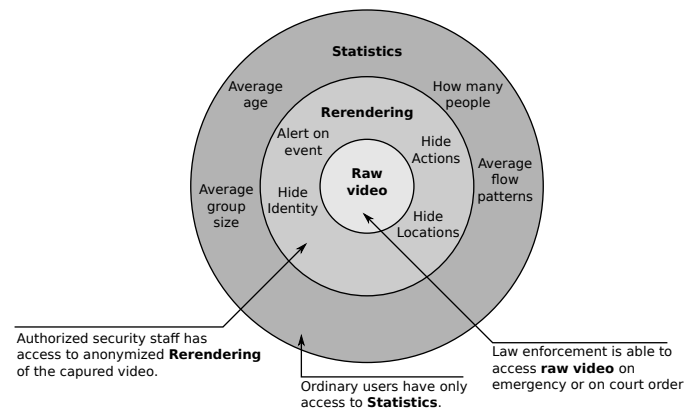


Figure 1. A layer based system architecture approach – based on [6]

Senior et al. [6] defines three basic groups with different access rights on the surveillanced data:

- **User with access to statistical data**
This data could be published riskless from a privacy point of view. All information has been anonymized, the user does not receive any image data.
- **User with access to rerendered data**
The security personell which has been introduced to the CCTV system has access to an anonymized video image to recognize possible conspicuous activity.

¹<http://www.thisislondon.co.uk/news/article-23412867-tens-of-thousands-of-cctv-cameras-yet-80-of-crime-unsolved>.

- **User with access to the raw video stream**

Access to the raw video stream containing all information and possible matches of face detection algorithms is restricted to law enforcement agencies. Even for such agencies the access should only be granted in an exceptional emergency case or on explicit court order.

Figure 1 from Senior et al. [6] illustrates this layer-based access policy system. Most of the techniques described in the upcoming section are used to preserve one's individual privacy in the *Rerendering-layer*.

III. PRIVACY PRESERVING TECHNIQUES IN A NUTSHELL

Having implemented an underlying access system, such an system needs to be equipped with algorithms and methods to detect and mask pedestrians from the video streams.

A. Using Mobile Communications to Assert Privacy from Video Surveillance

One approach to use mobile devices, so called *Privacy Enabling Devices*, has been presented by Jack Brassil [2]. Everyone who is interested in being private has to take along his special *PED* device, which basically is containing a clock, a unique identifier, a GPS device and a data storage or a wireless data connection. Equipped with these components every *PED* is able to locate and transmit/store the owners position at every point in time. The amount of data collected by the *PED* is fairly small (e.g. < 100 bytes/s), thus any built-in memory with up to 32 Mb is sufficient. For immediately transmitting these location information even GPRS with a bandwidth of up to 21.4 kb/s is adequate.

Entering a private state is simple for the user, once he is turning on some kind of *private switch*, his *PED* device is continuously sending the captured timestamp/location and identifier tuples to a so called *clearinghouse*. If the wireless link is temporary down, the *PED* is storing the information locally for a later synchronization with the *clearinghouse*.

The *clearinghouse* is collecting the trajectories from all *PED*'s; ordered by current location. Once a camera operator wants to disseminate his captured videos, he needs to make sure the privacy settings of all observed individuals are respected. He first needs to query the *clearinghouse* if an active *PED* was located in the monitored area. Once the *clearinghouse* is signaling at least one active *PED* the associated sequence and the recorded trajectory has to be processed by a *sanitizer*.

The *sanitizer* now needs to detect the *PED*'s trajectory in the supplied video sequence. This is basically done by extracting all distinct trajectories in the image sequence with not clarified computer vision techniques. The *PED*'s 3D trajectories now have to be mapped to image 2D trajectories by calculating a transfer function using various camera parameters.

1) *The sanitizing process*: The sanitizing process for one camera over the period $[t_0, t_0 + \Delta t]$.

Input:

- Imagesequence $v : [t_0, t_0 + \Delta t]$ with dimension $\begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix}$ at position (x_0, y_0, z_0) and camera image size (c_x, c_y)
- *PED* trajectories $p_1 \dots p_n$ of private *PED*s
- 3D to 2D camera back projection $back_proj(x) = y : x \in \mathbb{R}^3; y \in \mathbb{R}^2$
- trajectory error heuristic: $cmp(a, b)$, with $cmp(a, b) = 0 \Leftrightarrow a = b$ with a and b 2D trajectories. Popular choices are *Mean Square Error (MSE)* or *Linear Error (LE)*.

Algorithm III.1:

```

1 for  $t = t_0 \dots t_0 + \Delta t$  do
2    $\lceil$  Computer-vision methods are detecting all  $k$ 
3     individuals  $q_1 \dots q_k$  from  $v$ .  $\rceil$ 
4   //holds coordinates of all actual pedestrians
5    $act\_peds = []$ 
6   for  $p_t = p_1 \dots p_n$  do
7     if  $p_t$  is in  $(x_0, y_0, z_0) + \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix}$ 
8        $img\_coord = back\_proj(p_t)$ 
9        $act\_peds.push\_back(img\_coord)$ 
10    end
11  end
12  foreach  $p_z = act\_peds$  do
13     $\lceil find\ q_i : cmp(p_z, q_i) \rightarrow min \rceil$ 
14    mask  $q_i$  in image
15  end
16  end
20 end
    
```

The simulation of this algorithm has shown that this kind of approach is usable, even with async clocks, many *PED*s in a small area and imprecise location information.

2) *Conclusion*: The systems is based on cheap hardware, integrated in cell phones or PDA's, therefore the concept of a *PED* could be very successful. Simulation results have shown that various error sources do not impair the matching process. Because the user has to trust the *clearinghouse* and that privacy is only guaranteed if camera owners collaborate with the *clearinghouse* this approach raises new privacy concerns. Any intruder could retrieve the users exact location in every point in time.

B. Object-Video Streams for Preserving Privacy in Video Surveillance

Faisal Z. Qureshi et al. [5] has presented an approach how to make individual privacy possible by using *video-object streams*. A *Video-object* is basically any object in the scene which is not considered to be part of the background, mostly the *video-objects* are referring to the people in the scene. Having the raw video decomposed in the background and n object streams builds the foundation for the *Rerendering* layer from Figure 1. With one stream for each person or object

hiding or blurring one's part of the video is an easy job. In particular it is possible to:

- obscure the identity of an individual by replacing one's stream with a colored blob
- mark selected streams with special colors to indicate a special condition while preserving the identity of the individuals.
- reconstruct the whole image from all streams.

1) *The Vision Pipeline:* The whole computation process from the raw image to a pedestrian tracking video stream with privacy features is organized in a *vision pipeline* (see Figure 2). The computer vision algorithms and methods for all the presented steps are well-known and fast enough for a real-life scenario.

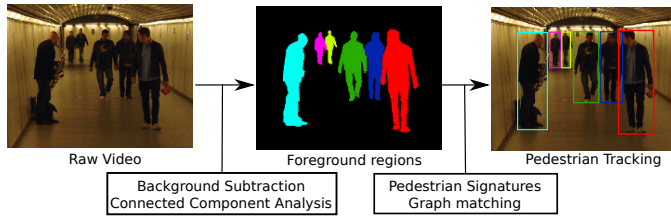


Figure 2. The vision pipeline¹, from Qureshi et al. [5].

a) *Background Subtraction:* At first, the camera has to learn how the background looks like. Having learnt the scene without any pedestrians present, the camera is now able to compute at every moment in time a (quite noisy) background mask. With such a mask all foreground objects can be distinguished from the background and used as possible pedestrian candidates. A real system in this case also needs to handle shadows, changing backgrounds and different lighting conditions.

b) *Pedestrian Tracking:* To obtain a list of all pedestrian candidates the foreground mask is getting applied to the whole image. All returning blobs are pedestrian candidates or pedestrian group candidates. [5] has implemented a state-of-the-art pedestrian tracker using a 3D-Histogram as pedestrian identification signature and the concept of the bipartit graph matching problem described in [3].

The tracker assigns one or more pedestrians to one blob in one frame, if the tracker could not find any appropriate blob in a frame the pedestrian is mapped to the whole image.

2) *Object-Video Streams:* By using previous mentioned Pedestrian Tracking and Background Subtraction methods the decomposition of the stream is done in the following way:

Algorithm III.2:

```

1 Extracting all background pixels  $F_t^B$ 
3 for  $t = t_0 \dots t_n$  do
4   for  $\forall$  Blobs  $C_i \in F_t$  do
5     [Construct Blob Mask  $M_t^i$  with 1
6       for all pixels enclosed in Blob  $C_i$  and 0
7       for other pixels]
```

¹"Saxophonist in einer Grossstadt." [Online] Available: <http://www.pixelio.de/details.php?imageid=406000&mode=search>

```

8   Construct  $F_t^i(x)$  as  $F_t(x)$  masked with  $M_t^i(x)$ 
9   end
10 end
12 for Objects  $O_i = O_0 \dots O_k$  do
13   Objectstream  $OS_i = []$ 
16   for  $t = t_0 \dots t_n$  do
17     if Tracker maps  $O_i$  to any Blob  $j$ 
18        $OS_i = OS_i \oplus F_t^j$ 
19     else
20        $OS_i = OS_i \oplus F_t$ 
21     end
22   end
24 end
```

The original image can always be recovered by calulating:

$$F_t = F_t^B \cup (\cup_i F_t^i)$$

After one run the whole sequence is decomposed into k Objectstreams $OS_1 \dots OS_k$. All intermediate steps have been illustrated in Figure 3. These streams can now be used in the previous mentioned ways to preserve one's privacy in the sequence.

3) *Conclusion:* The author has presented a way how a video stream from a CCTV can be decomposed loss-less into *Object-video-streams* by using state-of-the-art computer vision and tracking methods. Having k object streams instead of one fixed streams brings new opportunities for improving the privacy capabilities of a video surveillance system. This concept is in my opinion a good way to implement the *selective rendering* approach already presented in II. On the other hand, the success of this method is strongly tied on the capabilities of the underlying computer vision and tracking techniques.

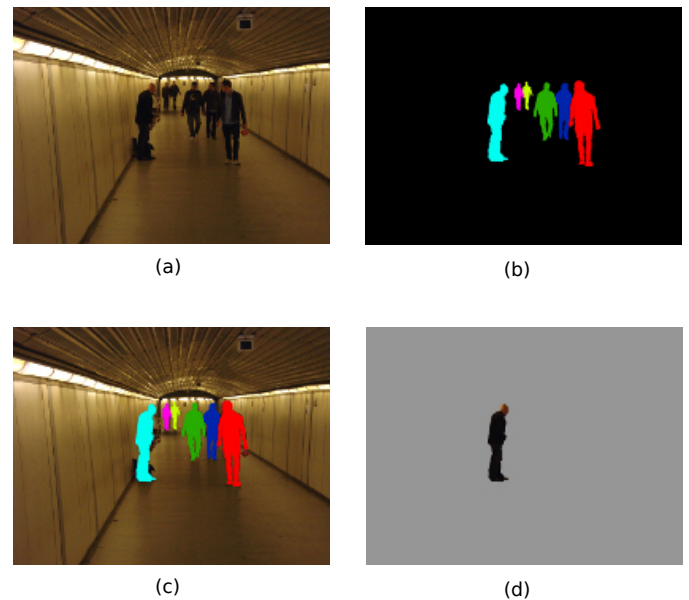


Figure 3. All algorithm steps for one frame: (a) the original image, (b) each color encodes one pedestrian, no background, (c) background image with color encoded blobs, (d) one pedestrian extracted from the image, based on [5]

Bad tracking or segmentation will lead for sure to bad object streams.

C. Preserving Privacy by De-Identifying Face Images

Since CCTV-s are not only standalone systems anymore face recognition software has become a major threat for one's privacy. Given an existing database of passport images the operator could use off-the-shelf systems to recognize and identify people in video images or streams. Newton et al. [4] is presenting a technique how on the one hand one's privacy is preserved while on the other hand the loss on image information is minimized.

1) *Face Recognition*: Before sketching how today's base-line face recognition algorithm called *Eigenfaces* works some definitions are necessary:

Definition 1: (Face) A Face Γ is the image of a face with a fixed width w and height h saved as an vector of pixels.

$$\Gamma = (px_{0,0} \dots px_{0,w} px_{1,0} \dots px_{1,w} \dots px_{h,w})$$

Definition 2: (Face Set)

A face set is a set of n face images:

$$G = \{\Gamma_i : i = 1, \dots, n\}$$

If a face set is called *person specific* when no two or more face images in the face set are belonging to the same person.

Definition 3: (Face Recognition Software) Given a face set G and a face image Γ a *face recognition software* F is matching Γ to the closest $\Gamma' \in G$.

Given these definitions we can describe the face recognition algorithm *Eigenfaces* presented in [4]. This algorithm will be later on used as test implementation for the presented de-identifying algorithms.

Having face images $\Gamma_1 \dots \Gamma_n$ given as training set

- 1) Compute the average face:

$$\Gamma_{\emptyset} = \frac{1}{n} \sum_i \Gamma_i$$

- 2) Subtract the mean from all face images:

$$\bar{\Gamma}_i = \Gamma_i - \Gamma_{\emptyset}$$

- 3) Build the Covariance matrix by using:

$$A = [\bar{\Gamma}_1 \dots \bar{\Gamma}_n] : Cov(A) = AA^t$$

- 4) Having built the Covariance matrix, basic math tells us that the Eigenvectors are the principle components of $Cov(A)$. We choose the n eigenvectors with the largest eigenvalue: $u_1 \dots u_n$. These vectors are called *Eigenfaces*. Some tricks are necessary to make this computation come out fast. For more details see [4].

Now having trained the system with all face images to get the eigenvectors $u_1 \dots u_n$ we can now determine how influential which eigenface on a particular face image is. This gives us the opportunity to run any traditional pattern matching on the problem. [4] has shown that this approach is both fast and reliable.

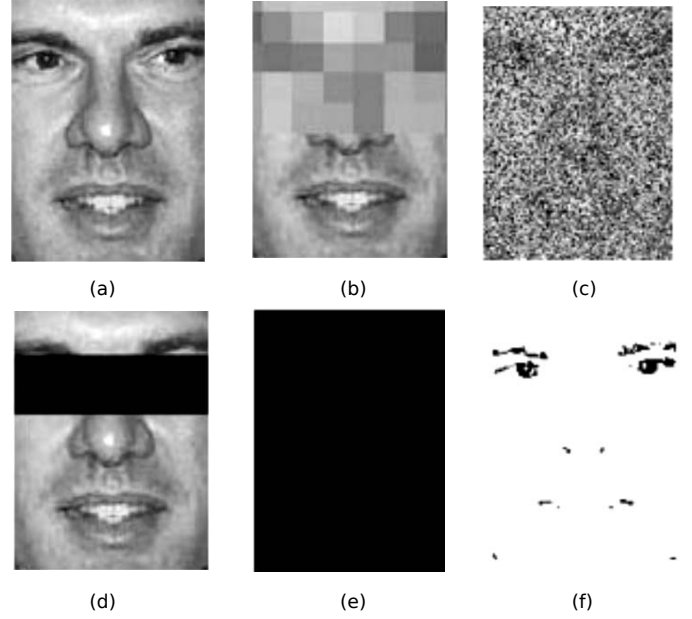


Figure 4. Example for a face de-identification of (a): (b) pixelation, (c) random noise, (d) bar mask, (e) black-out and (f) threshold, image from Newton et al. [4]

2) *Face De-Identification*: Face de-identification is already common in today's media. Whenever the identity of a person in a picture or video stream needs to be anonymized face de-identification is the best quality. Today these methods are mostly common (example in Figure 4):

- *pixelation*: The person's face or eye region is undersampled and therefore looks "pixelated".
- *random noise*: The person's face or eye region is overlaid with a random noise.
- *bar mask*: The person's eye region is overlaid with a black bar, the remaining face is visible.
- *blackout*: The person's face is overlaid with a black box.
- *threshold*: Every pixel in the face image is either mapped to black or white, depending on a greytone threshold.

All of these methods are destroying image information to achieve anonymity for the masked out individual. But for the viewer the video does not look unaltered anymore. A major goal of [4] is to find an algorithm that on one hand is as effective as these algorithms and on the other hand preserves a natural unaltered look of the image.

3) Formal Face De-Identification:

Definition 4: (Face De-Identification) Given two face sets H, H_d a function $f : H \rightarrow H_d$ is *de-identifying* a face image $\Gamma \in H$ by mapping $f(\Gamma) = \Gamma_d \in H_d : \Gamma_d \neq \Gamma$. Such a de-identification is called *effective* if it complies to a previous stated claim C .

An easy example for a de-identification function f is the *BlackOut()* function, this function is just replacing the face with black pixels. The de-identification process itself does not guarantee privacy at all, as long as more features are present an identification could still be possible. A drawback of *BlackOut()* is the fact that the image has to be modified in a very destructive way. To measure how a de-identification

algorithm is destroying information in the face image Newton et al. [4] has defined the metric $loss()$.

Definition 5: (loss()-metric)

$loss()$ is quantifying the amount of information a face de-identification algorithm has removed from the image. This could be for example based on:

- entropy
- percentage of changed pixels

For example the use of $BlackOut()$ does remove all information from the face image, therefor $loss() \rightarrow max$. The goal is to find an *effective face de-identification function* with $loss() \rightarrow min$.

Definition 6: (k-anonymity on face images) Given a person-specific face-set H and a de-identified face-set H_d with an de-identification function $f : H \rightarrow H_d$ and an inverse de-identification function $f^{-1} : H_d \rightarrow H$. H_d is *k-anonymized* if for $\Gamma \in H$:

$$f(\Gamma) = \Gamma_d \in H_d \Rightarrow |f^{-1}(\Gamma_d)| \geq k$$

If additionally exists no face recognition function h , which is capable of unveiling the identity of $\Gamma_d \in H_d$ with a probability of $> \frac{1}{k}$ then f is a *k-Same de-identification function* and H_d is a *k-Same de-identification*.

Clustering of face-images into groups still leads to a gain in privacy, every image in the cluster would be replaced with the new created anonymized image Γ_d .

4) *Algorithms for k-same de-identification* : The authors of Newton et al. [4] are suggesting two algorithms for protecting one's privacy against detection with any face detection software (not only when *Eigenface*-based tools are used). Given is a function

$$[img, ev] = recognition(\Gamma)$$

which is returning two lists in descending order of relevance: one list img with all matching face images and one list ev with in each element a vector of factors. Every factor belongs to one *Eigenvector*.

The function $kSamePixel()$ has two parameters:

- face set (images) H
- privacy constant k

The function's task is to de-identify all faces in H with respect to the privacy constant k .

Algorithm III.3:

```

1 funct  $kSamePixel(H, k)$ 
2   for  $\Gamma \in H$  do
3      $(img, ev) = recognition(\Gamma)$ 
4     //no. of matches
5      $assert(\|img\| == \|ev\|)$  //is the same
6     if  $\|img\| \leq k$ 
7        $k = \|H\|$  //cluster remaining face images
8     end
9      $closest = (img[1 \dots k], ev[1 \dots k])$ 
10     $avg = \frac{1}{k} \sum_i closest.img[i]$ 
```

```

11   for  $1 \dots k$ 
12      $H_d.push\_back(avg)$ 
13   end
14   for  $1 \dots k$ 
15      $H'.remove(closest.env)$ 
16      $H.remove(closest.img)$ 
17   end
21   end
23   return  $H_d$ 
24 end
```

The eigenface-space, also called *facespace*, equivalent to $kSamePixel()$ is $kSameEigen()$ which works not directly on the images but on the eigenvector factors. The differences to $kSamePixel()$ are marked with \blacktriangleleft , the function $kSameEigen()$ has two parameters:

- face set (images) H
- privacy constant k

Algorithm III.4:

```

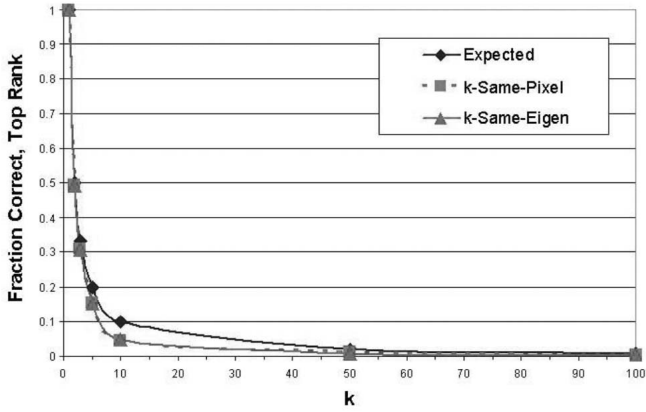
1 funct  $kSameEigen(H, k)$ 
2   for  $\Gamma \in H$  do
3      $(img, ev) = recognition(\Gamma)$ 
4     //no. of matches is the same
5      $assert(\|img\| == \|ev\|)$ 
6     if  $\|img\| \leq k$ 
7        $k = \|H\|$  //cluster remaining face images
8     end
9      $closest = (img[1 \dots k], ev[1 \dots k])$ 
10     $avg = \frac{1}{k} \sum_i closest.ev[i]$   $\blacktriangleleft$ 
11    for  $1 \dots k$ 
12       $H_d.push\_back(avg)$ 
13    end
14    for  $1 \dots k$ 
15       $H'.remove(closest.env)$ 
16       $H.remove(closest.img)$ 
17    end
22   end
24   return  $H_d$ 
25 end
```

5) *Correctness*: Newton et al. [4] have proofed the correctness as well as the fact that there can't exist any face detection software to recover the identity information out of the face images. The basic idea is straightforward. Both algorithms are averaging over k faces and replacing these k faces with the average. The full formal proof can be found in the work of Newton et al. [4]. The authors also evaluated the $kSamePixel()$ and $kSameEigen()$ functions against other commonly used privacy preserving techniques, in particular with Mask, Pixelation, Random Noise and Threshold (for examples see Section III-C2).

It turned out that for the naive recognition (altered probe image against original image) case all algorithms with exception of the Pixelation are achieving correctness rates of 0% – 2%. Pixelation on the other hand turned out to still match with 99% correctness and is therefor not a recommended choice for altering the images, although it is used widely in television.



(a)



(b)

Figure 5. (a) Face images on top processed by $k\text{SamePixel}()$, on bottom by $k\text{SameEigen}()$. Both algorithms with naive recognition performance for $k = 2, 3, 5, 10, 50, 100$. (b) performance of naive recognition on faces de-identified by $k\text{SamePixel}()$ & $k\text{SameEigen}()$, from Newton et al. [4]

The correctness rates for different privacy constants k are plotted in Figure 5. On the x -axis the privacy constant k is displayed, which equals the number of averaged images. The y -axis shows how accurate the matches of an eigenface face-detection algorithm have been, with 1 being totally correct and 0 being totally incorrect. This clearly shows that this state-of-the-art face-detection algorithm is not able to identify something significant when a privacy constant k of at least 50 is used.

6) *Conclusion:* With $k\text{SamePixel}()$ and $k\text{SameEigen}()$, the authors of [4] have presented preservative de-identification functions with minimal $\text{loss}()$, which are as effective as today's commonly used techniques. From all current used approaches these algorithms are preserving most part of the face images by still guaranteeing anonymity from current and upcoming face recognition systems. The authors did not elaborate how this system could be used in a real-life environment, especially performance issues are an open question in this case.

IV. CONCLUSION

The protection of privacy in a video surveillanced environment as research field will gain importance when CCTV's presence is going to increase as it did so far. Even areas which are currently not under surveillance are going to be equipped with CCTV equipment soon.

While most of the research in this area is still in early stages, IBM for example has already released² a privacy aware surveillance system consisting of a framework together with privacy aware cameras. Equipped with many state-of-the art privacy preserving techniques (e.g. pedestrian black-out) this surveillance system has introduced privacy awareness to mass market of surveillance technology.

Every technical computer vision/signal processing based privacy algorithm needs as foundation an underlying infrastructure to maintain security and privacy. Having equipped any video surveillance system or network with such an infrastructure (see Section I) various ways of filtering the video streams are possible. The use of mobile devices (see Section III-A) as source of one's trajectories induces structural problems which are making this kind of approach almost infeasible. Decomposing the video into object streams (see Section III-B) is a practicable approach which fulfills the most requirements for a privacy-aware rerendering layer. When it comes to de-identifying faces in video streams the approach presented in Section III-C does lead to anonymity while minimizing the loss of image information.

REFERENCES

- [1] G. Armstrong and C. Norris, *The Maximum Surveillance Society: The Rise of CCTV*, 1st ed. Berg Publishers, Sep. 1999.
- [2] J. Brassil, "Using mobile communications to assert privacy from video surveillance," in *Parallel and Distributed Processing Symposium, 2005. Proceedings. 19th IEEE International*, 2005, p. 8 pp.
- [3] H. Chen, H. Lin, and T. Liu, "Multi-object tracking using dynamical graph matching," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 2, 2001, pp. II-210-II-217 vol.2.
- [4] E. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, no. 2, pp. 232-243, 2005.
- [5] F. Z. Qureshi, "Object-Video streams for preserving privacy in video surveillance," in *Advanced Video and Signal Based Surveillance, 2009. AVSS '09. Sixth IEEE International Conference on*, 2009, pp. 442-447.
- [6] A. Senior, S. Pankanti, A. Hampapur, L. Brown, Y. Tian, A. Ekin, J. Connell, C. F. Shu, and M. Lu, "Enabling video privacy through computer vision," *Security & Privacy, IEEE*, vol. 3, no. 3, pp. 50-57, 2005.

²<http://www.research.ibm.com/peoplevision/videoprivacy.html>

Ubiquitous Devices: History and Research Trends

Hendrik Burmann – {firstname.lastname}@uni-ulm.de

Abstract—As we take a look at today’s world, we will find more and more computers around ourselves. The vision of computers embedded in the real world, fitting and disappearing in our environment, while not even having to think about them when using them, is called ubiquitous computing. This paper introduces the basic idea and the key elements of ubiquitous computing by giving an overview on some developments on ubiquitous devices like smart everyday objects, communicating with people and with other objects.

I. INTRODUCTION

IN 1991 Mark Weiser, US-american scientist at Xerox, introduced the idea of *ubiquitous computing* [23] by prognosticating computers to get smaller and smaller and finally disappear in conventional objects and things.

Hardware gets cheaper and many devices are already connected via wireless network or Bluetooth. We also have the ability to make objects identifiable with tags like RFID or barcodes. So Weiser’s vision gets more and more applicable these days.

Like in the mid-1970’s, when personal computing was becoming popular, we are now at a stage, where a new era of computing is coming up. Networks and the internet allow us more mobility and communication between people. This stage can be seen as an interstage between the age of personal computing and the age of ubiquitous computing.

Ubiquitous devices can be seen as computational objects which are smart enough to analyze their situational context, communicate with other devices and support the user without requiring too much focus on the technology.

Today, many objects are already equipped with sensors, processors and very small chipsets. In the age of the internet, people usually work and communicate with computers. Furthermore, the idea of the *semantic web* is making web content readable and understandable to machines, so this may be a further step to a machine-to-machine or thing-to-thing communication. If we imagine people communicating with everyday objects and objects communicating with other objects, we will reach a scenario called *intelligent environment* or the *internet of things*. [18].

Many applications are developed for mobile devices, like PDAs and smart phones which communicate together via Bluetooth, IrDA (infrared port) or WLAN. These applications are assembled under the idea of mobile computing. Other devices are integrated in clothing and accessoires, which covers the field of wearable computing [14].

II. THE IDEA BEHIND UBIQUITOUS COMPUTING

A. Mark Weiser’s vision

Ubiquitous computing is the vision of computers appearing almost everywhere, of people surrounded by hundreds and



Figure 1. The three waves of computing (based on Leimeister, 2002 [16])

thousands of computational devices. Weiser wanted technology to disappear in the background, which is supported by the miniaturization of hardware, chips and devices. This also leads us to new concepts of user interfaces, as common menu and dialog systems, mouse and keyboard input are not very applicable in this domain [14].

B. The three waves of computing

Mark Weiser described the *three waves of computing* [16]: Many decades ago the predominant way of data processing work was done via *mainframe computing* with one mainframe computer and many users working on it. This way of working was popular until the mid-seventies, when *personal computing* was starting to rise. Personal computing means that one user is working with one computer, an idea which amongst others was pushed by Bill Gates’ vision of “a computer in every home” [3]. Finally Weiser saw a third wave which he called *ubiquitous computing* having many computers (e.g. smart devices and systems) delivering services for one user. Another terminology for ubiquitous computing is *pervasive computing*, established by IBM [9]. Pervasive computing is focussed on industrial and commercial aspects. Its aim is to use existing technology to *pervade* business processes and areas of life, in order to deliver e-business solutions that satisfy customers.

Nowadays, we seem to be positioned between the wave of personal computing and the wave of ubiquitous computing with a network built between personal computers for communication through the internet. In the last years the internet itself became of major interest to the user, and many personal computers and mobile devices are mainly used to access the internet, retrieve information from the world wide web and communicate with other people [18]. While today’s multimedia machines still have

a strong focus on the user's attention, Mark Weiser wanted technology to fade into the background while still offering service and support to the user, who now can fully concentrate on the solution of his tasks.

C. Tabs, pads and boards

During their research work the people at Xerox PARC developed three fundamental device types: The tabs, which are very small, inch-scale devices like post-it notes; the pads, which are about the size of a paper or a book and finally the boards, which are thought of displays up to big screens like a blackboard. Looking around we can see hundreds of objects which could be realized as a tab device, like door signs, labels on books and folders, notes, ID cards, clocks, tags and sensor processors of all kind. Tabs should be locatable, users should be able to read information from them and communicate with them or use them to control other devices, like pads or boards.

A pad could be thought of as a sheet of paper on a desk, like the application window on the desktop of a personal computer. We may find about 10 to 20 objects surrounding us which could be imagined as a pad, for example our mobile phone, personal data assistant (PDA), a picture frame, newspapers, books and magazines, notepads and calendars. They are spread around on our tables and desktops, hanging on the wall or resting on the shelf. Every pad object has its own appliance, a book for example is usually used for reading and perhaps making notes in it.

If we look for a board device we are able to find two or three candidates around, like a television, a computer monitor or a whiteboard in the office.

In Mark Weiser's vision all these tab, pad and board devices have been designed to communicate. While tabs are primarily used to label or identify something, pads are thought of a network application which is used by one person and boards finally can be used for working together on a problem or displaying information to a group of people. Later we will see how these three types of devices have been realized in different ubiquitous projects.

III. MAKING DEVICES SMART

There are many ideas how smart devices can support the user's life, like food products which send recipe recommendations to the fridge door, smart clothing monitoring a person's heart frequency and body temperature up to lawn sprinklers, which receive weather informations from the internet, measure the ground's wetness to finally control the water output, based on these informations [18].

How can we develop everyday objects and existing technology like mobile phones, subnotebooks or PDAs to become smart devices? There are certain key elements we should pay attention to.

A. Interfaces

First we need to think of a device as a system with interfaces to the world around it. A smart device needs to have three interfaces: One interface to the real world consisting of sensors

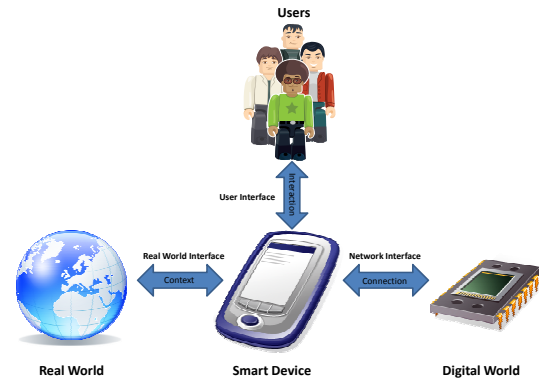


Figure 2. Smart devices and their interfaces (based on Leimeister, 2002 [16])

and processors, another interface to communicate with a whole network of other devices and finally a human-machine-interface to interact with the user consisting of actuators, displays and input controls [16]. User interfaces are often realized with technology like touchscreens, voice detection or gesture detection [14]. Polymere technology enabled the production of very thin and flexible displays [19] (e.g. E-Paper), today's E-Book readers use *electronical ink* to display information with a very high energy efficiency and allow the user to make notes on it with a touch-pen (e.g. smart paper) [18].

B. Processing technology

Computing power is an important aspect. Moore's Law [4] says that the power of microchips doubles every 18 months, which increases the efficiency of smart devices. But higher computing power of microchips and processors leads to higher energy consumption, so energy-efficiency is still a big challenge [17].

C. Miniaturization

Technology gets smaller and smaller and can be mass-produced very cheaply (e.g. RFID chips). The size of hardware is an important aspect to make devices disappear in the background. Latest developments lead to very tiny sensors not larger than a grain of sand, which are called smart dust. These wireless sensor-networks consist of a big amount of smallest ad-hoc networking sensors which are placed into the environment, e.g. for the appliance of monitoring [18].

D. Identification

In a system with multiple objects sharing information, it is necessary to apply unique identifiers to each object. The two most important technologies are Radio Frequency Identification (RFID) and visual barcodes [17].

RFID tags are tiny microchips with a transponder. Wireless RFID readers can read and process the information on these chips. They operate without the user's intervention and combined with RFID readers embedded in mobile phones, they

build a fundamental element of ubiquitous technology to make tagged objects identifiable.

Visual barcodes are 2D graphical pictures with encoded information in it. Visual barcodes hidden in the background have already been developed, like the Fujitsu Fine Picture code [21] which is embedded into photographs. Another example is an audio code developed by NTT DoCoMo [6], which allows text information to be embedded into music.

E. Context awareness

The real world interface is an important aspect in the design of smart devices to ensure context awareness. This means that applications and devices are able to identify the context of their use and offer the proper service to the user. This for example can be done by identifying the user and his role, the location where the device is used, the situation in which the device is used or the position between multiple devices in a system. Usually smart devices have multiple sensors to gather information for specific functions (like GPS-location, visual context, loudness, pressure, gravitation, ...).

For example, with the Global Positioning System (GPS) a device (e.g. a mobile phone or a PDA) can locate its own position [17]. This is already often used in car navigation systems and also in pedestrian navigation systems. As this only works outdoors, there are alternative ways to locate a device. A mobile phone's location can be detected by location the nearest antenna via the GSM signal strength. An alternative way is Wi-Fi location tracking, which works in a similar way, but delivers more precise results, due to the range of the Wi-Fi access points, which is smaller than the range of the cell phone antennas.

The values of these multiple sensors are usually put together in relation to make context analysis possible [14, 13]. This, for example, can be done by a sensor network. Usually each sensor node consists of sensors, microprocessors, memory and components to build a wireless peer-to-peer connection to other nodes [17]. These sensor networks for example are being used in cars for collision detection and airbag controlling. These networks can be based on existing technology like Bluetooth, WLAN, GPRS, UMTS and ad-hoc networking. Also, in mobile phones and game controllers like the Nintendo Wii we can find gravity and acceleration sensors, sending their data via Bluetooth [14].

IV. DEVELOPMENTS AND RESEARCH PROJECTS

A. Mediacup: Context awareness in non-computational objects

The Mediacup project was developed by Michael Beigl, Hans-Werner Gellersen and Albrecht Schmidt at University of Karlsruhe [11]. It can be seen as a complementary trend to mobile device projects, where computational devices build an interface to the real world. In this project, a non-computational real world object has an interface to the computational world through multiple integrated sensors. The Mediacup is based on the idea of the *active artifact*. Active artifacts are objects which are equipped with sensors to analyze their individual situational context. They also have the ability to autonomously share their context information with other devices within a



Figure 3. The Mediacup - An ordinary coffee cup equipped with sensor technology [1]

local range and allow other devices to use this information for their context analysis [13].

The Mediacup is an ordinary coffee cup coming with implemented hard- and software for context analyzation and sharing information. The hardware is embedded on a board of only 3 mm height, without affecting the cups shape, size, weight or use. Its energy source is wirelessly recharged when putting the cup on its saucer. So it meets the requirements of disappearing hardware.

The sensors embedded in the cup are used to analyze the temperature of the cup, to detect if the cup is being moved around or placed on a surface.

The software can compute four movement states of the cup:

- 1) *Cup is stationary*
- 2) *Someone is drinking from the cup*
- 3) *Someone plays with the cup*
- 4) *Someone carries the cup around*

As movement is a very quickly changing parameter it needs sensors with very fast reading time, in this case the sensor delivers new values every 20 ms after the first movement is triggered.

The temperature sensor delivers the current temperature and two states of the cup:

- 1) *Cup is filled up*
- 2) *Coffee is cooled off*

As mentioned before, energy efficiency is an important aspect in the design of ubiquitous devices. The problem is addressed by putting the processor into sleep mode whenever possible and interrupt it when the cup is moved. This way, 15 minutes of recharging will last for 12 hours of operation with the cup.

Information sharing is message based. Every two seconds the cups broadcast their context information together with a unique ID via IrDA. IrDA receivers are for example built into laptops connected to the ethernet or doorplates (so called *SmartPlates*) connected to a bus system. The communication model in this project is called the RAUM system (see figure 4 on page 37), where the cups communicate with other active artifacts, like the before mentioned SmartPlates. For example, the doorplate can detect that several hot cups are in a room and display that there is a meeting in this room. A coffee machine connected

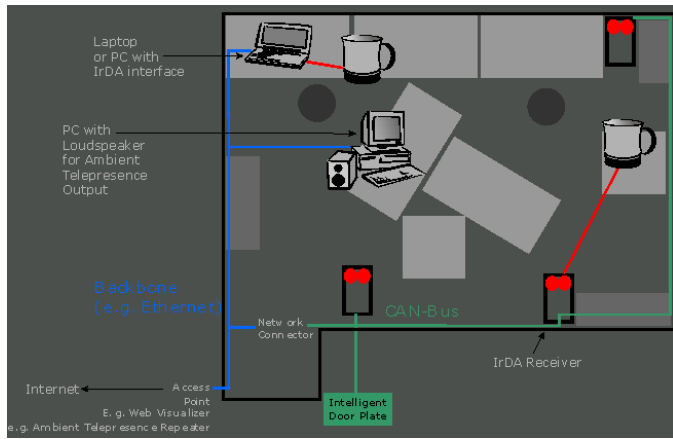


Figure 4. The RAUM system enables communication between Mediacyups and other devices [11]

to the Ethernet can detect how many empty cups are around and start making fresh coffee [11].

B. A mobile product recommendation system interacting with tagged products

This project introduces a consumer product recommendation system based on a model called APriori [22]. It shows how mobile devices can be graded up to support the user's everyday life. The system uses mobile phones equipped with RFID or barcode scanners to detect tagged products and to receive and send product information by connecting to a database server via internet. The project benefits from the fact that over 90% of the people living in well-developed countries are in possession of a mobile phone, which are able to build up internet connections. Some up-to-date mobile phones also support near-field-communication like RFID scanning or pattern recognition like barcode scanning via built-in cameras. Besides that, almost all of today's shopping products are tagged with barcodes or RFID chips.

The project was realized with a Nokia 6131 which includes a Java runtime for mobile applications and a RFID tag reader. The product server runs a Java Enterprise Edition application with a underlying JavaDB database.

Today especially younger generations are willing to share their experiences with other users on the web. Usually web shops today are supplied with recommendation and review functionality which have a strong effect on the sales of a product.

Product recommendation can be divided into two types:

- 1) *Product rating* by measuring product properties on a numeric scale
- 2) *Product reviewing* by describing experiences on a product as continuous text

Product rating is the preferred type in this project, because the user does not have to type or read too much on his mobile phone. Primarily typing is not very comfortable on a phone keyboard.

On his shopping tour the user is now able to use his mobile phone to scan a product's tag and receive ratings made by other



Figure 5. Receiving product recommendations with APriori [22]

users, helping him to decide which product to buy. Alternatively the user can rate a product himself by choosing rating criterias linked to the scanned product and entering rating values.

The project is not entirely fulfilling the criteria of disappearing hardware, since the mobile phone is always needed as a computational device to receive or send recommendations, but it shows how existing technology can be enhanced to interact with the real world.

C. SmartTV: A multi-person user's behaviour analysis and program recommendation system with iTV

The SmartTV [12] research also focusses on a recommendation system, in this case for the consumer of digital television. It is based on a multi-person behaviour analysis. TV watching analysis and recommendation is a well researched topic. The new aspects of this research project are the support of multiple users at the same time and the almost invisible login process. In other projects usually one user at a time is logged in for example by typing in a code on the remote control.

The project uses a set-top-box for receiving digital television program, which also implements the behaviour analysis and the interaction with the users. For the identification of the users sitting in front of the television, each user must be tagged. Instead of sticking RFID chips on each user, this is done by detecting the users' mobile phones via Bluetooth. This of course requires every user to carry his phone with him, but as we all know, this usually is the case, anyway. The user is identified by the *Bluetooth address* of his phone, which is a unique ID, similar to MAC addresses on network cards.

After identifying the users in front of the TV, the set-top-box can update each user's watching history, on which the recommendations are based. Of course, the system frequently is looking for users entering oder leaving the room. The users' IDs and watching history are sent via the digital TV return channel to a database server which is doing the analysis in order to recommend each user programs available in the electronic program guide (EPG).

Recommendations can now be sent back to the set-top-box and distributed via Bluetooth to each mobile phone, where the user can decide if he wants to watch the recommended TV program.

The project shows how common technology can be used for tagging users, interacting with them and simplifying technical

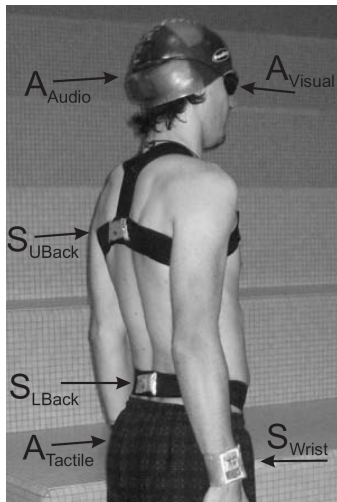


Figure 6. Sensors and actuators of the SwimMaster [10]

processes like login procedures to avoid interception of the user's common behaviour.

D. SwimMaster: A wearable assistant for swimmers

The SwimMaster project is an example for a wearable computing device. It was built to assist a swimmer achieving his training goals by monitoring his swimming style and giving feedback [10]. Its real world interface is consisting of acceleration sensors and micro-controllers. The user interface consists of visual, acoustic and haptical feedback actuators.

Three parameters are monitored by the acceleration sensors to improve the swimmer's reduction of resistance and to increase his propelling force:

- 1) *body balance*
- 2) *body rotation*
- 3) *stroke efficiency*

Nowadays, the common methods for a swimmer to improve his swimming style are *self-perception*, *supervision by a trainer* and *video-analysis*. The first method allows the swimmer to react instantly to faults, but maybe he might not be able to detect the faults by himself. So it may be better to be assisted by a trainer, but a trainer can only observe one swimmer at a time and also might have problems to detect faults happening under the water surface. Finally, video-analysis might solve the underwater problem, but the video-review can only be done after the training, so the swimmer cannot react instantly. At last, there is always the problem that a swimmer might need to train alone, because there is no trainer available. At this point the SwimMaster comes into play to directly help the swimmer keeping track of his swimming technique.

For the monitoring, three acceleration sensors are fixed to the swimmer's body. One is placed on the right wrist, like a wristwatch, the others are tight to the swimmer's back with belts. For the feedback system there are three different actuators. The visual actuator consists of two bi-color (red and green) LEDs, placed in the swimmer's glasses. The audio feedback is given through a 4 kHz beeper besides the ear and the tactile actuator is a vibration motor like those in mobile phones, placed on the swimmer's wrist (see figure 6 on page 39).

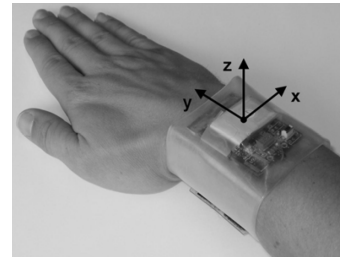


Figure 7. Orientation of the axes of the SwimMaster's wrist [10]

The analysis of the swimmer's parameters is done by a micro-controller, equipped with 1 GB of flash memory and a rechargeable battery.

The wrist sensor (see figure 7 on page 39) is monitoring the arm movements to extract the stroke parameters. The two sensors at the back measure the body movement, balance and rotation. Also, push-offs and turns on the wall of the swimming pool can be detected from the recorded values.

The wearable swim assistant allows the swimmer to instantly react to swimming faults, indicated by the feedback system. The sensor and feedback system could be imagined to be integrated in a swim suit to be applicable in everyday's swimming training.

Experiments showed, that the visual feedback given by the LEDs resulted in best reaction times, while the worst reaction times were observed for the acoustic feedback, because of the amount of background noise caused by the flowing water and the public swimming pool environment.

E. Pleo: A popular intelligent toy robot

The common ground of robots and ubiquitous devices is that they both own sensors and actuators and micro-controllers to read, analyze and react to their environment. While military, medical or industrial robots are programmed to automate processes and support workflows, a toy robot's goal is entertainment [20]. There are simple toy robots up to more sophisticated ones. The more human-like or pet-like a robot looks like, the more obvious it is for the user to compare the robot's behaviour to its real world counterpart. Examples for human- or pet-like toy robots can be found in the Sony QRIO and AIBO robots, but the development of both products has been stopped in 2006 [5, 15]. Clever enough, the developers of Pleo decided to give their toy robot the appearance of a baby dinosaur, so it is harder to compare its behaviour, sounds or colour of skin to any known real world animal.

Scientific philosophy and psychology books define an intelligent system as one which implements learning methods and develops complex decision rules based on sensory information [20]. Honda developed the perhaps most intelligent robot these days, which is called *ASIMO* and reacts on speech recognition, gesture recognition, is able to memorize faces and can even walk stairs [20, 2]. Of course, this high developed robot is not being sold as a toy, so Pleo stays the toy robot of our choice.

Pleo is thought of as a robotic life form based on a one-week-old *Camarasaurus*. The appearance of Pleo is very natural, its robotic skeleton is covered by a smooth elastic thermoplastic

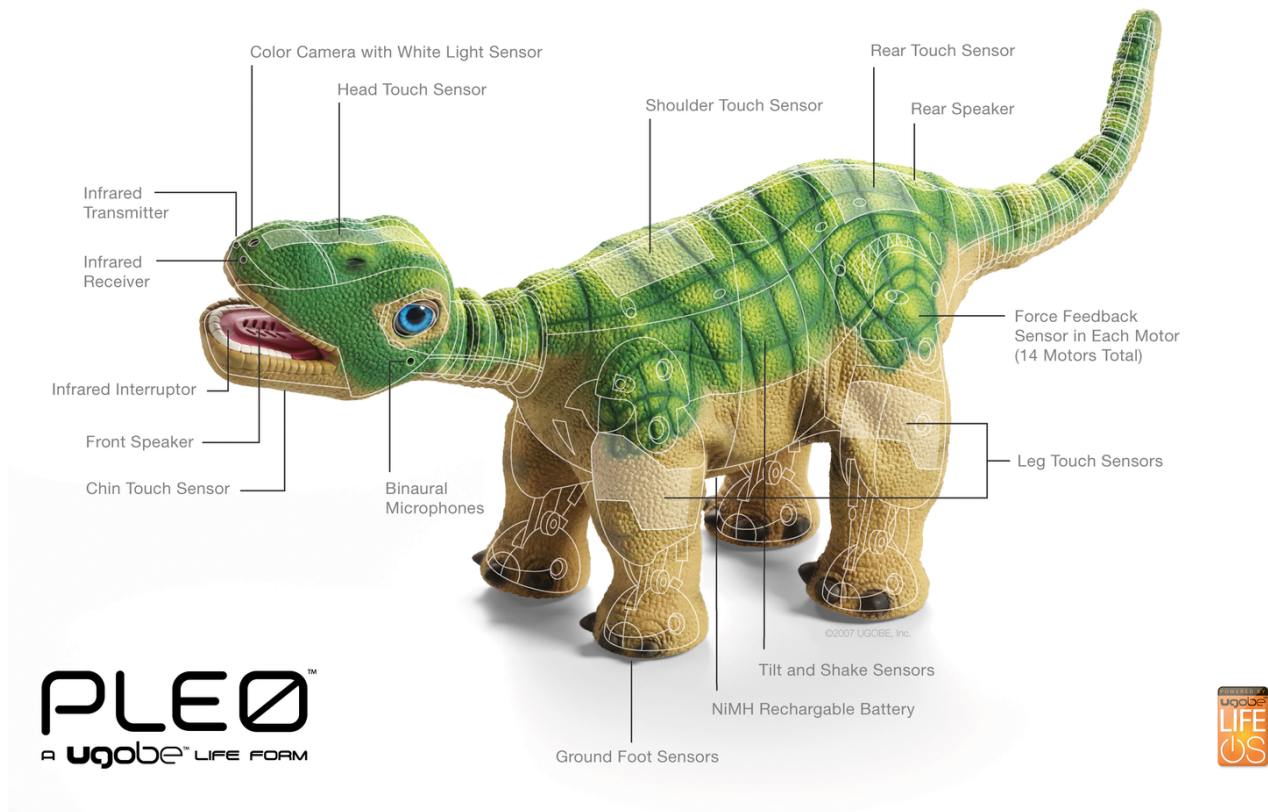


Figure 8. Sensors and actuators of the Pleo dinosaur toy

skin, its movements and actions look very fluid and spontaneous. The context awareness is realized by overall 36 sensors. Touch detection and force feedback sensors are placed all over the body, a color camera and infrared sensors at the head to detect light, motion and movement. Stereo microphones make it hear sounds and measure loudness. Contact sensors at the feet tell it, if it stands on a surface (cf. Mediacup) and acceleration and gravity sensors are inside to give it orientation. The firmware is called LifeOS and it can also read custom personality files via USB or SD-slot, influencing the toy's behaviour. Pleo's brain consists of a 32-bit ARM7 microprocessor and 4 motor control processors [20, 8, 7].

Unfortunately and despite all smart technology in it, Pleo is – aside from entertainment purposes and maybe educational aspects – comparatively useless to support people's life¹.

V. DISCUSSION

Going through this overview of research works on ubiquitous devices, some open questions may have appeared.

The development of technology is rapidly in progress, an increasing number of objects and devices are equipped with sensing, identification, processing and wireless connectivity. Ubiquitous computing might help to minimize the digital divide, when we manage to build technology operating in an invisible

way [17]. Also, feedback collected by the devices and sent to a central database server might increase product quality and satisfaction of the user [22].

But how invisible, how embedded can a device be? If we imagine the SwimMaster with all its sensors and actuators, maybe being built into a swim suit, it is questionable, whether the hardware may be a handicap to the swimmer and if it is perhaps influencing his swimming performance. User interfaces will have to be rethought. Though mobile phones and devices provide a perfect gateway between the user, the virtual and the physical world, they still draw much of the user's attention while he is operating the mobile phone's keys or handling the navigation.

Besides usual problems in software and hardware projects, like the reliability and dependability of systems, the user interface design or the system architecture, there are big concerns about the amount of collected data, privacy infringement and data security. For example, public concerns about privacy already affected the development of the RFID chips: Earlier versions of the RFID chip transmitted their informations in plaintext, while later versions already use encryption to ensure privacy. Another problem might be the challenge of controlling an ubiquitous system. Error traceability gets much harder in a system, where multiple devices influence each other [17].

Also, the problem of energy sources and battery life has to be solved. It's easy to imagine a system collapsing, because some of its devices ran out of battery power. The more complex

¹See also: "Pleo, useless robot dinosaur", YouTube video clip, April 2009, Link: <http://www.youtube.com/watch?v=D0KXGZgrXvA>

Table I
COMPARING THE EXAMPLES

	Mediacup	Mobile recommendation system	SmarTV
Interfaces	User: Ordinary cup Real world: Sensors Network: IrDA infrared port	User: Mobile phone keyboard and display Real world: Camera, RFID reader Network: HTTP	User: Mobile phone and TV Real world: Bluetooth to identify users Network: Bluetooth and TV return channel
Processing technology	Microprocessor: PIC 16F84 with reduced clock speed of 1 MHz Temperature sensor: Dallas DS1621 chip Accelerator: Three ball switches Battery: Two capacitors (1F each) Charge: 15 mins, Operation: 12 hrs	Nokia 6131 NFC with Java mobile application environment	Processing is done by the set-top-box and the program database in the TV-provider's back-office
Miniaturization	Hardware board height: 3 mm	Depends on the size of the mobile phone (Nokia 6131: 92 mm × 49 mm × 21 mm)	Depends on the size of the set-top-box and the mobile phones
Identification	Unique ID for each cup	Product identification via barcode or RFID	Bluetooth address to identify users
Context awareness	Temperature, acceleration, location awareness via IrDA	Barcode scanning via built-in camera, RFID scanning	User recognition
Classification	Active artifact	Pervasive computing	Pervasive computing
	SwimMaster	Pleo	
Interfaces	User: LEDs, beeper, vibration motors Real world: Three acceleration sensors Network: None	User: Pet-like behaviour, sounds, movement, haptics Real world: 36 sensors, camera, microphones Network: Infrared sensors to communicate with other Pleos, USB port, SD/MMC card reader	
Processing technology	Microcontrollers: Atmega644P for recording, Atmega48 for feedback Battery: 3.7 V, 370 mAh, Operation: 15 hrs	Microprocessor: One 32-bit Atmel ARM7 processor, four 8-bit motor control processors, one 16-bit camera processor Battery: NiMH package with 4 hours recharging for 1 hour of operation	
Miniaturization	Size of the SwimRecorder on the swimmer's wrist: 36 mm × 42 mm × 12 mm with a weight of 34g	52.5 cm long, 15 cm wide, 19 cm tall (a full sized adult Camarasaurus would be 18 meters long)	
Identification	None	Recognition of other Pleos	
Context awareness	Acceleration sensors to analyze the swimming movements	36 sensors for touch detection, detection of light, motion and movement, hearing sounds, measuring loudness, detection of surface contact, acceleration, gravity and orientation.	
Classification	Wearable devices	Smart toy	

a system is, the more energy is needed. While the Mediacup only needed 15 minutes of charging to operate 12 hours, the Pleo toy robot already needs 4 hours of charging for only 60 to 90 minutes of playing [8].

VI. CONCLUSION

Now we have got an idea how Mark Weiser and other pioneers of ubiquitous computing imagined a world of new computing devices. We learned what is needed to make devices smart and to let them disappear in the real world. Some points to focus in the future may be the miniaturization of hardware and, even more important, the exploitation of new, more efficient energy sources. We also have seen how these

visions were implemented in various projects and researches. Current developments deal with household objects, toys, mobile devices up to smart clothing. Today, there is a strong focus on the advancement of mobile devices, because they are very widespread in today's population. We also have got an idea where problems may occur, like concerns on privacy and data collection. Reconciling technology development with data protection may be one of the biggest challenges in this domain. If we find an agreement between the technology-based support and the privacy protection of the user it might encourage the acceptance of these new technologies.

REFERENCES

- [1] “mediacup@teco,” 2001. [Online]. Available: <http://mediacup.teco.edu/overview/german/uebersicht.html>
- [2] “Honda Worldwide | ASIMO | Technology,” Dec. 2002. [Online]. Available: <http://world.honda.com/ASIMO/technology/intelligence.html>
- [3] “Microsoft’s tradition of innovation,” Oct. 2002. [Online]. Available: <http://www.microsoft.com/about/companyinformation/ourbusinesses/profile.mspx>
- [4] “Moore’s law,” 2005. [Online]. Available: http://download.intel.com/museum/Moores_Law/Printed_Materials/Moores_Law_2pg.pdf
- [5] “heise – Sony schläfert Aibo ein,” 2006. [Online]. Available: <http://www.heise.de/newsticker/meldung/Sony-schlaefert-Aibo-ein-169676.html>
- [6] “Audio barcode | technologies | NTT DOCOMO global,” Jul. 2007. [Online]. Available: <http://www.nttdocomo.com/technologies/future/audio/index.html>
- [7] “Pleo technical specs at pleo HQ,” 2009. [Online]. Available: <http://pleohq.com/pleo-101/technical-specs/>
- [8] “Pleo website FAQ,” 2009. [Online]. Available: <http://www.pleoworld.com/Support/FAQ/FAQGeneralPleo.aspx>
- [9] W. S. Ark and T. Selker, “A look at human interaction with pervasive computers,” *IBM Systems Journal*, vol. 38, no. 4, pp. 504–507, 1999.
- [10] M. Bächlin, K. Förster, and G. Tröster, “SwimMaster: a wearable assistant for swimmer,” in *Ubicomp ’09: Proceedings of the 11th international conference on Ubiquitous computing*. New York, NY, USA: ACM, 2009, p. 215–224.
- [11] M. Beigl, H. Gellersen, and A. Schmidt, “Mediacups: experience with design and use of computer-augmented everyday artefacts,” *Computer Networks*, vol. 35, no. 4, pp. 401–409, 2001.
- [12] A. G. Foina, “SmarTV - a multi-person user’s behaviour analysis and program recommendation system with iTV,” in *Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on*, Mar. 2009, pp. 1–2.
- [13] H. W. Gellersen, A. Schmidt, and M. Beigl, “Multi-sensor context-awareness in mobile devices and smart artifacts,” *Mob. Netw. Appl.*, vol. 7, no. 5, p. 341–351, 2002.
- [14] C. Huhn, “Killerapplikationen des ubiquitous computing,” 2001. [Online]. Available: http://www.informatik.tu-darmstadt.de/BS/Lehre/Sem00_01/Ausarbeitungen/Christopher-Huhn-Ubicomp.pdf
- [15] M. Koelling, “Focus Online – Nachruf: Sony lässt Aibo und Qrio sterben,” 2006. [Online]. Available: http://www.focus.de/digital/diverses/nachruf_aid_104293.html
- [16] J. M. Leimeister and H. Krcmar, “Ubiquitous Computing und die Herausforderungen für das Informationsmanagement,” *wisu-Das Wirtschaftsstudium*, vol. 31, no. 10, pp. 1284–1294, 2002. [Online]. Available: [http://www.krcmar.in.tum.de/lehstuhl/publikat.nsf/intern01/C3528422BBDD765EC1256CD40040AF52/\\$FILE/02-34.pdf](http://www.krcmar.in.tum.de/lehstuhl/publikat.nsf/intern01/C3528422BBDD765EC1256CD40040AF52/$FILE/02-34.pdf)
- [17] D. Ley, “Ubiquitous computing,” *Emerging Technologies for Learning*, vol. 2, pp. 64–79, 2007. [Online]. Available: http://schools.becta.org.uk/upload-dir/downloads/page_documents/research/emerging_technologies07_chapter6.pdf
- [18] F. Mattern, “Ubiquitous Computing – Die Vision von der Informatisierung der Welt,” 2003. [Online]. Available: <http://www.vs.inf.ethz.ch/publ/papers/UbicompLogin.pdf>
- [19] K. H. Shin, “United states patent application: Flexible display,” United States, Jun. 2007. [Online]. Available: <http://appft1.uspto.gov/>
- [20] J. Somló, A. Molnár, A. Kutyepova, and D. Olchanski, “How intelligent the toy robots are?” Balatonfüred, Hungary, Jun. 2006. [Online]. Available: http://autonomousmobilerobots.co.uk/wp-content/uploads/articles/01_Somlo.pdf
- [21] G. Suzuki, N. Yamaguchi, S. Nakamura, and H. Chiba, “Mobile interaction using steganographic image on mobile display,” in *MobileHCI ’08: Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*. New York, NY, USA: ACM, 2008, p. 507–510.
- [22] F. von Reischach, D. Guinard, F. Michahelles, and E. Fleisch, “A mobile product recommendation system interacting with tagged products,” in *Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on*, Mar. 2009, pp. 1–6.
- [23] M. Weiser, “The computer for the 21st century,” *Scientific American*, no. Communications, Computers, and Network, Sep. 1991. [Online]. Available: <http://www.ics.uci.edu/dutt/ics212-wq05/weiser-sci-am-sep-91.pdf>

Market-ready Ubiquitous Systems

Sven Reichel – {firstname.lastname}@uni-ulm.de

Abstract—Ubiquitous systems are often future visions, which show perfectly integrated circuits in the daily life. But some approaches in varying sectors are already available today. The five domains emergency, automotive, energy, medical and payment are pointed out in this paper with different systems each. It presents major systems which can be bought already or will be on the market soon. This article attempts to provide a short introduction on the most innovative and less known systems rather than giving a comprehensive survey of all existing market-ready ubiquitous systems.

I. INTRODUCTION

ONE short ride on a bus or train at the rush hour is enough, to see at least one person working with his mobile phone. A lot of them are connected to the Internet and they are checking mails and appointments. As you can see every day, the little mobile devices become integrated in our daily life more and more and the acceptance increases. Smaller and increasingly powerful gadgets are available each year to fulfill the human desire to be online everywhere.

But there are not only mobile devices, like Apple iPhone, Amazon Kindle, iPod, Sony PSP or so on. Already today there is a huge market of other ubiquitous systems available and there will be more in the next years. There is also a tendency to connect every object to another as described in “The Internet of things” by J.P. Conti [5].

Mark Weiser used the term Ubiquitous Computing the first time in his well known article “The computer of the 21st Century”[25]. He describes a vision of many small intelligent devices, connected to each other and integrated in the daily life. There are two major goals of ubiquitous computing: First of all, the technological aspect to develop new powerful hardware, being small and having all necessary communication ability available. The other crucial point is the seamless integration into the human life. Therefore it is necessary, that the user does not know he is interacting with a computer, but after all he will receive even a better support at the appropriate time and place than from a classic computer.

I would like to point out different systems in varying sectors to give a short overview of the currently available systems. For sure it is far away from being complete, but it should provide a summary of unfamiliar systems. There is a short introduction to the emergency system e-Call which calls automatically an ambulance and will be in every car soon. Furthermore I point out automotive HMIs, with BMW ConnectedDrive as an example. A huge influence to our daily life will be intelligent energy solutions, like SMART Meters or virtual power plants. Ubiquitous systems can save lives too, as it is described in the medical section of this paper. Last but not least I will show some efforts to get rid of wallets and pay with mobile phones.

II. EMERGENCY SYSTEMS

In an emergency it is crucial to react as fast as possible, even one or two seconds can save lives. It does not matter whether it is a car accident, a fire, a terrorist attack or anything else where lives are at stake. There are lots of different approaches to minimize the time which elapse between the incident and the arrival of the rescue workers at the place of accident. Another important point is the preparation for the operation; to overlook the chaos and react as efficient as possible.

A. *electronic emergency Call*

Each year there are a lot of people dying in road accidents. In the twenty-seven countries of the European Union for example, died 42.955 people in 2006 [7]; which means every 12 minutes someone dies. Often a live cannot be saved because the ambulance is called too late. For Instance, somebody who approaches the scene of accident has to secure first and call the ambulance or police second and it is not unusual providing incomplete information to them. As a result the European Union funds the e-Call system[6], which calls an ambulance directly after an accident and provides system information like GPS, severity and much more on a specified basis. According to the European Union it is possible to reduce the response time up to 50 percent, hence save 2500 lives each year and reduce the severity of injuries by 10 up to 15 percent.

The eCall controller is connected to a GPS-navigation module, a mobile phone and the car’s system. When the car is involved in an accident, eCall gets activated by the airbag sensors if the severity is high and injuries are likely. It calls automatically the number 112, submits the exact geographical location and maybe some information from the car’s sensors, depending on the available data. Via the audio system and a microphone the emergency hotline can communicate with the people in the car if they are conscious to provide more information. In addition there is a key to activate the eCall system manually.

In every European Country it is possible to reach the emergency hotline free of charge by dialing 112. Even in Non-European-Countries this is often possible. Ecall uses the E112 system, a 112 call enhanced with a set of data, for example location based. Half of the European-Countries already support these calls in the mobile sector¹. Currently most of them are based on the mobile network with GSM-Localization, but it is quite possible to use GPS-data as well. Even without E112 support the eCall system can be used, for example it can read the set of data over every voice connection.

In a report [2] the European Commission introduced eCall to the European citizens in August 2009. In addition to the mentioned life saving aspects there are several other benefits.

¹Europe ertico-its: ESafety Support: <http://www.esafetysupport.org>

For example the traffic-incident management by authorities will be more efficient and as a result there will be less congestions caused by accidents. Not to mention the failure to communicate abroad; foreigners can get help without knowing the language of the country they are traveling in.

Thanks to mass production of the eCall devices it will cost less than 100 Euro each [2]. Additionally the system can synergize with the car's equipment. For example many cars have already a navigation system on board and the number will increase dramatically in the next years.

Another major aspect is data privacy. A lot of people are afraid to be tracked by an electronic device. Theoretically it won't be a problem for the eCall system to send the data continually. This would definitely improve the traffic management systems and reduce jams. But it could be used to collect personal data as well. To respect this concern the eCall system will only submit data in the case of an accident or by activation through the manual key.

In 2005 the European Commission and the automotive industry agreed to provide an option for eCall in every new car build after September 2009. This release was postponed because of national problems to handle eCalls at the emergency hotline in some countries. The first available systems are scheduled for 2011 and from 2014 onwards every car should be equipped with an eCall system.

The Fraunhofer Institute has developed another use case for the eCall system [10]. In former times there was a nationwide warning system for the German population. This system was uninstalled after the cold war. As a result the population can only be warned through radio, television and recently mobile phones. The researchers suggest to blow the horns of parking cars in the case of an emergency, thus everybody can turn on their radio or television. To activate the horns they use the eCall system, but whether this will be an option for the final system is questionable.

Till the final release of eCall, many vehicle manufacturers provide their own systems. For example BMW has a system called BMW Assist [3] to provide its customers with a lot of location based services and personal assistance in case of an accident or breakdown. Mercedes Benz has got a similar system as eCall called "mbrace"².

B. Siemens Video Fire Controller

Not only in the case of a car accident are lives at stake, but also in a burning house. There are a lot of different devices to alert the inhabitants and call automatically the fire fighters. In a little house it is quite easy for the rescue workers to identify the source of fire and get an overview of the situation. But if there is a huge public building with many people inside burning, it is quite challenging to get an overview and react efficiently.

Siemens has developed a Video Fire Controller for integration in a fire alarm system[20]. If there is an incident the system can stream the video data to the control center and allow the fire fighters to prepare themselves on their way to the place of action. So they can start to plan their actions in advance and

rescue people efficiently, because they know where to look for them.

The live-streaming is done through a LAN-interface compressed as MPEG-4. It is possible to access the stream locally or remotely with every browser. Additionally the system saves images on a SD-Card helping investigations afterwards.

Another common problem are false alarms. In Germany for example every third alarm occurs without an incident. The Siemens video fire controller enables the control center to identify false alarms and save a lot of money and time in which the fire fighters may act in a real situation.

III. AUTOMOTIVE

There is another major sector for ubiquitous systems as well: the automotive one. There are up to 70 computers in a single car which support the driver mostly unnoticed. In addition to driver assistance systems like ABS, ASR, ESP, et cetera, the multimedia parts are getting more and more important and are already a crucial point selling a car.

Due to the slash size of multimedia parts despite increasing the power, it is possible to integrate a whole entertainment system in a car. Some decades ago only the high-end cars had a navigation system, DVD-function and a surround speaker set. By increasing the quality of the provided systems and spreading of mobile devices which can store and play music, the desire arose to integrate these devices in the car environment. A lot of different accessories like adapter cassettes flood the market. Nowadays, even small cars do have different options for entertainment systems. In the high-end class there is already the rolling office available, where you can check your mails, book a hotel, surf the web or watch a movie in a quality comparable to your home entertainment system. For sure everything is connected to your mobile phone, office and computer; your life perfectly integrated in your car. Nearly every car manufacturer has these options in stock, but I would like to point out BMW ConnectedDrive as one of the mostly integrated ones.

BMW ConnectedDrive

BMW promotes its ConnectedDrive[4][17] system as the perfect fusion between human and machine. The driver is not replaced by a system, but he keeps always the control though the car supports the driver with as many information as he likes. Thus BMW created a symbiosis between car, driver and environment.

As mentioned before, BMW offers its customers BMW Assist to get a connection with the service center. But it is not just for making an emergency call, more likely it is for getting location based services over the in-car telephone. For example by looking for a good restaurant, the service center knows where the car is located, can support the requested information and in addition transfer this directly into the navigation system. However, this search can be made by the driver on his own, because there is a service called BMW Online which allows a direct connection with GPRS to the Internet. Additionally to find the interesting points with Google local search, it is possible to check emails, weather information, news, dates and even find a free parking place, because it is connected to around 2.850 car parks and supplied with the currently

²<http://mbrace.mbusa.com>

occupancy. BMW also connects the driver's computer at home with his car. For example, it is possible to export the local address-book and even trips, planned at the computer, to the car. Another bridge is built to the mobile phone; some services of BMW Online can be accessed outside the car by a mobile phone connected to the Internet. For instance, if the driver does not know where he parked his car, a mobile phone with GPS can direct him.

In 2008, 16.134 cars were stolen in Germany [8]. Even the best security systems can not prevent every theft. BMW equips their cars with a system called Tracking, which sends continually its GPS location to a service center after it was activated online or detected a theft. This may help finding the car by the police very fast.

Another connection is made to a certified BMW workshop. If a service is necessary, the car calls automatically the workshop and submits all relevant data to them. So the dealer knows exactly what to repair and may order the spare parts in advance. When everything is ready, the customer is called and an appointment can be arranged. In advance of this information policy the customer gets a service as short as possible and the dealer can plan its workflow much better.

In the future, BMW plans to connect the car to the driver's home as well. There are already tests in Munich where a house is equipped with some hardware which can be controlled from inside a car. For example, it is possible to turn on the lights or heater before arriving at home[4].

IV. ENERGY SECTOR

Every country, even the developing ones, depend on electricity nowadays. Just a short period without will show how often in our daily life we are using electrical devices. At the 14th of August 2003 over fifty million people had no power in North America. Only a little incident in Ohio caused an economic deficit of nearly one billion dollar simply in New York [22]. Not only the companies had to suffer, also the general public had problems to keep their food refrigerated and compete without public transport. In the future we rely even more on the electricity network, because the oil resources are getting smaller and for example all major vehicle manufacturer are working on electrical cars. According to a survey from WWF the energy demand in Germany will increase in 16 percent if every car will be powered by electricity [11]. For sure it is crucial for a world leading country to supply a reliable power network. To achieve this and save energy at the same time, SMART Meters are a big step forward.

A. SMART Meter

These days every house or flat has its own electric meter to measure the amount of power a consumer uses over a period of time. Normally the billing is carried out once a year. Therefore the consumer does not know exactly when he uses its power or, if at the worst, will be surprised or even not able to pay his amount. For the supplier it is also a problem not knowing in advance when the peak times are. Additionally they have to employ a lot of people for meter-reading.

To benefit both sides, the SMART Meter[13], an electronic

metering device connected to the power supplier, was introduced. Ideally the data exchange is done through powerline communication where the data will be sent directly in the power network, therefore there is no special wiring or a wireless communication module necessary. There is a two-way-data connection, so it is also possible for the supplier to send data to the consumer or even control a device remotely. For example it might be possible that the consumer just defines to run the washing machine at night and the exact moment is controlled by the supplier to optimize the utilization of his network. Furthermore the costs of labor decrease because they do not need people for meter-reading anymore.

Normally the consumer does not know when peak times are and when the power is expensive. Furthermore, this is irrelevant because the installed power-meters cannot determine dates the power is used at. Thus the user is not interested to pay attention when he switches on his electrical devices. For example there is no difference whether the washing machine is running right in the middle of the day or at night. As a result, the suppliers have to hold a significant amount of power for peak times to reduce up and downturns of the current and frequency. If frequency varies too much, the danger of a black-out increases. Most of the power suppliers use pump storage hydro power plants to save energy in low demand and regulate peak times with them. This problem will get more and more important by establishing renewable energy facilities, because for example there is just sun at day times and the wind also varies. Thus it is crucial for the suppliers to know exactly the state of their power network for using it efficiently and minimize the costs. By collecting the data from every SMART Meter that will be possible or even by controlling devices assessable. All in all the power consumption will smooth and there is a chance to save 3 up to 15 % of energy which will reduce the CO₂ emission and even some power plants could be shut down[23]. For consumers the transparency will increase and they always know how much and when they consume their power. They can display their current power consumption and may adopt their usage to the low cost periods. For sure the power suppliers will provide different price lists according to the time of use. As a result, each consumer can save money. Additionally he always knows how expensive his current bill is and may plan in advance for paying it.

The exchange of power-meters is already ongoing in different countries. For example EnBW ODR, a subsidiary enterprise of EnBW, installs SMART Meters in Stuttgart already[13]. In 2010 every power-meter in a new built house has to be intelligent. The European Union regulated to exchange every power-meter till the year 2020.

This change is just a beginning to introduce Smart Grids: to make the power network intelligent it is essential to know where and when power is consumed. So the power supplier can react and distribute their power dynamically. There are a lot of different approaches to buffer the electricity that is not used in low times, for instance in an electric car or water heated in a block heat and power plant as suggested by VW and Lichtblick[18].

B. Virtual Power Station

Another approach to save power and compete with renewable energy is a distributed system with lots of small producers, controlled by one control center. Using photovoltaics or wind energy result in many different small power suppliers, which have to arrange each other to feed enough and smooth energy into the power grid. One major problem of these systems is the unpredictability. Sometimes there is no wind or sun to provide wind power stations or photovoltaics, but the power is needed nevertheless. With a virtual power station it does not matter what kind of energy it uses at a specific moment, the supplier just guarantees that there is a certain amount to a particular time available.

Siemens developed a distributed system called DEMS for monitoring, controlling and scheduling a lot of mini power plants and appear as one normal to others[9]. It can predict the power infeed by taking account of the weather forecast or power demand by SMART Meters. DEMS is used and tested already in Sauerland with nine water power plants. The communication is done through standard mobile communication technology, thus it is quite easy to connect all power-plants over huge distances without wiring them.

A different problem is that the power cannot be transported without wasting a certain amount. As a result it is better to produce the energy where it is used. Additionally the investments in expensive land lines decline. Such as one public service in the Ruhrgebiet has not extended their power network, but installing mini power plants and connected them to a virtual power station instead.

In the future these systems will be more and more important, because a lot of regenerative little power stations will emerge. Figure 1 shows the different parts of a virtual power plant and a first approach to connect them for exchanging data. Every consumer and supplier is connected to a centralized communication network, from which it is controlled. As a result, this judgmental override provides the possibility to save energy and build up a reliable power network.

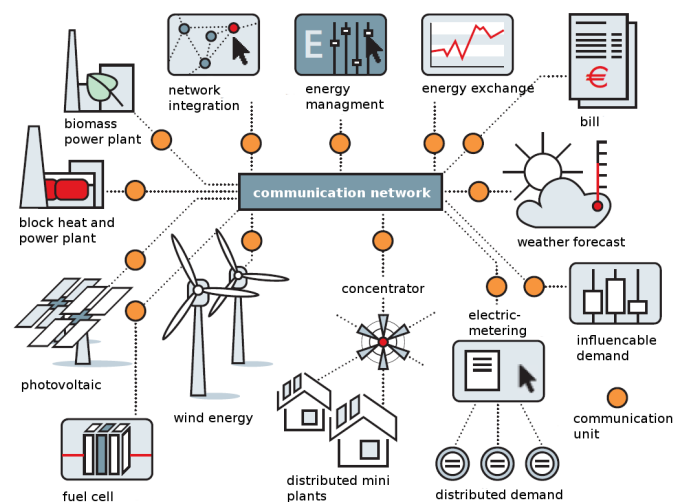


Figure 1. Virtual Power Plant [9]

V. MEDICAL

Nobody likes to be in a hospital with a serious disease. As a result, people react often very stressful or are afraid and not unusual there are life-threatening situations. To help people efficiently it is important to minimize their stress level and introduce a well organized workflow at the hospital. To obtain that there are systems already available supporting doctors and nurses in their daily work.

A. Augmented Reality

Often operations are critical in terms of time to save the patient. For example, cancer or cardiac arrhythmia have to be treated as soon as possible. But quite often the patient has to be x-rayed multiple times in advance to plan the operation perfectly. Augmented Reality is one approach to minimize examinations and support physicians in the operation room with multiple input from different devices like CRT, MRT or ultrasonic. To do so, these devices have to be connected and need a common data basis.

The surgical hospital at the Ludwig-Maximilian-University in Munich installed a system[21], which allows the doctors to view the real image of a patient, overlaid with the anatomical structure of an X-Ray (figure 2). For achieving this important feature they equipped an X-Ray apparatus with a normal camera. Next to the operating table there is a monitor the overlaid images can be displayed on. As a result the doctor can view a patient's anatomical structure, before cutting the skin and without x-raying multiple times. This easy system helps to reduce operation times and avoids mistakes.



Figure 2. Leg with overlaid anatomy [21]

B. RFID: surveillance of blood donations

Exchanging blood donations can be life-threatening if they will get swapped. Additionally, it is essential to keep them at a certain temperature anytime; in Germany for example many blood donations get wasted each year by not requiring the unbroken cool chain. Sometimes nurses and doctors are

stressed and making mistakes, then they need a system which will monitor their actions and react if a warning is necessary. Siemens introduced a system monitoring blood donations by RFID chips[19]. That for, every transfusion is attached with a unique RFID chip. Additionally to inform about blood type, the system monitors the blood's temperature in all its stages from donation, processing, testing, distribution, storage to the final infusion. Even sterilization, pasteurization or centrifugation does not damage the chip.

The system was tested at a hospital in Saarbrücken and was ready for common use in 2008. But still today it is waiting for certification through the German Regulatory Authorities.

C. Next Generation Hospital

In a hospital there are a lot of documents of different kinds. For example, every single inpatient has its own file with records about therapy, medication, treatment, x-rays, ultrasonic images, CRT diagrams and much more. Most of the time all these documents are not digital available and the nurses have to keep an eye on them. Sometimes they will get lost and the patient has to be examined a second time. Furthermore the required space is enormous.

At the Millenium Wing of the Southern Medical Clinic in Trinidad and Tobago, the paper-free century has already begun. Recently Siemens equipped the ontology station with up-to-date hardware and optimized the process workflow [12]. As a result, the inpatient's report is completely digital and can be exchanged easily between the threatening stations. The time for a treatment declined dramatically thanks to the efficient workflow.

In spite of the digital century it is unavoidable that patients have to carry their documents from one physician to another. Sometimes this is quite annoying, because a relevant document is still at another doctor and the examination has to be repeated. Additionally to financial costs, it is often stressful for the patient and may be result in side effects. The Fraunhofer Institute developed an electronic case file, which is used at the HELIOS-Clinic in Erfurt currently [16]. It is not a patient file where everything about this person is documented, but only containing the files necessary for one case, for example medical findings, operation reports, therapy advices, blood count and many more. The data is always up-to-date, so the family doctor can follow his patient at hospital.

At Erfurt the files are kept on a centralized server at the hospital and every physician threatening a patient with an electronic case file, can access it over the Internet. It does not matter whether the doctor is at hospital or not. Data security is very high, despite using the Internet for transferring information. The data connection is encrypted and everyone who wants to access it needs a hardware token for identification.

The project was initiated by the three major private Clinics Asklepios, Phön-Klinikum and Sana-Kliniken in 2006. One year later the first ten hospitals got equipped with the necessary software and hardware. The system is welcomed by nurses and physicians; only in Erfurt around 20.000 case files were created already.

VI. PAYMENT

Ages ago, people traded their goods with other products on markets. Nowadays nearly every trade is done with money. Therefore it is essential to take along any payment instrument. But normally it is not always the same, some vendors just take coins, others credit cards, cash cards or only cash. Additionally, every person poses a wide variety of discount cards, public transport tickets, identification card or driver licence and many more. Usually it is necessary to have them with you all the time. But maybe with Near Field Communication technology this will change soon. O2, Visa and DB are ready to introduce a payment system for near field communication devices. Near Field Communication (NFC) [14] is a standard defined in 2002 by NXP Semiconductors. It is based on the RFID technology mixed with the infrastructure of a smartCard. A NFC ready smart-phone can read data from RFID tags and you can bind them to applications on your mobile phone. For example, it is possible to touch a RFID tag with your mobile phone and it will send automatically a short message or initiate a call. Additionally it can communicate with other mobile phones in its range, usually only a few centimeters around the phone itself. Furthermore it can communicate with NFC ready card readers, to pay with your mobile and maybe get access to restricted areas.

There are just a few mobile phones available which support NFC, for example Nokia 6212 Classic + NFC, Samsung SGH-X700n, Motorola L7 (SLVR) [1], but the supporting devices will increase dramatically in the next years.

A. O2 Wallet

At November 2007 O2 started a huge field test in London with its application called O2 Wallet[15]. In a wallet people do not just have cash, normally they have different cards for using as a bank, identification, discount or credit card as well. A lot of people have a bunch of cards and often the wallet is too small to held all of them. O2 and Nokia introduced with O2 Wallet a system to solve this problem by integrating wallets into mobile phones. Like a real wallet it holds different cards, but virtual ones as applications. To use one card, the proband just selects it in a menu and uses for example an oyster card to pay public transport, a Barclaycard to use mobile payment or get access into a O2-bar reserved for O2 customers. Thus, nearly every action you normally use your wallet for, can be done through O2 Wallet. Almost everybody has already a mobile phone with him and with O2 Wallet it would be possible just to take the mobile phone and leave the wallet at home.

B. DB Touch & Travel

Travelling by train or bus can be very challenging. First you may have to book your seat, find the right ticket, struggle with the selling machine and finally have the suitable change with you. All in all this may take long enough to miss your train. This experience made nearly every traveler at least once. But maybe this will change, because the Deutsche Bahn, Vodafone, T-Mobile, O2, BVG, et. al. introduced a system called Touch

& Travel where everybody with a NFC ready mobile phone will be able to use it for paying his ticket.

It will be possible to pay every train, bus and metro ticket with Touch & Travel³. Therefore in every station and in every bus there is a Touchpoint to activate or deactivate the application on the mobile phone by keeping it to the touchpoint. Every point is a passive RFID chip providing its location to the mobile phone and the processing is done on the phone by a special application. Traveling in public transport with such a system is quite easy. Whenever the user starts a trip by entering a bus, train or metro he activates his mobile phone by touching the point and after finishing his ride he deactivates the application by touching another point. The background system will automatically calculate the distance, provide a valid ticket and will send each month a bill to pay. So it is not necessary to buy a ticket in advance or knowing the different rates. The application also regards combinations of different trips, for example if you use the bus in the morning it will calculate a single ride and if you use it at the afternoon again, it can cumulate the single trips until a day ride is cheaper and change all single rides to a day ticket. As a result, the mobile phone can be considered as the only device necessary to travel in public transport. By connecting to the Internet or a special application it can provide timetable information, book a seat and with Touch & Travel also choose and pay the ticket fare. It is not quite sure when Touch & Travel will be available for everybody on every station and bus. At December 2008 a second pilot test in Berlin, Frankfurt/Main, Hannover, Kiel, Lübeck and Eutin was started. In 2010 there will be another one in Nordrhein-Westfalen.

C. Visa payWave

Visa has already started a new kind of payment in the United States; the contactless paying system payWave [24]. It does not require to insert credit cards into a card reader, instead a single swipe over a special reader initiates the paying process. It is designed to fasten the payment and allows paying smaller amounts without cash. In addition a new credit card as a key ring was introduced.

The system uses a NFC compliant technology, so the communication between a credit card reader and a NFC mobile phone gets real. In collaboration with Nokia, Visa developed an application to pay at every contactless credit card reader by a NFC mobile phone. As a result it is not necessary to take your credit card with you, instead a mobile phone is enough to pay the bills.

VII. CONCLUSION

For sure this was just a short introduction to market-ready ubiquitous systems. As pointed out, there are many different dimensions with countless use cases. But as varying as these systems are, there are also common properties. As seen in the definition of ubiquitous, every system is connected to at least one other, rather to many or even to a distributed network with countless devices.

³Deutsche Bahn Touch&Travel: <http://www.touchandtravel.de>

Furthermore, same technologies for completely different systems are being used. Every system, which is shown in this paper, needs a communication device for Internet access or to communicate with another system. Some of them need a location independent module, like the mobile network. Mentioned systems are for example the emergency call system, BMW ConnectedDrive or the Near Field Communication projects. The others, like SMART-Meter, Virtual Power Station or the Next Generation Hospital, are stationary systems, which can connect over ethernet or a common landline. In addition, another crucial aspect are location based services, for example BMWConnected Drive or Emergency Call. These systems are equip with a GPS-Receiver to determine their global position. Last but not least, there is often an identification necessary. It does not matter what device, object or person has to be recognized by a system, as long as it knows how to match a unique ID to it. In the majority of cases this is done by an RFID tag, respectively an RFID reader.

It is impossible to count the current research projects and many of them are already at a late stage and may be available soon. Thus it is not sensible to draw a clear line between research projects and finally market-ready systems. For sure, the market will be flooded by countless ubiquitous devices in the next ten years. But the all-dominant point is not what will leave the laboratories and become available to the public, rather it is the acceptance by human beings. After all, users make the decision which system they will use and what will be a failure. A common cited example is the automatic gear shift. Introduced in the forties, it is still not completely accepted. There is no question it alleviates driving a car, but many people still like to shift gears manually, especially in Germany. In the United States or Japan this is completely different. Thus there are local discrepancies to compete with by introducing a ubiquitous system as well.

The aim and intention of each system is different, some of them can save lives others are just simplifying our daily lives and help us to do things more efficiently. As a result, some of these systems are in a common interest. For example the emergency system, medical systems or power grids are essential and thus are needed to push on the market. Often this is done through regulations, so the user is not asked whether he wants to use it or not. On the other hand, he can judge if he wants to use a system to simplify his daily life. All in all the next years will be challenging and interesting which systems will be accepted and which will not.

REFERENCES

- [1] Near field communication research lab hagenberg - devices. [Online]. Available: <http://www.nfc-research.at/index.php?id=45>
- [2] "Citizen's summary eu proposal - road accident alert system (ecall)," Aug. 2009. [Online]. Available: http://ec.europa.eu/information_society/activities/esafety/doc/comm_20090821/citizens_sum_en.pdf
- [3] BMW. Bmw assist. [Online]. Available: http://www.bmw.com/com/de/owners/navigation/assistvehicle_1.html
- [4] —. Bmw connected drive. [Online]. Available: <http://www.bmw.com/com/en/insights/technology/connecteddrive/overview.html>
- [5] J. Conti, "The internet of things," *Communications Engineer*, vol. 4, no. 6, pp. 20–25, Dec.-Jan. 2006.
- [6] European Commission, "ecall - saving lives through in-vehicle communication technology," Aug. 2009. [Online]. Available: http://ec.europa.eu/information_society/doc/factsheets/049_eCall_august09_en.pdf

- [7] EurostatEuropäischeKommission, *Verkehr*, ser. Europa in Zahlen. Eurostat, 2009, ch. 11 Verkehr, pp. 395–413, statistik zu Verkehr.
- [8] GesamtverbandderDeutschenVersicherungswirtschaft. (2009, Oct.) Pressemeldunge: Autodiebstahl 2008 nord-süd-gefälle bei.. [Online]. Available: http://www.gdv.de/Presse/Pressearchiv_Linkliste/Pressemeldungen_2009_Uebersichtsseite/inhaltsseite24596.html
- [9] H. Hassenmüller, “Gemeinsam stark,” *Pictures of the Future*, vol. Herbst 2009, pp. 40–42, 2009.
- [10] S. Heyduck, “Alarm per hupe,” *weiter.vorn*, no. 4, p. 50, 2009.
- [11] J. Horst, G. Frey, and U. Leprich, “Auswirkungen von elektroautos auf den kraftwerkspark und die co2-emissionen in deutschland,” Mar. 2009.
- [12] A. Kleinschmidt, “Medizin auf weltniveau,” *Pictures of the Future*, vol. Herbst 2009, pp. 59–60, 2009.
- [13] —, “Schluss mit dem blindflug,” *Pictures of the Future*, vol. Herbst 2009, pp. 36–37, 2009.
- [14] Nokia. Nfc-technik (near field communication). [Online]. Available: <http://www.nokia.de/service-und-software/technologie/nfc>
- [15] —. (2007, Nov.) O2 announces o2 wallet with transport for london,... Press Release. [Online]. Available: <http://www.nokia.com/press/press-releases/showpressrelease?newsid=1171383>
- [16] H. Nordwig, “Kein arztbrief mehr per post,” *Fraunhofer Magazin*, no. 2, pp. 28–29, 2008.
- [17] OSW-Medien. Bmw connected drive - das intelligente auto. [Online]. Available: <http://www.auto-entertainment.de/Navigation-News.58+M5c8840f7b9f.0.html>
- [18] S. Schultz, “Allianz mit lichtblick: Vw strickt das volksstrom-netz,” *Spiegel Online*, Sep. 2009. [Online]. Available: <http://www.spiegel.de/wirtschaft/unternehmen/0,1518,647858,00.html>
- [19] Siemens. (2007, Jan.) Lückenlose überwachung von blutspenden mit rfid. [Online]. Available: <http://tiny.cc/1aJk7>
- [20] —, “Info per video vor dem ausrücken der feuerwehr,” Oct. 2009, reference Number: IN 2009.10.2d. [Online]. Available: http://w1.siemens.com/innovation/de/news_events/innovationnews/innovationnews-meldungen/2009/021_ino_0925_2.htm
- [21] R. Sterbak, “Die vieläugigen,” *Pictures of the Future*, vol. Herbst 2009, pp. 92–94, 2009.
- [22] Tagesschau.de, “Blackout-opfer reichen sammelklage ein,” Aug. 2003. [Online]. Available: <http://www.tagesschau.de/wirtschaft/meldung319660.html>
- [23] A. Vaughan, “Smart energy meters in every uk home by 2020,” May 2009. [Online]. Available: <http://www.guardian.co.uk/environment/2009/may/11/smart-meters-energy-efficiency>
- [24] VISA, “Visa pay wave.” [Online]. Available: <http://usa.visa.com/personal/cards/paywave/index.html>
- [25] M. Weiser, “The computer for the 21st century,” *Scientific American*, 02/1991 1991. [Online]. Available: <http://www.ubiq.com/hypertext/weiser/SciAmDraft3.html>

Pervasive Gaming

Rainer Möhle – {firstname.lastname}@uni-ulm.de

Abstract—This paper overviews the large topic of ubiquitous or rather pervasive gaming. It will discuss the differences between ubiquitous and pervasive gaming, their importance in today's life and the hardware and social requirements (divided into four dimensions) for playing these types of games. Furthermore, it will provide an overview on pervasive gaming by introducing and describing existing projects for each game genre. Each project realises certain aspects and faces several problems based on hardware or social issues. This paper will give a review on the advance in solving these problems and where further research is required.

I. INTRODUCTION

IN recent years, among purely stationary games for personal computers and game consoles, a new field of games developed – ubiquitous or rather pervasive gaming. The purpose of these games is to transfer classic games into a new interactive environment or to develop new types of games with the help of computer hardware. The importance of this new field grows by the spread of mobile and ubiquitous hardware in everyday life. It provides a complete new opportunity, for example to carry game ideas into natural environments or to explore new ways of interaction with other players.

The importance of pervasive gaming can be estimated by taking a look at the support of the European Commission in 2004. For nearly two years the Commission's IST Programme founded a project called "IPerG" (Integrated Project on Pervasive Gaming) with an amount of six million euros. The involved industry partners raised another four million euros [21].

Beside this project there have been a lot of other projects carried out until today. All of them have one in common: they faced difficulties with hardware, software and the human factor. The goal of this paper is to overview the topic pervasive gaming and to give a review on the most promising projects so far and what the main challenges were they had to face.

Following now is a differentiation between ubiquitous and pervasive gaming.

II. PERSVASIVE GAMING: FUNDAMENTALS

Distinguishing between the terms "ubiquitous" and "pervasive" gaming is not simple. Pervasive gaming is a subcategory of pervasive computing. The word "pervasive" itself means "to go through" or "to spread through". It is often equated by researchers with ubiquitous computing [18]. Because this is not quite correct, it will be accurately described in the next section.

A. Pervasive versus ubiquitous gaming: an approach

To understand the difference we have to go back in time when Mark Weiser was the manager of the Computer Science Lab in the Palo Alto Research Center and shaped the term

"ubiquitous computing". In his probably most cited article "The Computer for the 21st Century" he described in [24] future developments:

"We are therefore trying to conceive a new way of thinking about computers, one that takes into account the human world and allows the computers themselves to vanish into the background."

It becomes clear that the goal of ubiquitous computing is on the one hand a technological aspect by trying to develop small and mobile computers and integrate them into the everyday world. On the other hand the challenge is to take the computer out of his technical and often abstract background and integrate it into everyday life in such a way that the user doesn't recognize it as such.

The first time using the word pervasive in the context of pervasive computing was probably in 1998 by IBM. It was more or less used in a business than in a research or technical context. The idea of pervasive computing was [12]:

"Pervasive computing is about enabling people to gain immediate access to information and services anywhere, anytime, without having to scrounge for a phone jack."

From this follows that at this time pervasive computing was seen to provide an opportunity to access information anywhere at any time. Over the years, as technology advanced, ubiquitous computing as well as pervasive computing has been redefined. Today the differences between ubiquitous gaming and pervasive blur. One approach to differ between them is that ubiquitous gaming has a more technological background and pervasive gaming uses these proceedings but focuses on gameplay and interaction [18]. In my mind, this is the most suitable term to determine the difference.

B. Definition of pervasive gaming

As seen in the previous section, pervasive gaming uses the advances of ubiquitous computing to create new kinds of games. But what exactly is pervasive gaming?

First of all, pervasive games take the game itself into the real world. The real world can be an outdoor area, like streets of a city or just the living room at home for example. Pervasive games use sensors to get information about the surrounding area, the location or player's movements and use them for frequent adaption of the game state. Most of the currently existing games are designed for multiplayer purposes. The player himself is not forced to stick to his personal computer or gaming console in order to play games. Pervasive games connect the virtual world with the real world which creates a new gaming experience [3]. According to [11], this experience can be divided into four dimensions:

- Physical
- Mental/Intellectual
- Social
- Immersion

The success of a (pervasive) game depends on how well these dimensions of gaming experience are implemented and how deep the players are mentally involved in the game.

The physical dimension is the kind of feedback the player gets when he touches or uses objects in his gaming environment and accordingly the awareness of the physical presence of other players in the real world.

Mental or intellectual experience includes the challenges and concentration. A game has to be challenging to all types of players. The game has to adapt itself to the players' skill and has to be adjustable enough that the player can create his own path in order to complete his tasks. Also the player's concentration has to be kept alive in the course of the game. Therefore it has to provide incitations that are worth attending and demanding enough. It should also offer a helping hand when the player tries to map the ingame tasks into his surrounding environment.

The social dimension is of particular importance. It comprises social interaction and communication with others players. The game should include riddles, events or other corporate situation that encourages interaction and communication between players inside or outside the game.

Last but not least the fourth and probably the most demanding dimension is the immersion into the game. The player should be involved enough to lose his sense of time and the awareness that he is just playing a game. It has to make him feel like being a part of the game and losing the focus on the problems and difficulties he has in the everyday world. Besides being involved emotionally into the game, the experience or actions a player faces may not violate social norms [14].

These four dimensions are fundamental for developing a pervasive game. Over the years many genres of pervasive games evolved. In [16], there is a differentiation into five genres:

- Smart Toys
- Affective Gaming
- Augmented Tabletop Games
- Location-Aware Games
- Augmented Reality Games

The following section introduces for each genre the most significant projects so far.

III. EXISTING PROJECTS

Besides the already mentioned "IPerG" project, which brought out numerous pervasive games and software, there has been a large amount of projects by universities, commercial providers and private groups. For a better understanding, this section is divided into the genres of pervasive games. Each genre will be introduced with one or two outstanding respectively recent projects so far.

A. Smart Toys

Smart Toys are normal toys which are upgraded with further technical hardware. Most common smart toys are equipped with sensors and a computer logic which interprets the detected surrounding environment.

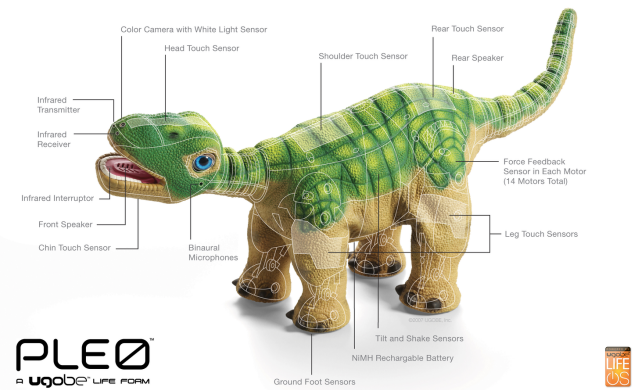


Figure 1. Pleos integrated sensors

1) *Pleo - the robotic dinosaur*: Pleo is a smart toy in the form of a baby dinosaur developed by a company called Ugobe. It has plenty of sensors, microprocessors, and motors integrated. With the help of these sensors and its operation system Pleo acts like a one-week-old dinosaur. Most of the behaviours are triggered by events captured by motion detection, infrared, tilt, camera and microphone sensors (see Figure 1).

Pleo adapts its movements and behaviour to his environment [19]. Recently, a new generation of Pleo has been released with improvements like longer lasting paint, improved mechanics, skin and battery charger. Despite some enhancements there are still unsolved issues. The battery life is way too short. After about four hours of charging time, Pleo works for approximately 30-45 minutes at a stretch. There are three stages of evolvement to pass. From the stage of birth to hatchling and finally to juvenile, Pleo enhances its movement and behaviour. Though it seems that there is a possible extension of Pleos knowledge, it is not implemented until today. Actually there are software updates for changing his behaviour but no potential of learning something new [6].

B. Affective Gaming

This genre offers a young but very interesting research field. In contrast to location-aware games using the location of a player, affective gaming concentrates on the players emotional state and integrates this data into the game flow. Rosalinda W. Picard described affective computing in her book [20] which carries the same name:

"When will we have affective computing?

This book has defined key issues in "affective computing", computing that relates to, arises from, or deliberately influences emotions."

Using player emotions will add new possibilities in games. Affective games measure for example the heart rate, the galvanic skin response, brain waves or muscle activity. One possible field of application could be the adaption of skill levels in games. When a game detects frustration it could simplify the game difficulty and raise it if a player gets bored.

1) *SmartBrain Technologies*: SmartBrain Technologies has an exclusive partnership with NASA and received the NCTI Award for the “Brightest Idea in Assistive and Learning Technology”[8] in 2007. The idea is to use a normal racing game in combination with the detection of brain waves via a sensor integrated into a visor. If the brain waves match a specific predefined pattern, the player is able to handle the car and increase its speed. If not, the car control is disabled and it decelerates. Besides visual the player also gets feedback via audio speakers and gamepad vibration (see Figure).

There is a long list of supported Sony Playstation and Microsoft XBOX racing and jumping games for this affective gaming system. This project was developed for people with attention deficit hyperactivity disorder. Instead of boring exercises people with ADD/ADHD can play interactive games for training [23].

C. Augmented Tabletop Games

Despite the astonishing development of computer hardware and games, classic tabletop games are still an integral part of everyday life. The social experience sitting face-to-face with other opponents, feeling their reactions in every step of the game or just engaging in small talk cannot be replaced by a virtual representative like network multiplayer games. So augmented tabletop games take up the idea of the classic variants and extend them by adding the advantage of computer hardware to create new and inspiring game situations.

1) *Poker Surface - Combining a Multi-Touch Table and Mobile Phones in Interactive Card Games*: This project transfers the classic Poker tabletop game into the digital world. Instead of a normal table and poker cards this game uses a multitouch table and mobile phones. Like in normal Poker games the players are seated around the table. The multitouch table arranges a private area for the players’ digital chips and playing cards. Poker chips and playing cards can be dragged and moved (see Figure 2). A chip with high value can also be splitted into chips with lower values. A player can check via knocking on the table. Players can decide whether they play with the table by itself or extend it using mobile phones. Mobile phones add new possibilities to the game play. The phone displays the playing cards and gives the players a feeling of holding those cards in their hands. Also it provides a hidden area for their playing cards. The mobile phones have built-in accelerometers which are able to detect user gestures. Those gestures are used for looking into cards, folding and checking [22].

2) *STARS - Towards the Next Generation of Tabletop Gaming Experiences*: STARS is a platform for augmented tabletop games. It consists of a multitouch table used as playing ground, public vertical display and integrates PDAs. It offers the possibility to place tokens onto a (multitouch) table like in classic tabletop games. These tokens as well as the player’s position are detected by an overhead camera. With the help of this framework it is easy to implement new games [17].



Figure 2. Poker surface with private areas, digital chips and playing cards



Figure 3. Runner with earphones holding a PDA

D. Location-Aware Games

While tabletop games add new pervasive game experience to tabletop games, location-aware games use pervasive technology to examine the surrounding world like for example streets or buildings of a city. Now a player gets really part of a game. Location-aware games use techniques to locate players for example via GPS, GSM or WiFi or others methods and transfers the location into the game continuously [11]. The game isn’t tied anymore to a certain place or rather an object. The real world gets interwoven with the virtual world. This creates completely new possibilities but also challenges for game developers.

1) *Can You See Me Now?: Can You See Me Now?* created by the artist group “Blast Theory” in collaboration with the University of Nottingham, transfers the simple idea of cops and robbers into a mixed virtual/real world scenario. There are two types of players. The runners who move through the streets of a city and online players. Runners use a PDA with WiFi and GPS connection and walkie-talkies (see Figure 3).

The PDA shows their own location and those of the online players. With the help of walkie-talkies they can communicate with others runners for example to tell about

traffic jams or to arrange a corporate chase of online players. The communications between runners are streamed over the Internet and accessible to online players as well as the runners' locations. Up to 15 online players move virtual representations of themselves at a given speed through a generated 3D model of the city. They can see the location of runners and send text messages to other players which are also visible to runners. The trick is not to be caught by runners using background information like traffic jams which exist in the real world but not in the virtual world. Runners can use the communication with other runners to send confusing messages about their plans. If another runner is in range of sight they can irritate online players making wrong plans about their next actions while they communicate offline with the help of hand signals. If a runner is within five meters of an online player, he is caught and a photo of the spot where he was seen is taken. This game has toured around the globe in cities like Sheffield, Cologne or Barcelona [3] [2].

2) *NavBall*: NavBall is a game of speed. There are two teams consisting of eleven players with GPS phones. Every team has a ball and a goal placed on the map of a city. This map is displayed on their mobile phones including their team member's positions. The intention is to score a goal by lining up the team members in the shape of an arrow behind the ball and heading the arrow towards the goal. After scoring a goal, another ball and goal is displayed in another position on their map. The team scoring the most goals within 45 minutes wins [7].

E. Augmented Reality Games

In augmented reality games the real world is extended by virtual objects. Using head-mounted displays, projecting images on real world surfaces or hand-held devices the real world environment is augmented with virtual content. This content could be virtual objects like additional buildings in a city, descriptions on famous buildings or altered buildings for example in TimeWarp [10] which overlays today's buildings with textures of medieval times.

1) *Human Pacman*: a mobile, wide-area entertainment system based on physical, social, and ubiquitous computing: Human Pacman carries the classic Pacman game into augmented reality. In contrast to the classic version, Human Pacman consists of three types of players: Pacman, Ghost and Helper. Pacman and Ghost players wear head-mounted displays which display virtual objects into the real world environment. These displays are connected with a wearable computer which is connected to a central server via WiFi. Pacman's challenge is to collect all cookies in order to win the game while the Ghosts try to catch him. The 2D cookies from the classic game are transformed into virtual objects in the environment, for example on real footpaths (see Figure 4).

In order to increase communication experience another player type is introduced - the Helper. Helpers are connected via desktop computers and have a complete view of the PacWorld. A Helper is assigned to Pacman or Ghost and transmits

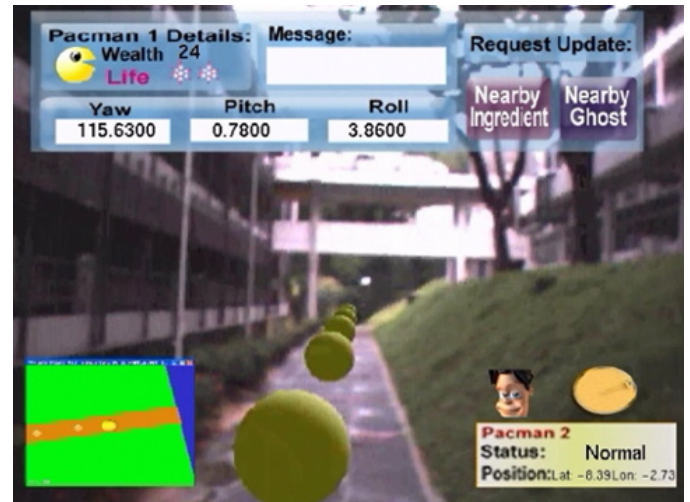


Figure 4. First-person view of a Pacman player

informations about enemies via text messages. Those text messages are displayed on the player's hand-held computer. Like in the classic game there are also boxes lying on the ground. Every box has a Bluetooth device attached. With the help of these boxes, Pacman achieves the possibility to capture a Ghost for a short time. Finally there are two teams: the "Pacman team" consisting of two Pacman and two Helpers and the "Ghost team" consisting of two Ghosts and two Helpers [5].

2) *TimeWarp*: TimeWarp is an augmented reality game which takes place in the city of Cologne. The goal of the game is to walk through the city, solve riddles and finally bring back the "Heinzelmännchen". A legend says that the "Heinzelmännchen" worked for the citizens of Cologne while they were sleeping and one day the "Heinzelmännchen" abruptly disappeared. The player walks through the city with augmented reality equipment and a handheld device. By taking a walking within the city additional medieval textures are laid over existing buildings and streets as well as "Heinzelmännchen" [10].

IV. STATUS QUO

Over the years there has been a lot of interesting and promising projects in the field of pervasive gaming. By means of progress in hardware, projects could be realised which were unthinkable years ago. A closer look into existing pervasive games reveals, that the used hardware has not technically matured yet. The developers meet problems which they try to solve but some still remain unsolved due to limitations of hardware. In the following, different aspects of current pervasive gaming projects are highlighted.

A. Hardware requirements

As seen, there are many interesting and promising projects in the field of pervasive gaming. Now it's time to show the used hardware in a short summary.

Table I
COMPARISON - UTILISATION OF DIMENSIONS IN PERVASIVE GAMES

Game	Physical	Mental	Social	Immersion
Pleo	++	—	—	+
SmartBrain	—	+++	—	++
Poker Surface	++	+	+	++
STARS	++	++	++	+++
CYSMN?	+	++	+	+
NavBall	+	++	++	++
Human Pacman	+	++	+	+
TimeWarp	+	+	—	—

- *Sensors*

Any type of pervasive game requires sensors. There are a lot of different sensors used in previous projects like those detecting motion, touch, audio, video, infrared or brainwave or sensors like GPS or GSM used for locating the player's position. Sensors are one of the main requirements for realising the idea of pervasive gaming. The collected data helps to adapt the game to the current state of the player and his surrounding environment.

- *Classic computers*

There is still a need of desktop computers in many pervasive games. In some projects like "Can You See Me Now?" online players can participate in a game. Most of the projects need a central server who cares about updating the game state and implementing the game logic. A multitouch table or a tabletop game also needs a maybe smaller version of desktop computers to analyze the data from sensors.

- *Mobile computers*

Hardware like hand-held computers or mobile phones are necessary in many projects to display the current player's state or to communicate with other players. In addition there are head-mounted displays which show a virtual world and additional information.

- *Connectivity*

Connectivity is an important aspect if a pervasive game takes place in open space or buildings. It is fundamental to connect mobile computers like hand-helds with those of other players. It is also important to establish status updates between mobile computers and central servers.

B. Aspects solved

Creating a pervasive game is always a span between the theoretical idea on paper and hardware preconditions. As introduced in the first part of this paper, there are four dimensions of pervasive gaming. These dimensions remain part of the detection which progress has been made and – following in the next passage – what kind of issues remain. Pointed out in the earlier section there is a classification into different genres of pervasive gaming. Each game genre tries to include all dimensions but focuses on certain ones. Following now is a reflection on each dimension. In table I you can find a

summary of the utilisation of each dimension in the previously addressed games.

The physical dimension, for example the experience by touching objects and the awareness of other players in the gaming environment are realised in several genres. One example for the awareness of touchable objects are "smart toys" or "tabletop games". "Pleo" is one physical object which provides the ability to interact directly with. There are also other examples for tabletop games like "STARS" which offers the ability to place objects on an tangible table similar to classic tabletop games. In augmented reality games like "Human Pacman" there is the opportunity to grab a real box to influence the course of the game. Location-aware games like "NavBall" or "Can You See Me Now?" take the experience out on the streets. In "Can You See Me Now?" the player interacts online on a desktop computer. The creators made some effort bei placing computers with public access throughout the city to get a more physical feeling. The players in NavBall do physically manipulate their shown position and have a better interaction with physical players.

Challenge and concentration sum up the mental or intellectual experience. Games like "Human Pacman" or "Can You See Me Now?" need elevated concentration because one moment of inattention can be the end of the game. Tabletop games offer more challenge but do not need continuous concentration. Challenges in augmented reality games depend on how good the riddles are.

The social dimension stands for the social interaction and communication between players. In augmented tabletop games the social interaction stays nearly the same as in classic tabletop games. By facing the other players in the real world the player gets a good feeling for their strengths and weaknesses. Games like "NavBall", "Can You See Me Now?" or "Human Pacman" need the interaction with other players in order to win the game. May it be via text messages or direct interaction with other players, they encourage social interaction. Text messages may be an appropriate solution for some games but weaken social communication.

The immersion into the game depends on how the players get emotionally involved into the game. A good game has to hide discrepancies and errors. In "TimeWarp" the user study showed that many users concentrated on the errors and problems [10]. The same was in "Can You See Me Now?". At one time some of the users thought that the runners would turn off their GPS to get an advantage [2].

C. Open issues and challenges

Every game genre faces hardware and software issues in which further research has to be done. Software issues can be solved by putting more effort on programming, but current hardware reaches its limits.

- *Battery Life*

In contrast to Moore's Law which says that processor speed and memory capacity doubles each 18 month, battery life has no exponential growth [15]. This leads to a common problem the hardware industry has to face. Battery life cannot keep up with the developments made

in processor speed and memory capacities. For example smart toys like “Pleo” struggle with this issue. After four hours of charging the battery lasts about 40 minutes. One common solution are battery saving hardware and software components but that’s no final solution to the battery life problem itself.

- *Localisation*

Pervasive games mostly need to locate the position of players or objects. Some games only require rough position data, others need the exact location of a player. Augmented reality and location-aware genres can be listed as examples for the need of exact data. Games like “TimeWarp” and “Can You See Me Now?” expose the limitations of current hardware. Players struggled in both games with uncertainties [10] [2]. In “Can You See Me Now?” the greatest problem was the localisation with GPS which is a common outside tracking method. Though the action took place outside there still were connection issues. The main reason was the location where the game took place - the city. In contrast to open space the streets of a city bury many black spots where no GPS signal can be received or cause errors [1] [4]. Black spots can be produced by narrow streets, known as the “urban canyoning effect” [4]. As already mentioned this caused uncertainties in the game flow. Players were disappointed because a runner appeared out of nothing and caught them all of a sudden or they thought that the runners were manipulation their GPS receiver in order to distract them [2]. Similar problems existed with the WiFi connection. One approach to solve this problem is to combine different tracking methods besides GPS in order to receive a reliable position [2]. Another approach can be the integration of errors like incomplete WiFi coverage as a game component [16]. Players can hide in those uncovered regions or sneak up on their enemies, for example.

- *Communication*

One part includes the known connection problems discussed in the previous localisation section. Communication between players is an integral part of the most pervasive games genres. Outdoor games, which also use WiFi as a communication channel, face these problems. Besides WiFi, there are additional possibilities like GPRS or UMTS if global communication is needed. If players are in short range they can use for example WiFi, Bluetooth or RFID [4]. Another problem is the joining and leaving of players during the game. Due to battery or connection problems a player could leave the game. In “Can You See Me Now?” a player would propose that a runner probably thought that a runner left the game or took a coffee break because his signal was lost [2]. The game server should handle these issues.

- *Devices*

Devices like PDAs, head-mounted displays or mobile phones are necessary for example for location-aware and augmented reality games. There is a need of standard for all existing PDAs if a game should take place outside of strong orchestrated field tests. Also head-mounted displays

are not always the best choice for augmented reality games even if they are partially replaced by PDAs. Many people may struggle with the imagination of walking through a city with such unusual hardware. In this field there has to be put more effort on research alternative types of devices. Also hardware like head-mounted displays or multitouch tables are still too expensive for potential private gamers.

V. CONCLUSION

Pervasive gaming made a lot of progress in the last years. Though developers and researchers could implement important aspects of pervasive gaming or rather pervasive computing, there is still a lack of further research. Despite the fact that a lot of interesting and promising projects have been developed there are less pervasive games who made it to market. Two well-known representatives for location-aware games are GPS Mission [9] and Geocaching [13]. Most of the games developed by universities face the problem that whether the hardware is too expensive nor the games are developed for urban scenarios. Except smart toys or augmented tabletop games which mostly are bound to a specific type of hardware, the other genres meet the same problem. Especially for location-aware games like “Can You See Me Now?” or augmented reality games like “Human Pacman” there is a high initial effort on integrating the virtual world into the real world. Afterwards one specific area, like the one was developed in, is integrated but potential buyers cannot adapt it to their own game area they want to use.

The next issue is that there are still uncertainties with methods like WiFi or GPS in cities. In my mind, putting pervasive games on the market from today’s point of view seems nearly impossible. Efforts are made but there is still a lack of existing hardware in private households. Pervasive gaming is a very attractive field but still needs further research. It will take years until pervasive games are ready for mass-market.

REFERENCES

- [1] S. Benford, R. Anastasi, M. Flinham, C. Greenhalgh, N. Tandavanitj, M. Adams, and J. Row-Farr, “Coping with uncertainty in a location-based game,” *Pervasive Computing, IEEE*, vol. 2, no. 3, pp. 34–41, 2003.
- [2] S. Benford, A. Crabtree, M. Flinham, A. Drozd, R. Anastasi, M. Paxton, N. Tandavanitj, M. Adams, and J. Row-Farr, “Can you see me now?” *ACM Trans. Comput.-Hum. Interact.*, vol. 13, no. 1, pp. 100–133, 2006.
- [3] S. Benford, C. Magerkurth, and P. Ljungstrand, “Bridging the physical and digital in pervasive gaming,” *Commun. ACM*, vol. 48, no. 3, pp. 54–57, 2005.
- [4] W. Broll, J. Ohlenburg, I. Lindt, I. Herbst, and A. Braun, “Meeting technology challenges of pervasive augmented reality games,” in *Proceedings of 5th ACM SIGCOMM workshop on Network and system support for games*. Singapore: ACM, 2006, p. 28. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1230040.1230097&coll=ACM&dl=ACM&CFID=60270734&CFTOKEN=36140616>
- [5] A. D. Cheok, K. H. Goh, W. Liu, F. Farbiz, S. W. Fong, S. L. Teo, Y. Li, and X. Yang, “Human pacman: a mobile, wide-area entertainment system based on physical, social, and ubiquitous computing,” *Personal Ubiquitous Comput.*, vol. 8, no. 2, pp. 71–81, 2004.
- [6] I. L. Corporation, “PleoWorld: the home of pleo, the robotic baby dinosaur from innvo labs,” 2009. [Online]. Available: <http://www.pleoworld.com/Home.aspx>
- [7] T. S. C. Creators, “NavBall – next level mobile GPS game,” 2008. [Online]. Available: <http://navball.wordpress.com/>
- [8] N. C. for Technology Innovation, “Brightest ideas in assistive and learning technology - winners for 2007,” 2007.

- [9] O. GmbH, "GPS mission - mobile GPS game," 2009. [Online]. Available: <http://gpsmission.com/>
- [10] I. Herbst, A. Braun, R. McCall, and W. Broll, "TimeWarp: interactive time travel with a mobile mixed reality game," in *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*. Amsterdam, The Netherlands: ACM, 2008, pp. 235–244.
- [11] S. Hinske, M. Lampe, C. Magerkurth, and C. Röcker, "Classifying pervasive games: On pervasive computing and mixed reality," IN *PERVASIVE GAMES, C. MAGERKURTH AND*, vol. 1, 2007. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.66.6807>
- [12] IBM, "How pervasive computing will personalize E-Business," 2004. [Online]. Available: http://domino.watson.ibm.com/comm/wwwr_thinkresearch.nsf/pages/bergman398.html
- [13] G. Inc., "Geocaching - the official global GPS cache hunt site," 2009. [Online]. Available: <http://www.geocaching.com/>
- [14] K. Jegers, "Pervasive game flow: understanding player enjoyment in pervasive gaming," *Computers in Entertainment (CIE)*, vol. 5, no. 1, p. 9, 2007.
- [15] F. Kargl, "Mobile and ubiquitous computing," 2009. [Online]. Available: http://www.uni-ulm.de/fileadmin/website_uni_ulm/iui.inst.100/lehre/mobile_ubiquitous_computing/ss09/muc-2-devices.pdf
- [16] C. Magerkurth, A. D. Cheok, R. L. Mandryk, and T. Nilsen, "Pervasive games: bringing computer entertainment back to the real world," *Computers in Entertainment (CIE)*, vol. 3, no. 3, pp. 4–4, 2005. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1077246.1077257&coll=Portal&dl=ACM&CFID=67562406&CFTOKEN=17824029>
- [17] C. Magerkurth, M. Memisoglu, T. Engelke, and N. Streitz, "Towards the next generation of tabletop gaming experiences," in *Proceedings of Graphics Interface 2004*. London, Ontario, Canada: Canadian Human-Computer Communications Society, 2004, pp. 73–80. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1006058.1006068&coll=Portal&dl=ACM&CFID=67562575&CFTOKEN=95893050>
- [18] E. Nieuwoudorp, "The pervasive discourse: an analysis," *Comput. Entertain.*, vol. 5, no. 2, p. 13, 2007. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=1279540.1279553>
- [19] T. Perry, "Tools & toys: Pleo, the Poop-Free pet," *Spectrum, IEEE*, vol. 45, no. 4, p. 24, 2008.
- [20] R. W. Picard, *Affective Computing*. Massachusetts Institute of Technology: The MIT Press, 2000.
- [21] C. Schulzki-Haddouti, "EU-Kommission foerdert pervasive gaming," 2004. [Online]. Available: <http://www.heise.de/newsticker/meldung/EU-Kommission-foerdert-Pervasive-Gaming-105815.html>
- [22] A. S. Shirazi, T. Döring, P. Parvahan, B. Ahrens, and A. Schmidt, "Poker surface: combining a multi-touch table and mobile phones in interactive card games," in *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*. Bonn, Germany: ACM, 2009, pp. 1–2. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1613858.1613945&coll=Portal&dl=ACM&CFID=68913948&CFTOKEN=77715728>
- [23] S. Technologies, "SmartBrain technologies," 2008. [Online]. Available: <http://www.smartbraintech.com/>
- [24] M. Weiser, "The computer for the 21st century," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 3, no. 3, pp. 3–11, 1999. [Online]. Available: <http://portal.acm.org/citation.cfm?id=329124.329126&coll=Portal&dl=ACM&CFID=68577703&CFTOKEN=46729797>

Alternate Reality Games

Ferdinand Deger – {firstname.lastname}@uni-ulm.de

Abstract—Alternate Reality Games (ARGs) are a recent form of collective gameplay, where many thousands of players try to solve a fictional problem or mystery by using real-world media, like email or phone calls. The non-existence of a gameplay interface and the absence of fixed rules make ARGs look like not being a game. The fictional universe of the game is created by a small design team, the puppet masters. The team has to assure a compelling story and a fluent gameplay. The players keep this fictional world alive by a strict separation between the in-game fiction and the real world and a collective suspension of disbelief. This paper presents an overview over Alternate Reality Games. Various examples will clarify different aspects of the game genre ARG.

I. INTRODUCTION

ALTERNATE Reality Games (ARG) are a new form of interactive gaming, that was established in the last couple of years. The easiest way to introduce ARGs is to present an example. It will show basic concepts and important components, that are explained in this paper in-depth.

Example: The Lost Experience

Imagine you are watching the second season of the TV show *Lost* and a notice a spot for the Hanso Foundation during the commercial break. As a frequent *Lost* viewer, you probably know this company, as it is a part of the fictional plot of this mystery show. The spot says: "Since the dawn of time, man has been curious, imagining all that is possible. The Hanso Foundation: reaching out to a better tomorrow. Discover the experience for yourself." and presents you a toll-free phone number. This took place in 2006 and if you had called this number you would have been in the middle of a highly complex ARG. The game called *Lost Experience* and lasted for several months [19]. Within this game the players followed Rachel Blake, an ex-employee of Hanso Foundation, who tried to uncover the truth about that company. The Players had to solve many puzzles and especially cooperate with each other to unravel the mystery. Hints were planted in a variety of different places. For example: One time the location of a new website was encoded in the html-source of a page, that was already known before, another time a URL was molded into chocolate bars, that were specially manufactured for the *Lost Experience* ARG. An important component of the game was the decryption of secret messages. In one stage of the game hieroglyphics were hidden in 70 different locations — online as well as offline. With them it was possible to decipher a video, that explained unknown parts of *Lost*'s background story. The complexity of The *Lost Experience* made it impossible to play alone. The puzzles were hard to solve. Only few people have the patience and skills to decrypt pictures or to uncover the sense of a strange clue. And additionally all offline clues were

quite limited. For example, figure 1 shows a hieroglyphics on the side of a car. By coincidence some players are at



Figure 1. These hieroglyphics are examples from the *Lost Experience*. They were hidden in many different places and were necessary to decipher a mystical video. The left glyph is code number 29, it was spotted on the side of a car driving past the EW offices in Los Angeles. The right one is code 70, it was hidden as 14 separate pieces in the blog of Rachel Blake (she is the main character in this ARG) [19].

the right place at the right time. But certainly this does not apply to all of them. The remaining ones rely on second hand reports. ARGs are a team sport. All players have to bunch their informations and create a collective intelligence [12]. For this reason the fans created external web sites to discuss their theories and exchange new informations. According to HiRes!, the responsible design agency, thousands of weblogs, forums and hundreds of podcast were created to share the experience of this ARG [5]. Unfortunately most of the original puzzles are no longer available. Therefore some cites of this paper go to fan-sites and wiki-systems, as they happen to be more persistent, than the official resources.

The *Lost Experience* was designed in London and deployed to the participants in the UK, USA and Australia simultaneously. It took place between May and September 2006 and, according to HiRes! [5], it was a great success, with over 30,000 active regular users — only in the UK. The *Lost Experience* gave fans an opportunity to uncover further informations about the *Lost* background story. For the television company this ARG was a marketing instrument, to ensure, that people kept on watching.

A Definition

Jane McGonigal, an alternative reality game designer and one of the leading researchers on the field, defined ARG in a

presentation for the MacArthur Foundation¹. ARGs are [10]:

”an interactive drama played out in online and real spaces, taking place over several weeks or months, in which dozens, hundreds or thousands of players come together online, form collaborative social networks, and work together to solve a mystery or problem ... that would be absolutely impossible to solve alone.”

This definition was distributed on variant slides in that presentation, to point out the key elements of ARGs. First of all a ARG is a drama, that means a story has to be told and this is done in an interactive way. This invokes, that the audience is not just consuming a product, but able to play a role within the drama. Another part of this definition clarifies the number of players and the amount of time. With a few thousands players she actually stated a small number. For example: According to the responsible designers, the successful ARG *The Beast*, which launched 2001 and therewith three years before that presentation, had more than three million people actively participating [1]. The lead designer Elan Lee even declared a million participating players as a critical mass for a successful ARG [7]. But certainly fewer participants also enable fluent gameplay, as the definition indicates.

Due to large player mass, ARGs have to rely on supportive social networks for coordination. Modern organization tools like forums, weblogs and wiki-systems are mandatory, to keep the players up to date over such a long period. Many play just casually and depend on summaries to keep up to date. It is irrelevant for the player if these social networks are hosted by the design agency or if they are independent. The drama itself is obviously not limited to the online space, as we saw by the offline clues in the *Lost Experience* and in McGonigal’s definition. The online discussion forum helps to optimize the workflow and creates a collective intelligence, that facilitates the problem solving process [11]. The Collaboration of players, that work together, combine their skills and resources, is one of most central aspects of ARG. It is impossible to solve an ARG alone [10].

One of the reasons ARGs gain so many player is its accessibility. The game uses real world communication devices, like websites, blogs, SMS messaging. It is not necessary to get used to a special device, like a joystick or game controller. Players only need what they are highly likely have used before: search engines, email programs, phones or postal services [3]. The use of real world media to transport a fictional plot is often characterized as blurring the boundaries between fiction and reality [12]. This blurring spawned the expression *This is not a game* [15]. A game — ARGs are obviously games — is defined as not being a game. What maybe sounds strange in the first place, describes actually a lot about ARGs and the self-image of ARG-players, as we see later.

Origins and boundaries

Alternative Reality Games are a relatively new game. Because the necessary technology to connect the collective has

just been available for the last decade. The first notable ARG was *The Beast* in 2001. It was launched to promote the Steven Spielberg movie ”AI/ Artificial Intelligence” and was a great success, according to the design agency [1]. Later other movies (e.g. *Matrix*), TV shows (e.g. *Alias*, *Lost*) and computer games got their ARG. Most of them were official ARGs and used as marketing tools. But there are also independent, grass root ARGs, like *metacortechs* — based on the movie *Matrix*. There are even ARGs, that are not based on an existing universe, like *MeiGeist*. *MeiGeist*, *The Beast* and *IloveBees*, a promotion for the Xbox game *Halo2*, are presented later to clarify certain aspects of ARGs.

ARG are quite distinct from other forms of online gaming. First of all, players do not have a virtual character as it is common in massively multiplayer online games (MMOG) or role-games (MMORPG). As mentioned before ARG are using real world media and therefor do not require a 3D digital world [7]. Furthermore most ARGs do not have fixed rules — in contrast to MMOG and MMORPG. That does not imply, that ARGs have no rules at all. But they are rather implicit. Players have to find the limitations and boundaries in a process of trial and error [8]. Problems, that are caused by this approach will be treated in section III, *ARGs in Context*

II. DESIGNING AN ARG

How do you design an ARG, when you cannot write down just the rules of the game, as they are not explicitly given? How can you attract the attention of the people and introduce them to a new ARG, without mentioning, that its a new *game*. This section introduces into the design process, and shows how the design team is usually working.

Elan Lee, the chief designer of the *The Beast* and *IloveBees* points out two main features to keep in mind, when designing an ARG: A compelling story and a collaborative gameplay [7]. In order to obtain a fluent gameplay it is important to know about the players. Therefore the designers have to constantly monitor the actions of the players and the status of their discussions, to determine at what stage of the game they are. When one puzzle is solved, the storyline has to move forward to sustain the interest of the players. The necessary updates must be deployed in real-time to avoid gaps in the gameplay. This does not imply that the design team is giving away the control of the timing. For example in the ARG *IloveBees*, the intended update circle was taking a week [7]. This gave players time to reflect and rethink the plot. But when users uncover the secrets faster, the designers must be able to react immediately. A fluent game is urgent for a good collaborative gameplay: ”timing is key” [7], as Elan Lee expressed.

One of the difficulties, that come along with collaboration of thousands of players is the variety of their characteristics. A puzzle that seems easy for one player might be impossible to solve for another. An ARG has to find a balance between the levels of challenges. But a compelling story makes it easier for the design team to deal with different types of players. For many players it is enjoyable to just follow the plot of an ARG, rather than actively participating. Sebastian and Kinzie quoted an ARG player, who just kept reading *The Beast* [14]:

¹A private foundation, with a program for public interest media

"The beast was really REALLY well-written. One time when I 'quit', I told myself I'd stop playing it but just keep READING it. And it was good enough just to read."

The game designer Brooke Thomson calls casual players *readers* [4]. These players browse the various websites, both in-game and out-of-game, following up on the story narrative and reading what other players are doing and saying. And they are by far the biggest group within the players [4]. Readers are not unimportant, because they increase the scope of the ARG and they maybe change to active players. This can happen for example quite casually by starting to comment results of other players and then to participate more and more. ARGs need both components, they need active players, to bring forward the plot and they need *readers*. ARGs are collaborative games as well as spectator sports [7].

Design team

The design team of an ARG is called *PuppetMaster* (PM) by most ARG fans. For them, the puppet-master lives behind a curtain and manipulates the puppets. The players cannot see behind this curtain and therefore do not know anything about the PM [17]. In reality there is a small group responsible for an on-going ARG. For *IloveBees* the design team had 30 members. *IloveBees* was an ARG in 2004 to promote the Xbox-game Halo2 and it was designed by a group around around Elan Lee, also responsible for the ARG *The Beast*. Due to the experience of that team it is interesting to look at their structures and procedures, to learn something about the workflow behind the curtain of an ARG. The three main responsibilities behind *IloveBees* were [7]:

1) *Storyteller*: The storytellers are responsible for a compelling story. Therefore they have to get their work done, before the game is launched. In *IloveBees* the ARG plays in the Halo2 universe. The storytellers were in charge for the consistency with the Xbox-game. As *IloveBees* was launched to promote the game, Halo2 was still under development, when the *IloveBees* plot was written. The storytellers were in close contact with the Halo2-Developer team.

2) *Community lead*: As the number of players is hard to predict, monitoring is a key element. The community lead (in the case of *IloveBees*, it was Jane McGonigal) has to check every available information. This includes, beyond the official sites, private forums and weblogs. Most of the communication between the players takes place in forums, that are outside the control of the design team. There are examples of other ARGs, where the team even monitored chat room communications outside the official websites, to to gather information on the current theories of the ARG-players [12]. The monitoring process is most important. There is usually no beta-testing on the target group before an ARG starts. Everything has to work properly at the first attempt. In *IloveBees* the community lead wrote daily reports. Based on this reports the design team decided, how to update the story.

3) *Technology support and sound effects*: This group managed technical issues, like web page design and web security. Because *IloveBees* depended on audio-files from the Halo2

universe, the necessary audio files were also produced in this group. However Technology support is no key position of an ARG. It is not necessary to create all puzzles within the team. This can be easily outsourced to other design agencies [7]. Note that the whole design team consisted only of 30 members. This is quite small in comparison to other online games, like MMOGs and especially very small compared to the numbers of players, that are cooperating within the game. On the one hand, this implies that the design team cannot take care of any single player. It will stay an illusion, that the players can directly communicate with every main character of the ARG. On the other hand a small design team keeps the budget low and makes it possible for an ARG to be free of charge, as marketing budget is sufficient to cover the costs.

Launching an ARG



Figure 2. The rabbit hole of *IloveBees* was a bottle of honey from Margaret's Honey in San Francisco. Suspended in the Honey was a letter spelling *I - L - O - V - E - B - E - E - S*. This honey was sent to approximately 20 persons, who are active in the digital gaming industry[7].

ARGs start with a so called *rabbit hole*, a starting hint to attract attention and recruit the first players. There are unlimited possibilities for a successful rabbit hole. For instance, the commercial for the Hanso Foundation was the rabbit hole of the *Lost Experience*. Two more examples: In *The Beast*, hints were planted in the trailers for the movie Artificial Intelligence, as it was designed as a promotion for that movie. To ensure sufficient success, more than one rabbit hole was planted independently. First there was a name hidden among the regular credits for the movie - "Jeanine Salla". When people started to search for this name on the internet, they were directed to the entrance page of the ARG. To make it easier to find this name, the team placed a phone number in some of the movie trailers. When people called that number and after following some instructions they received an email saying: "Jeanine is the key" and "you've seen her name before", which direct the players to the first rabbit hole again [18]. The rabbit hole is a critical part of a ARG. It is important to have a fast grow in the first weeks, to gain a critical mass. For example, *IloveBees* had exponential growth.

It started with 100 players, then increased to 500, 10,000, 250,000, and eventually three million, within 12 weeks [7]. The rabbit hole of this ARG was sending packages with bottles of real honey to a limited number of people, who were active in digital gaming industry. You can see some of these bottles in figure 2. It directed the first players to the corresponding website <http://www.ilovebees.com/> [7]. In addition they also presented the website into the trailer of Halo2. This section intended to be an eye opener, how rabbit holes can be designed. As it is a highly creative moment, with the approach to attract attention, no general rules can be derived by this examples. Furthermore, it must be possible for new players to join the game and follow the plot at a later stage, too. In a successful ARG, the community offers commonly some tutorials and quickguides[3].

Collaborative Gameplay

The story of an ARG is delivered in several fragments, which players have to assemble to the narrative. To construct an ARG, means to design a variety of possible interactions with the game. In an interview McGonigal discussed in what different way players are addressed in a well designed ARG [16]:

"First, you try to create an extremely diverse range of participatory opportunities. Some are online, text-based. Some are through vocal interaction, like phones or Skype. Some are real-world, face-to-face. Some are puzzles. Some are stunts. Some are problems. Some are literary. You just throw out a really wide range so people feel hailed by something, and something specific."

The players usually assume, that an ARG follows a certain plan and the *man behind the curtain* knows exactly how the story will unfold. But the design team has often fragments of story at the beginning and develops the game while it is being played. It is impossible to make a meaningful beta-test and to predict the exact direction of the game. Elan Lee surprised many players when he admitted, that they had control over how the plot unfolds[7]. O'Hara essentially calls the core players *co-creators* [12], although they are never on the same status as the puppet master.

III. ARG IN CONTEXT

Although ARG are blurring the boundary between fiction and reality, they still retain a certain narrative embedded in a purely fictional context. For example: It would be easy to unmask the Hanso Foundation as an imaginary construct². The users have to accept the content as fictional content, similar to when they are watching a TV program.

Verisimilitude and Suspension of Disbelief

ARG differ from most other games, by there urge towards realism. They are using real world media to transport the plot and removed hereby all fiction identifying frame [3]. For instance,

²A WHOIS Lookup reveals, that the website is owned by Hi-ReS!

a television screen makes it obvious, to distinguish fiction and reality. Everything inside the screen is fiction and everything around is reality. This is different for ARGs. Aesthetically, technologically and phenomenologically speaking, there is no difference between the look, function or accessibility of the in-game sites and normal websites [9]. A visitor that stumbles upon an ARG site by coincidence would not recognize that he is in a game. However when players are actually playing the ARG, they are well aware of the fictional character. Stean Stewart, a science fiction author and lead writer of *The Beast*, laughed at the question, if players would mistake *The Beast* for reality. "The game is set in the year 2142 A.D.," he stated. "There are killer robots and sentient houses. How could anyone be confused?" [8] To separate the outer world from the ARG, players invented terms to distinguish between both. For the players Out-of-game (OOG) is everything, that occurs in the real world, while in-game is the opposite[17]. In order to be able to enjoy an ARG the players have to accept the boundaries and non-written rules of the game. Elan Lee explains [8]:

"Players were never meant to believe the *This is not a game* rhetoric. It was obviously a game. There was nothing we could do about that. What we could do was make it a game with an identity crisis. If I know it's a game, and you know it's a game, but IT doesn't know it's a game, then we've got a conflict."

The belief in in the credibility and verisimilitude of the game or in other words the suspension of disbelief, can be seen as a collaborative activity [12]. The community want to belief in order to enjoy a compelling narrative without disturbance of the real world. The conflict Lee described, that IT is not knowing about the game, is suppressed by the players. In a discussion group of *The Beast* someone expressed this principal [8]:

"I'll say it right out - I think that any use of WHOIS whatsoever detracts from the enjoyment of the game. It's simply akin to reading ahead in a novel."

But even when the players are willing to play by the rules, it is often not clear what information is OOG. Most ARG maintain indeed a discussion list, of sites deemed OOG. But this list is written by fans, who are not able to look behind the curtain. Every inconsistency might be the next hint from the PM — or coincidence and OOG. An example: In the ARG *MeiGeist* a player discovered, through his researches on the internet, that the actress playing the main character had the same name as another fictional figure in the ARG. This was a coincidence, but he posted his result as a potential new clue, nevertheless. By mentioning the real world name of the actress he was mixing the ARG-universe with the outside world. Some players react pretty harsh when they have the feeling that someone presents too much information and thereby destroys their illusion of the imaginary world. The separation of OOG-information and relevant clues can be seen as a *signal to noise ratio*. Hereby minimizing the noise means suppressing unnecessary elements, like information from the real world, which are useless within the ARG. On the other hand maximizing the signal, is equivalent to clarify the information made available by the puppet master [3].

Cultural Context

Many ARG are embedded into a bigger context. It might be harder, or even be impossible to follow the plot of the ARG without knowing other narratives related to the surrounding fictional. For instance: many circumstances in the *Lost Experience* were already known to followers of the TV show *Lost*. The enforcement to connect different story-lines is no specialty of the ARG genre, but rather a recent style in popular culture. As Jenkins describes [6]:

”The old Hollywood system depended on redundancy to ensure viewers could follow the plot at all times, even if they were distracted or went out to the lobby for a popcorn refill during a critical scene. The new Hollywood demands that we keep our eyes on the road at all times, and that we do research before we arrive at the theater.”

This can be transferred to ARG, as there are many cross-links, between the game and e.g. the TV-show. In an ARG for the TV series *Alias* it was even necessary to watch the show, to be able to solve a new clue [13]. In this context it is meaningful to reconsider, that most ARG are marketing instruments. Therefore, recent ARG often provide no independent interconnection of different narratives, but a *hierarchy of meanings* among these texts, with a clearly identifiable *ur-text* [13]. As Örnebring says [13]:

”The [...] ARGs might well refer to characters and events from the fictional universe of the TV series, but flows in the opposite direction (i.e. the TV series taking up narrative threads or using characters from other media texts based on the same fictional universe) are much rarer.”

You can follow the TV-show *Lost* without playing the *Lost Experience*, as Rachel Blake is not participating in the TV show and *Lost Experience* has not spoiled any mysteries of the TV series.

The *ur-text* has to fulfill a certain characteristic, to give space for additional narratives — among others ARGs. It needs syntagmatic gaps and areas for narrative exploration outside the main texts [13]. As mystery shows usually have these gaps in the narrative, and space for further story-lines. Therefore they are quite suitable for ARGs. The wide majority of ARG uses cultural context with a known *ur-text*, however this is not mandatory, as *MeiGeist* is not based on any other story.

IV. COOPERATION

The cooperation of players is essential, because a large number of players can form a collective intelligence together [3]. But there are many different forms of participation or engagement within the ARG. Someone just shoots an offline photo and posts it on a web forum. Another person might be a good clue solver that is capable of decoding complex hieroglyphics, but does not understand their meanings. And yet another person has a new speculation on how this piece might fit into the overall narrative. Players can contribute if they have the time, skill or inclination but were not penalized in terms of progression of the story if they fell short on one of these dimensions [12]. A contribution does not always means

a big breakthrough. Even small commitments might help the community. For example: A simple transcript of content, e.g. a video to text, may simplify the discussion about the message in a web forum. It is not surprising however, that cooperation between so many different personalities is not an idealistic world. Players do not only form a collective, by bringing together their skills, but also have tensions and competition among each other [12].

Types of Players

Players can be classified by their different level of participation. For example: The design agency 42 Entertainments classifies their players in: casual, active and enthusiastic [3]. As we have seen casual players previously, we focus now on the active players. According to Dena [3] these, more frequent players can be distinguished into several types. She is referring to them in the style of other digital games as *hardcore gamers*:

Puzzle Player: They are very good in solving cryptographic puzzles, like: morse code, anagrams, braille and many more. Some of them are not interested in the story and only visit related websites, to solve the puzzles. They are usually persons with elaborate analytic thinking capabilities.

Story Players: They spend much time on rethinking the story and speculate about where it might lead to in the future. They are sharing their ideas and make new connections between clues and theories other players posted.

Real World Players: Real World Players are researching locations that are offline. For example: In *IloveBees* players went to certain GPS positions on specific days to unlock important game information. Some game designers think in these categories. The game designer Brooke Thomson connected certain characteristics [4] to types of players. In his experience story players are the most passionate once, that hold onto that passion long after the experience has concluded. He has a slightly different characterization with additional types of players [4], like *community support* (players, that like to communicate and introduce new players), *Character Interactor* (players, that like to interact with fictional characters). There is no consistent classification in literature. These characterizations are tendencies and no strict separation. Players might be in various groups. Many of them have to adapt their involvement during the long period of an ARG, depending on the spare time they have. So they might switch between the different types of players, like being passive players for a while.

Between competition and cooperation

Practical reasons frequently forces the design team to send certain messages, like SMS, phone calls etc. only to a subset of players. For example the bottle of honey we have seen in figure 2 was only sent to a limited number of players. The general idea behind this strategy is, that these persons work like multipliers and share their information with the community. Normally all players do share their informations — there is no report about an unique lost clue, that ruined an ARG. However, when players post about the clues they have gained, there is clearly a sense of disappointment from other frequent players

who have not received anything. An example of such a reaction in *MeiGeist* can be found in the paper of O'Hara [12]:

"I have not received anything, no poster, no letter nor email. And I have done a lot to support this game. I signed up back in December and spread the word to other players and brought them to this game."

"I haven't had a poster either!"

"I think people who received them should post whereabouts they live and we'll see if there's any circles or something. We know they got one in Somerset too."

It is interesting to see a simple disappointment in the first two reactions, while the last tries to find a hidden clue, in the location of the posters. O'Hara reported another example of tension, as there was an ebay auction in *MeiGeist*. Six dolls were auctioned off, that obviously were representing new clues. Many players were curious about these dolls and wanted to observe them on their own, rather than reading a summery in a forum. But out-bidding another player is not really supporting the community. A discussion started about the goals of the group. One player stated there [12]:

"We should not turn this into a bidding war between us players. I personally will not be bidding as I live in the States and want the clues to get to us players asap. We already have to wait almost 5 days before they ship."

Besides a general advice this player makes a personal abandonment, because the shipping (from the UK) would obviously last longer than for a successful bidder within the UK. The player realized that a personal success would lead to a disadvantage of the group, as they would have to wait longer to receive the new clue. During that discussion some players, accused the puppet master of pitting the players against each others [12]. It is in the nature of an auction to bring out the competition, rather than the cooperation, however, it was not the intention of the design team to introduce tension.

Live Event

Some ARGs have special live events, where the players get the opportunity to interact directly with the characters of an ARG. As most alternate reality games happen to be international games, only a small subset of players can participate at the live event. In *MeiGeist* out of several thousand active players only eight players eventually turned up to participate [12]. In contrast to the ebay auction, a live event forms a new level of cooperation and communication, in stead of tension and competition. The players, who went to this event made a commitment to the group to keep records as accurate as possible. They even offered direct involvement [12]:

"Just an idea (I've played way too many adventure games and love set-ups like this) - but does anyone who is NOT going to be there, want to volunteer themselves as an on- call researcher if we need to look something up and can't actually find a computer or the time to use our PDA's?"

After the game the live-eventers posted their experience on YouTube to share their impressions. A Live Event demonstrates

the big advantages of a collaborative gameplay and is therefore a key element of many ARG [12].

V. TYPES OF ARG

From a commercial point of view ARGs can be separated into different groups. There are few ARGs, that charge their players. A rare example is *Perplex City*, which is selling puzzle-cards, that are necessary to solve the clues. But these are exceptions rather than a proper category. The wide majority and the most successful ARGs are free of charge. These games can be divided into two sections. There are marketing instruments and fan-fiction.

Marketing

In this paper, many ARGs, that have been named are examples of successful marketing instruments. Often players prefer this kind of indirect advertisement, as these games are exciting and provide them with more details about a certain universe. The content industry is combining different forms of media, in the style of the *new hollywood*, to animate the consumers explore the world of related products [17]. Therefor using ARGs as marketing can be often seen as a win-win situation. But there are also some critical points about this form of marketing. First of all the players have to give away lots of personal data in order to fully contribute to the game. How can they receive a real world clue without giving away their home address? How can they receive phone calls, without giving away their phone numbers? How can they receive emails, without handing out their valid email address? The players allow marketing companies to monitor their actions and behaviors — and this is may not be limited to in-game sites, as we have seen. The *this is not a game*-culture of ARGs, includes marketing, that is not declared as marketing. Players stumble upon an oddity and following hints without knowing where the narrative might end. For example, there has been an ARG *Last call poker*, where many people thought it was the marketing effort of an online gambling site [13], instead it turned out to be the promotion for "gun", a wild-west themed computer game. It might happen that ARG players support the campaign of companies they are not supporting, without knowing about it. Some companies are seeking a more serious appearance and are announcing their ARGs officially. For example, the television station ABC made a press release, to announce the *Lost Experience*, without saying any detail about the game [2].

Grass Roots ARG

For many popular science-fiction or conspiracy movies, like *Martix* or *Minority Report*, fan based ARGs were created. These are independent from the design team of the original story and, therefor often less connected to the direct story. The makers usually don't have any insider informations. One of these so called grass roots ARGs is *Omnifarm*, that plays in the universe of the TV show *Alias*. It was constructed around the 'Rambaldi mythology', which is part of the fictional universe, but rather as a background story in the TV show. There is another reason not to touch the main plot of TV show. The

characters and plotlines are generally protected by intellectual property law [13]. It depends on the content owners, to what point they are tolerating non profit fan fiction. Although these would have a benefit by grass root ARGs and free promotion by loyal fans, it is not yet clear how tolerant they are eventually going to be. The conflict between user generated content and copyright is a more general issue and out of the scope of this paper. But it will certainly affect future ARGs and the amount of independent productions.

VI. CONCLUSION

This paper provided an overview on Alternative Reality Games and pointed out different views about this recent game genre. For the players, an ARG is the opportunity to participate within the plot of a mystic story. Among many thousands of other players they can benefit from their collective intelligence and solve problems, they would not be able to solve alone. They are living in an imaginary world, although they are using everyday's resources like email services. The paper pointed out the benefits and problems that come along with this sort of cooperation and described the process of creating an alternate reality in a fictional world as the suspension of disbelief. The design team of an ARG has to assure a compelling story and a collaborative gameplay. One opportunity to gain a background for such a story is to use an existing universes, e.g. a TV show and to connect the ARG with it. This kind of connection is especially interesting for marketing purposes, but also independent production benefit from it. It will be interesting to see, how future promotional ARG will be designed and how ARGs can still be *not a game*, when more companies officially announcing their ARGs.

Another recent evolution of ARGs are Serious ARGs, who are trying to solve real world problems in stead of fictional ones. *World Without Oil*, was in 2007 the attempt to imagine a world without fossil energy and to solve future problems in advance. The performance of the collective intelligence, that an ARG provide is quite potent. But it is questionable if real world problems can be translated correctly for the game. It is up to further research if these Serious ARGs are capable of providing suitable solutions or just staying games.

REFERENCES

- [1] 42entertainment, "The beast | project page of the agency 42entertainment." [Online]. Available: <http://www.42entertainment.com/beast.html>
- [2] ABC-news "Lost Game Lets Fans Hunt for Clues" [Online]. Available: <http://abcnews.go.com/Entertainment/story?id=1881142>
- [3] C. Dena, "Emerging participatory culture practices: Player-Created tiers in alternate reality games," *Convergence*, vol. 14, no. 1, pp. 41–57, feb 2008. [Online]. Available: <http://con.sagepub.com/cgi/content/abstract/14/1/41>
- [4] B. T. et. al., "Alternate reality games sig/whitepaper." [Online]. Available: http://wiki.igda.org/Alternate_Reality_Games_SIG/Whitepaper/Understanding_your_Audience
- [5] Hi-ReS!, "Lost experience | project page of the responsible design consultancy hi-res! london." [Online]. Available: <http://archive.hi-res.net/thelostexperience/index.html>
- [6] H. Jenkins, *Convergence Culture: Where Old and New Media Collide*. NYU Press, October 2008. [Online]. Available: <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0814742955>
- [7] J. Y. Kim, J. P. Allen, and E. Lee, "Alternate reality gaming," *Commun. ACM*, vol. 51, no. 2, pp. 36–42, 2008. [Online]. Available: http://portal.acm.org/ft_gateway.cfm?id=1314222&type=html&coll=Portal&dl=GUIDE&CFID=60539624&CFTOKEN=97692757
- [8] J. McGonigal, "A real little game: The performance of belief in pervasive play," in *In Proceedings of DiGRA's Level-Up*, 2003.
- [9] —, "'this is not a game': Immersive aesthetics and collective play," in *MelbourneDAC*, Melbourne, 2003. [Online]. Available: <http://www.seanstewart.org/beast/mcgonigal/notagame/paper.pdf>
- [10] —, "Alternate reality gaming: Life imitates arg." in *Annual Board of Directors Meeting of the MacArthur Foundation, San Francisco, California.*, 2004.
- [11] —, "Why i love bees: A case study in collective intelligence gaming," *The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning*, pp. 199–227, 2007. [Online]. Available: <http://www.mitpressjournals.org/doi/abs/10.1162/dmal.9780262693646.199>
- [12] K. O'Hara, H. Grian, and J. Williams, "Participation, collaboration and spectatorship in an alternate reality game," in *OZCHI '08: Proceedings of the 20th Australasian Conference on Computer-Human Interaction*. New York, NY, USA: ACM, 2008, pp. 130–139.
- [13] H. Ornebring, "Alternate reality gaming and convergence culture: The case of alias," *International Journal of Cultural Studies*, vol. 10, no. 4, pp. 445–462, dec 2007. [Online]. Available: <http://ics.sagepub.com/cgi/content/abstract/10/4/445>
- [14] R. Sebastian and M. Kinzie, "Development of social problem-solving networks in the interactive web-based game the beast: A qualitative study," in *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2006*, T. Reeves and S. Yamashita, Eds. Honolulu, Hawaii, USA: AACE, October 2006, pp. 2350–2357. [Online]. Available: <http://www.editlib.org/p/24061>
- [15] D. Szulborski, *This Is Not A Game: A Guide to Alternate Reality Gaming*. New Fiction Publishing, 2005.
- [16] D. Terdiman, "Newsmaker: Drawing on the wii, reality tv." [Online]. Available: http://news.cnet.com/Drawing-on-the-Wii-%2C-reality-TV---page-2/2008-1043_3-6146088-2.html
- [17] B. Thompson, "Alternate reality gaming quick start guide." [Online]. Available: <http://www.mirlandano.com/arg-quickstart.html>
- [18] Various, "The beast (game) | from wikipedia, the free encyclopedia." [Online]. Available: [http://en.wikipedia.org/wiki/The_Beast_\(game\)](http://en.wikipedia.org/wiki/The_Beast_(game))
- [19] —, "Lostpedia - the lost experience clues | wiki-system of lost-fans." [Online]. Available: http://lostpedia.wikia.com/wiki/The_Lost_Experience_clues

Bio-inspired Networking

Sebastian Schimmel – {firstname.lastname}@uni-ulm.de

Abstract—This article is a survey about two works which refer to the topics “bio-inspired networking” and “self-organization”. The first article delineates rule-based sensor network. The basic idea is to transfer the knowledge of cellular and intracellular signaling into rule-based logic devices to work in self-organization in sensor and actor networks. The second work is based on the firefly synchronization which were observed in South-East Asia. This expertise is explained by bioluminescence. Previous works developed different frameworks for this synchronization scheme. The transmission delays and other necessary time periods are described specifically.

I. INTRODUCTION

Nature has a vast repertoire in relation to communication among each other. Consequently there are a many signaling ways to transfer information. When these information are received by another animal or even a cell, they have to decode the message. After decoding, the specific entity knows how it has to deal with the information and treats the instructions. The signaling and information transfer could happen through light sequences, neuronal stimulations, scents, sounds or through chemical constructions as it is handled in cells. There were a lot of intentions to investigate the biological behaviour. Now the attainments of these researches start to get combined with modern techniques to deskill everyday life.

A part of this field of research is the integration of biological communication systems into network communication. That is because the nowadays networks grow irresistible and need a new technique to control the information transfer. Also special networks like sensor and actor networks become more and more ubiquitous. Because of change, self-organization deployed for networking protocols is necessary. Self-organization means that entities organize themselves. This relates to a distributed behaviour without any external or central control unit.

To facilitate these properties there exist several possibilities. In this survey two works are depicted. At first the self-organization problem is solved by rule-based sensor networks in SANETs. In these networks scalability and energy efficiency are the most challenging topics. This solution is based on the signaling pathway of inter- and intracellular communication. The other option to guarantee the self-organization and network efficiency is the so-called “Firefly Synchronization”. Fireflies in South-East Asia are able to emit light flashes. After a few seconds they synchronize their flashing. This extraordinary behaviour, utilized on pulse-coupled oscillators, enables the self-organization in wireless networks.

How the signaling pathway operates is explained at “The cellular signaling”. Moreover, the firefly flash is explained in the section Firefly Synchronization, in detail.

II. USEFUL ESSENTIALS

At the beginning, it is advantageous to know a few fundamentals about *Dynamic Networks*, to understand the reason for the research in technology and biological ensemble. In addition to that, some basic biological areas will be explained, and we will take a look at several existing projects simulating biological environments.

A. New Network Dimensions

When the network became public, a normal static network sufficed in order to control the traffic inside. Nowadays networks increase immensely fast and require highly complex networks. Since complex static networks are intricate and more difficult to handle, it is unalterable to develop so-called *Dynamic Networks*. Dynamic Networks are a network architecture based on the construction of telephone network. Therefore it is possible to create connections between different processors as necessary. On top of that, in massively distributed systems, there has to be an efficient solution for coordination and communication. At first this could be controlled with routing algorithms from “outside” or by using so-called *Selfrouting Networks*. These are highly complex architectures which indicate convenient algorithms and logic devices. With such constructs the network is able to control the data to the endpoint, via processor controlled nodes, in dependence on the incoming header-information, without external influence. For this survey, another useful field of networking is the Sensor and Actor Networks (SANETs). This network consists of sensors to retrieve information e.g. the temperature, and actors, which are the process nodes to execute the incoming information. Due to that these SANETs are very comfortable for smart environments.

B. Biological Models

The biological sphere offers a lot of communication models which can be transferred into none-chemical techniques. With this knowledge, it is possible to develop self-organized algorithms that can be applied to *Dynamic Networks*. One of these models is the eminently effective swarm intelligence. This model describes the social behaviour and decentral communication e.g. in ant colonies or firefly swarms. In this algorithm, several so-called “agents”, i.e. logic devices, form a swarm. Each unit is autonomous and reacts to changes in the environment according to its own set of rules. Random interaction between these devices eventually leads to a form of global intelligence, without the need for a central organisation structure. A different biological technique can be observed at the intracellular communication. Here the behaviour of the cells can be analysed and it can be investigated how

the chemicals interact with different receptors. This form of communication is based on the receptors in the cell membrane which open “ports” in order to deliver signals according to certain conditions. With this, it is possible to construct a *Rule-based Sensor Network*, for example, as shown below. In the biological Activator-Inhibitor System, DNA is read according to substrate concentrations. These concentrations trigger the activator/inhibitor, thus decreasing or maintaining the substrate level. As can be seen, this system lends itself to being simulated in a digital environment. There are a lot of communication models in nature which can be used as a sequential logic system.

C. Existing Projects

Nowadays, this specific research becomes more and more important for nearly every network. Because of that the topic area *Bio-inspired Networks* was given a lot of attention. For that reason some related works were published in the last few years. A main topic for routing algorithms, task allocation or searching in *Peer-2-Peer Networks* is, for example, the so-called *Ant Colony Optimization*. This algorithm is part of the ant colony algorithms family, where the behaviour of ants in a colony is observed and digitally recreated. Specifically, in this example, the method of pathfinding of the animals is transferred to creating solutions for algorithmical problems. Additionally, there are works in reference to *Activator-Inhibitor Systems*, where mathematical models of chemical reactions concerning the space distribution of substances are researched, *Artificial Immune System*, where the aim is to simulate an immune system supervising an environment and detecting any changes from normal behaviour, *Homeostatic System* and a currently up to date topic, the *Epidemic Spreading*. Here, the observations of the spread of viral infections are used to distribute information in wireless ad hoc networks. These are only a handful of researches in the scope of *Bio-inspired Networks* [1].

III. SELF-ORGANIZED NETWORKING ALGORITHMS

Two different ways of *Bio-Inspired Networking* are discussed in the following. Both refer rather to a technique for self-organized networking algorithms than to *Bio-Inspired Networking*.

A. Rule-based Sensor Network

Rule-based Sensor Networks are based on the information transfer, specifically the cellular procedure. These techniques are converted to logic devices in Sensor and Actor Networks (SANETs) [2].

1. The cellular signaling

This section represents the information transfer in cellular environments. The exchange, which is the same approach as in network nodes, is called *signaling pathways*. Depending on the physiology of the mammal, cell neighbours have to inform each other for correct functionality.

An example for a receptor depending information transfer is the cell answer of e.g. steroid hormones. This small molecules

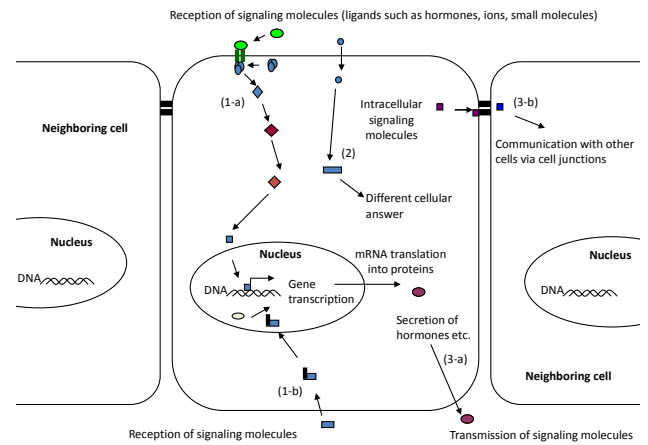


Figure 1. Cellular structure with intra and inter cellular communication [3]

become transmitted via the blood. With this pipeline the steroid hormones reach the correct destination cell. Now the molecule can bind on the specific receptor and is able to pass the cell membrane. After that the receptor-hormone liaison initiates an “cellular answer” across the *Protein Biosynthesis*.

The important fact of this signaling pathway is, that the information addresses the destination itself. During the differentiation the cells express their specific receptors, so that one cell never can have all receptors. The cell only includes the receptors for which it is originally constructed. After the conclusion of molecule, receptor and the protein biosynthesis, the cell produces its characteristic protein. The protein diffuses the membrane and is now able to induce signaling processes into the direct environment.

Another way of information transfer is the message information forwarding between neighbouring cells. At this intracellular transmission method, the cells are connected via junctions. One of these special junctions is the so-called “gap-junction”. This is a tunnel between two neighboring cells, where signaling molecules can pass through, as shown in Figure 1 (3-b). These are, of course, not the only signaling pathways. Certainly, these are the essentials of signaling pathways to develop careful rule-based logic devices with an adequate solution.

2. SANETs and the RSN model

With the knowledge about the signaling pathway in mammal cells and the essentials of former developed rule-based elements, the *Rule-based Sensor Network* (RSN) can be constructed. The RSN was developed for the so-called *Sensor and Actor Networks* (SANETs). SANETs are used wherever sensors and actors have to interact, as shown in Figure 2. These special networks can be appropriated for *Smart Environments*, medicine and similar fields.

The networks are deployed in the field level and connect the available sensors and actors. The connections between sensors and actors can be wireless but they do not have to be; furthermore, SANETs should be self-organized. To guarantee flexibility, easily transferable code and scalability

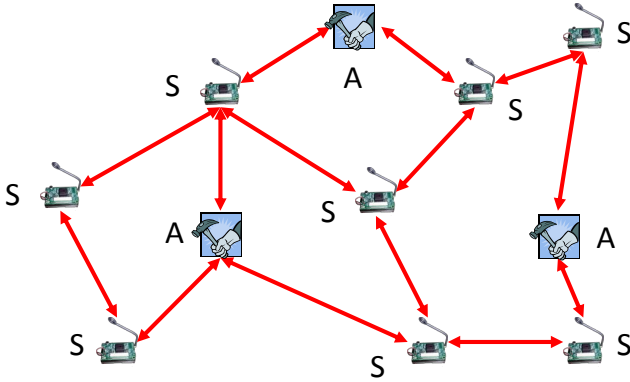


Figure 2. A network-centric operation and control model of SANETs [3]

for network-wide operations, RSN modules are deployed in SANETs.

RSN is based on the following three design objectives that enable the aforementioned objectives:

- *Data-centric communication* – Each message carries all necessary information to allow data specific handling and processing without further knowledge, e.g. about the network topology.
- *Specific reaction on received data* – A rule-based programming scheme is used to describe specific actions to be taken after the reception of particular information fragments.
- *Simple local behavior control* – We do not intend to control the overall system but focus on the operation of the individual node instead. Simple state machines have been designed, which control each node (being either sensor or actor)[3].

The messages in SANETs contain a special encoded pair with (type,content) next to the address information. The message and content are depicted through the type value. The data itself usually includes a value and application-specific meta information such as a geographical position or priority information[3]. This is the structure of a typical message which is used in SANETs:

$$M := \{ type, region, confidence, content \}$$

In every message, processed by RSN, the values type and content are mandatory. Every other value is useful for meta information which are needed for more detailed information, like geographical tags or confidence level. With such meta information the data process can be optimized. Of course, the meta information depend on the SANETs architecture and the needed information.

This is an example of how a message could look:

- $\{ temperature, [10, 20], 0.6, 20 \}$
A temperature of 20 °C was measured at the coordinates

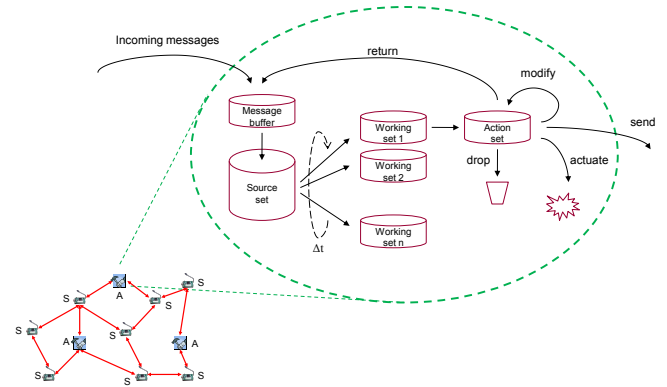


Figure 3. A single RSN node with basic components [3]

[10, 20]. The confidence is 0.6, therefore, a low-quality sensor was employed.

- $\{ pictureJPG, [10, 30], 0.9, "binary JPEG" \}$
A picture was taken in format JPEG at the coordinates [10, 30][3].

The RSN node is designed in form of a basic state machine and controlled with a rule interpreter. Furnished with a queuing subsystem, it is able to store the messages for a particular time and to act as a generic receptor for the messages. The time control prevents the system of a queue overflow because of unknown type messages. In Figure 3 the working behaviour of a single RSN node is depicted.

The incoming message is at first stored in the message buffer. The interpreter has a defined Δt and starts the process of the message periodically after Δt , or after a new message has arrived. Moreover, the Δt has to be determined in the following way: the longer messages are stored before being processed, the better is the possible RSN node behaviour (more messages can be aggregated into a single one). The longer the period, the longer the artificially introduced per hop delay [3].

After the message buffer, the messages are executed by rules composed of a condition and an action. Consequently, every rule is declared like **if** CONDITION **then**{ACTION}. Also, the rules select a message from the source set to apply the analogous action. The selected messages which contain specific attributes, are stored in "working sets". The so-called action sets define the resulting action after the right condition was found. The pattern of this construct is shown in Figure 4 and it depends on an action example. Furthermore, it is possible to access every value of a message.

For example: ($\$type == "Temperature"$) to use all messages with the chosen type, to use the priority value for identification ($\$priority > 0.9$) or to select any value bigger than the threshold, you can use this term. Moreover, every information extends a valid timestamp and a counter for the message buffer. The RSN contains the following three basic actions to control the node for the given situation:

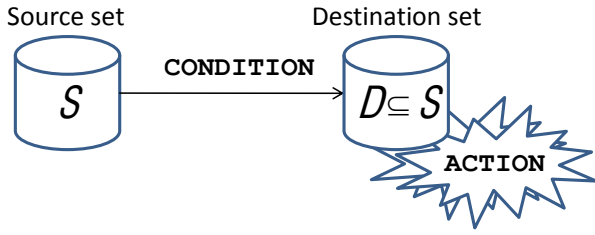


Figure 4. Each rule selects a number of messages from the source set (CONDITION) and applies a (set of) actions to the selected messages (ACTION) [2]

- *rule execution*
e.g. operations on the incoming messages
- *node control*
i.e. to control, how the node has to react
- *simulation control*
i.e. actions needed for experiment control without influence on the node behaviour.

Each of these three actions has its separate commands and works on the source message set that is constructed across the condition element. The rule execution action contains, not in detail, commands like !stop, !drop, !dropDuplicates, !return, !returnAll, !send, !sendAll and !actuate. The node control actions allow to change the current rule set and to enable or disable appended sensors and actuators. Actions to use these properties are !controlSensor - to control an appended sensor, !controlActuator - to control an appended actuator and !controlManagement, which defines the rule set itself. At last, the simulation control actions are components to deskill the control of simulation experiments. These actions do not work on a given set of messages, but it is possible to initiate these actions based on the current state of the node, i.e. after the reception of a specific message [3]. With !recordAll, statistics for each message in the working set should be recorded and !endSimulation stops a current running experiment.

The mentioned actions and conditions need local variables or message parameters to describe the node state. The most necessary attributes are *message* and *node attributes*. The first one allows to change the encoding for the specific encoded message according to the program scenario. Local state information about every node can be stored and updated. In the context of self-organization, this refers to the local state of an autonomous system. This information can be updated by other local observations.

The tables which contain these attributes are not listed in here because it would exceed this paper summary.

Nowadays, the RSN is implemented as a C++ library and allows to use all functions for a RSN statement. These statements are implemented in a flexible script language [2]. Even so, there are some open issues like the handling of unknown messages, the period of RSN execution Δt and the

rule generation and distribution which still subsist.

This was a short conclusion about the *Rule-based Sensor Networks*. In the next section, *Firefly Synchronization in Ad Hoc Networks* will be described in summary.

B. Firefly Synchronisation

Self-organization can be observed in nature, nevermind in which country or even perhaps in the sea. But in South-East Asia, for example, there exists a special variety of so-called *Fireflies* [7]. This species is able to emit a flash in different time intervals with their lower bodies. Luciferin and Luciferase are both the main components for that blinking. These two substances, in combination with oxygen and ATP (Adenosintriphosphate), result in a biological light, called *Bioluminescence*. There exist only theories, how the flashes can be controlled in duration and sequence. Even though, there is no explication for this behaviour; a mathematical model and theory has been developed by [6]. On this mathematical base, it is possible to describe the Firefly Synchronization via oscillators. The oscillators for this field of application and studies in similar biological systems are described as “pulse coupled oscillators”. [8] [6]

1. Mathematical model of firefly synchronization

These pulse coupled oscillators are constructed to observe the interaction with other neighboring oscillators at discrete times. The phase function $\phi(t)$ is a plain function to describe the oscillator in mathematics. The visualization of that effect can be described by a phase response curve (PRC). The phase function is a linear increasing function over time. If the function reaches a defined threshold ϕ_{th} , the growing will stop. In the same time, when the threshold is reached, the oscillator has to send out a pulse and to turn back to the starting point. This behaviour is called “fire”. Of course, it is possible to couple oscillators, but if this fact is not given, the standalone oscillator fires in a periodical time T . The reason for this is based on the natural definition of an oscillator, so that he has to fire in this way, when there is no external influence. This scenario is plotted in Figure 5(a) during a time T the phase function needs to reach the threshold. The firefly needs also a specific time period to emit a flash and that period until the next firing, is encoded through the phase function $\phi(t)$.

Mirollo and Strogatz analyzed spontaneous synchronization phenomena and also derived a theoretical framework based on pulse-coupled oscillators for the convergence of synchrony [8].

When two or more oscillators are affiliated with each other, they start to pulsate. That means, if one oscillator fires, it will pull the neighboring up by a fixed amount, or increment the others threshold ϕ_{th} till it will fire. Contingent on the current value of the function and an incoming signal this can be described by the following form:

$$\phi \rightarrow \phi + \Delta\phi$$

[8]

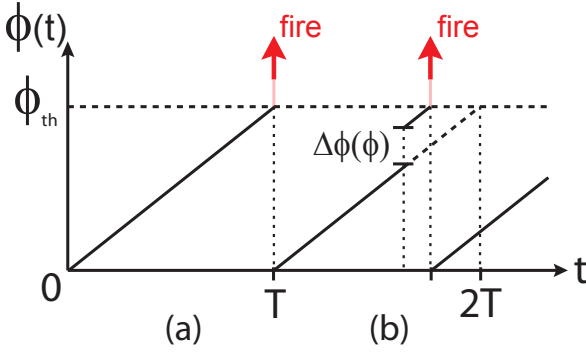


Figure 5. The plot of the phase function $\phi(t)$ [8]

By comparing the behavior and time response, in relation to the achievement of the threshold, of a single oscillator and pulse-coupled oscillators, the differences as shown in Figure 5 can be determined. In this case, you can see that the pulse-coupled one reaches the threshold before the normal time interval T is over.

This happens because of the incrementation of $\Delta\phi$ as shown in the upper form. By examining $\phi + \Delta\phi$, this leads to the following term:

$$\phi + \Delta\phi = \min(\alpha \cdot \phi + \beta, 1) \quad \text{with} \quad \begin{cases} \alpha = \exp(b \cdot \epsilon) \\ \beta = \frac{\exp(b \cdot \epsilon) - 1}{\exp(b) - 1} \end{cases} \quad (1)$$

[6]

The variables in this form are defined in the following:

- $\Delta\phi$ is the phase increment, depending on the current phase and determining the linear Phase Response Curve
- b is the dissipation factor
- ϵ is the amplitude increment

To specify the connection between neighboring oscillators, the factors b and ϵ are used. Also ϕ_{th} , the previously defined threshold, is normalized to the value one.

If the factor $b > 0$ and $\epsilon > 0$ and all nodes together in the network are connected to each other, the system reaches the convergence of synchrony. These facts are shown in [6]. This means e.g. a network consists of pulse-coupled oscillators which start to fire and convergence their firing until they fire all completely synchronized. Furthermore, the initial conditions are independent. In addition to that, the time that is needed to arrange the oscillators in synchrony is conversely proportional to the product $b \cdot \epsilon$.

Consequently the procedure of the synchronization of pulse-coupled oscillators relates to the instant of arrival of a pulse. Principally, it is to mention that every oscillator in this network has an integrated receiver. This receiver discerns the incoming pulses, processes the information and at last, adapts their phases. After observing the whole network, you can notice, that there are no typical interferences. It has been shown that when two simultaneous pulses emit, they can superimpose constructively. With this attributes, superimposed fire signals from oscillators, which are far away, they can be received even if they fire in the whole network as one oscillator.

2. Implementation in Wireless Systems

To apply the firefly synchronization, as described and defined previously, in wireless network systems, there are a few adaptations necessary. These adaptations pertain to the propagation delay. Every signal transfer entails a delay, because the information has to cover a distance. At first there is the synchronization through pulses and then there are the typical delays in wireless systems which are described in the following.

To combine the mathematical model of firefly synchronization with wireless sensor networks, there have to be some modifications to make it efficient [5]. Therefore the synchronization protocol for wireless networks has been modified. This modification was reached by accepting the so called propagation delays. A propagation delay, in this case, is the sum of time which a pulse needs to completely transfer the information from the oscillator to the receiver over a medium.

In the work from Ernst, Pawelzik and Geisel which occupies with “Synchronization induced by temporal delays in pulse-coupled oscillators”, a conclusion can be drawn [4]. This implies that a system consists of pulse-coupled oscillators proceed to an unstable system, when there arises a propagation delay T_0 between two oscillators. Adoptively, an oscillator starts to emit a pulse. This pulse reaches the receiver of another oscillator. Consequently, the pulse can effectuate, after the propagation delay time T_0 , the oscillator to fire. And this pulse anon initiates the oscillator, that emitted the first pulse, to emit a pulse after T_0 again. If this appears, the procedure will repeat all the time. T stands for the exact amount of seconds.

$$T_{\text{refr}} > 2 \cdot T_0 \quad (2)$$

This is the inequality for the refractory period time. The refractory time has to superinduce to obviate the previously described effect. This factor T_{refr} implies that a just fired oscillator cannot fire again as long as it is in the refractory time. Until the refractory period is over, the oscillator is able to emit a pulse again. At the same time the phase function $\phi(t)$ of the affected oscillator is during the entire T_{refr} set to 0. As explained, also an incoming signal cannot change the value of the phase function. The same principle can be observed in neurons.

Once the sensor network is synchronized, the protocol can also be used to propagate information in a similar fashion to the pulse-position modulation [8]. This means, that an incoming data can spontaneously be shifted from node to node what leads to a global shift. With this, the offset in the reference instant can be detected and that can be used to transfer specific information to a lurker.

The next section is about the delays in wireless systems in relation to the firefly synchronisation. To communicate in a pulse-coupled oscillator network based on the firefly synchronization algorithm, the pulses, emitted from oscillators, are decoded and received by other oscillators. In a pulse-

coupled wireless network it is normal that a few nodes fire in sequence or fire in a burst mode with a duration of T_{Tx} in the synchronization scheme. That implies in relation to the synchronization scheme that there is another timeslot where the oscillator cannot receive or fire. Furthermore, the system requires a decoding delay T_{dec} . The decoding delay is needed to handle the incoming messages in the right way and to signal that a synchronization message was received. Consequently the amount of $(T_{Tx} + T_{dec})$ describes the entire time a synchronization message needs to receive and decode.

The four delays for a wireless pulse-coupled oscillator network are:

- T_0 : Propagation delay - time taken for a burst to propagate from the emitting to the receiving node. This time is proportional to the distance between two nodes.
- T_{Tx} : Transmitting delay - length of the burst. A node cannot receive during this time.
- T_{dec} : Decoding delay - time taken by the receiver to decode a burst.
- T_{refr} : Refractory delay - time necessary after transmitting to maintain stability.

[8]

By comparing this scheme with the model developed by [6], the four delays were not mentioned or defined.

To complete the timeslots it is necessary to calculate the total delay time T_{del} .

$$T_{del} = T_0 + T_{Tx} + T_{dec} \quad (3)$$

With this form it is possible to destine the total delay. The total delay represents the period of time from the start until the end of a transmission without any mistakes. However, a computation to hold the accuracy is also necessary:

$$T_{wait} = T - (T_{Tx} + T_{dec}) \quad (4)$$

With this approach, if T_0 is neglected, the receiver will increment its phase exactly T seconds after the transmitter has fired [8].

In Figure 6, two previously synchronized pulse-coupled oscillators are plotted. At instant 0, oscillator 1 reaches ϕ_{th} . It waits until $t_1 = T_{wait}$ before starting to transmit a synchronization burst. At $t_2 = T_{wait} + T_{Tx} + T_{dec} = T$, oscillator 2 has successfully received and decoded the burst. As the two oscillators are already synchronized, it will follow the same scheme as oscillator 1, and wait until $t_3 = T + T_{wait}$ before transmitting [8].

At last, lookup at a pulse-coupled system with N nodes, every oscillator handles the synchronization messages as previously described. After a certain time two groups of oscillators can be declared. These two groups fire T seconds separately and support the synchronization of each other. With a consequential execution of the rules the network will reach the synchrony stadium within 15 periods.

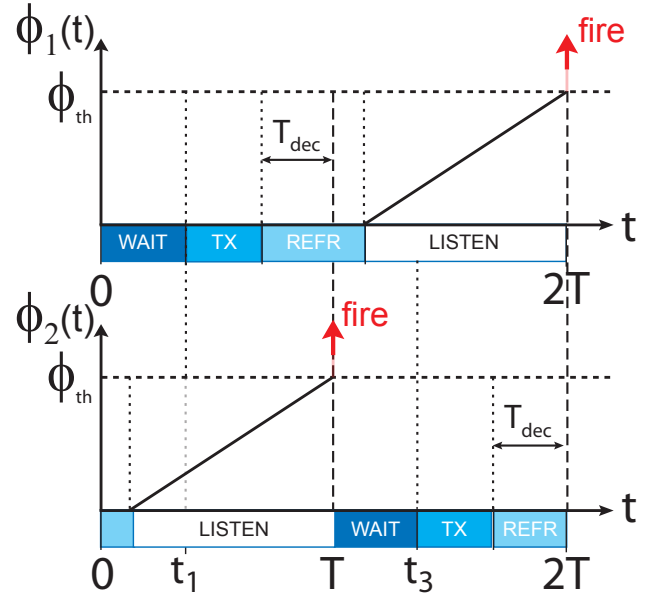


Figure 6. Time advance strategy for two previously synchronized oscillators [8]

IV. CONCLUSION

In the last chapter, two precepts were introduced. Both are inspired by biological behaviour. Starting with the essentials for different biological models, two patterns are chosen. The first one is based on the signaling pathway of cells. With this architecture it was possible to develop a logical device unit to process sensor data in sensor and actor networks, the so-called Rule-based Sensor Networks.

The other is a synchronization scheme for wireless networks. After observing fireflies this knowledge can be transferred to pulse-coupled oscillators. Also a part of neuronal stimulus forwarding is important to guarantee the frictionless execution. These two researches guarantee a network strategy with self-organization.

These are just two promising projects following the trend of combining biological knowledge with logical algorithms. As can be seen in these examples and throughout the rest of this survey, nature has a lot to offer us concerning the solutions to logical problems. In many ways, nature resolves conflicts logically and efficiently, lending itself to be simulated through algorithms. With growing knowledge in the field of computer science, we are more and more able to transfer increasingly complex biological scenarios to software. Further research into bioinformatics is a field of high interest and should be pursued in the future.

REFERENCES

- [1] O. B. Akan, “Bio-inspired Communication and Networking,” March 2009.
- [2] F. Dressler, “Bio-inspired networking - the basis for efficient communication in sensor and actor networks,” October 2007.
- [3] F. Dressler, I. Dietrich, R. German, and B. Krüger, “A Rule-based System for Programming Self-Organized Sensor and Actor Networks,” *Elsevier Computer Networks*, vol. 53, no. 10, pp. 1737–1750, July 2009.
- [4] U. Ernst, K. Pawelzik, and T. Geisel, “Synchronization Induced by Temporal Delays in Pulse-Coupled Oscillators,” 1995.
- [5] Y. W. Hong and A. Scaglione, “A scalable synchronization protocol for large scale sensor networks and its applications,” vol. 23, no. 5, pp. 1085–1099, 2005.
- [6] R. E. Mirollo and S. H. Strogatz, “Synchronization of pulse-coupled biological oscillators,” *SIAM Journal on Applied Mathematics*, vol. 50, no. 6, pp. 1645–1662, 1990. [Online]. Available: <http://www.jstor.org/stable/2101911>
- [7] S. H. Strogatz and I. Stewart, “Coupled oscillators and biological synchronization,” *Scientific American*, vol. 269, pp. 102–109, Dec. 1993.
- [8] A. Tyrrell, G. Auer, and C. Bettstetter, “Firefly synchronization in ad hoc networks.”

Whitespace Networking

Johannes Merx – {firstname.lastname}@uni-ulm.de

Abstract—Today’s wireless network systems are characterized by a fixed spectrum assignment policy. However, the available radio spectrum is limited and has become more and more inefficient concerning its utilization, due to consumers’ increasing interest in wireless services and the proliferation of wireless technologies, such as WLANs, Bluetooth and wireless mesh networks. This trend necessitates a new communication paradigm to exploit the existing wireless spectrum opportunistically. This new spectrum-aware technology is referred to as whitespace networking, which is subject of this paper. The architecture, novel functionalities and challenges of whitespace networking are explained in detail and finally applications and existing implementations of whitespace networks are presented.

I. INTRODUCTION

EXISTING networks for providing broadband wireless access employ communication technologies and spectrum policy in a very static manner. Radio spectrum is regulated by governmental agencies such as the Federal Communications Commission (FCC) in the USA, or the European Regulators Group (ERG). They assign spectrum to license holders and services, such like TV broadcast networks, cellular networks, or industrial, scientific and medical (ISM)-networks, often for a long period of time and for large geographical regions. This conventional approach to spectrum management is very inflexible since each operator requires an exclusive license to operate in a certain spectrum band and most of the useful radio spectrum is already assigned to various services. Additionally, as evidenced in recent measurements, a large portion of the licensed spectrum is rarely utilized continuously across time and space [4]. Figure 1 shows spectrum utilization in the frequency bands between 30 MHz and 3 GHz averaged over six different geographical regions in the US. As it can be seen, the spectrum usage is concentrated on certain frequency bands while most spectrum remains unutilized. Frequency bands which are not used locally over a long period of time are referred to as *whitespaces*.

Although the fixed spectrum assignment policy served comparatively well in the past, spectrum utilization in urban regions has become more and more inefficient in recent years. This is mainly caused by consumers’ increasing interest in wireless services and increasing demand for the limited radio spectrum in the license free ISM bands [7]. Moreover, with the emergence of new wireless applications and the compelling need for broadband wireless access, this trend will surely continue in the coming years and necessitate a new communication paradigm to exploit the existing wireless spectrum opportunistically. Whitespace networking is proposed to improve the inefficient usage of existing spectrum and provide high bandwidths to mobile users through heterogeneous wireless architectures and dynamic spectrum access. It is a radically different access paradigm used for wireless networks where unlicensed

users are allowed to access temporally unused but licensed frequency bands in an opportunistic way if they do not interfere with the existing users. The key enabling technology for whitespace networks is the cognitive radio, which enables wireless devices to detect whitespaces (*spectrum sensing*), select a channel according to availability and QoS requirements (*spectrum management*), coordinate access to the selected channel (*spectrum sharing*), and switch the frequency band used for communication when a licensed user appears (*spectrum mobility*).

Since a big range of frequencies will be freed up in the future due to a scheduled changeover from analogue to digital television broadcasts in the upcoming years, the FCC and ERG have already expressed their interest in permitting unlicensed access to whitespaces in TV band. This interest stems in part from the great propagation characteristics and bandwidth of the UHF spectrum, occupied by TV bands. In late 2008, the FCC has finally approved the unlicensed use of whitespaces for devices that support location finding. Building on this trend, the IEEE has formed a working group (IEEE 802.22) to develop an air interface for opportunistic secondary access to the TV spectrum [3][5]. The main goal of this endeavor is to provide wireless broadband internet access in rural areas. Finally, with KNOWS [7] and WhiteFi [2], there already exist some prototypes ready for use, which collaboratively detect whitespaces in the TV spectrum and enable opportunistic access and sharing of the whitespaces.

This paper presents an overview concerning the functions, implementation and challenges of whitespace networks. The architecture and functions of whitespace networks are presented in Section II. In Section III, a brief overview of the cognitive radio technology is provided. In Section IV, the basic spectrum sensing techniques and sensing challenges are discussed. Whitespace functionalities such as spectrum management, spectrum mobility, and spectrum sharing are introduced in Section V, VI, and VII.

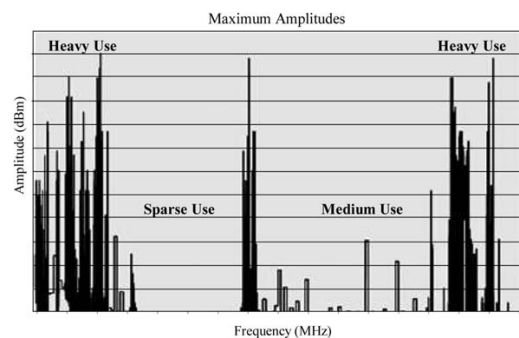


Figure 1. Spectrum utilization between 30 MHz and 3 GHz in the US. Spectrum usage is concentrated on certain frequency bands [1].

II. WHITESPACE NETWORKS

As explained in Section I, existing wireless networks employ communication technologies and spectrum policy in a static manner. Furthermore, licensed frequency bands often show unbalanced utilization in spectrum usage and unlicensed spectrum, like the ISM band, have become extremely crowded and inefficient in urban regions. Therefore, new technologies for wireless broadband access are necessary for keeping up with increasing user requirements.

Whitespace networking is a radically different and novel access paradigm for wireless networks where unlicensed (secondary) users are allowed to operate in temporally unused but licensed portions of the radio spectrum, so called whitespaces or spectrum holes, which is generally referred to as opportunistic spectrum access or dynamic spectrum access and can be enabled through cognitive radio technologies, as described in Section III. The most important challenge for whitespace networks is to use a certain spectrum band without interfering with the transmission of other users, as illustrated in Figure 2. In order to protect licensed network systems, or primary users, from the adverse effects of secondary users' interference, whitespaces need to be reliably identified by all secondary users across frequency, time, and space. Table I lists some approaches which may be employed for this purpose [4]. Whitespace identification via databases and broadcasting imply the task of providing unlicensed users with current spectrum usage information, like location and transmit power of nearby competitive network systems, as well as expected duration of spectrum usage by the primary users. Providing this information could be realized through a centralized database or by broadcasting this information on regional beacons. These methods lead to simplified unlicensed users but require modification to existing licensed systems and unlicensed users with positioning capabilities.

Spectrum sensing, on the other hand, solely relies on the unlicensed user to identify whitespaces through direct sensing of the radio spectrum. Unlicensed users are equipped with cognitive capabilities, which allow them to detect and react to changes in the radio environment autonomously. Low infrastructure cost, a high grade of flexibility, and compatibility with legacy primary systems are only a few advantages to be mentioned at this point. Section IV provides an overview of different issues associated with the implementation of spectrum sensing functionality in whitespace networks.

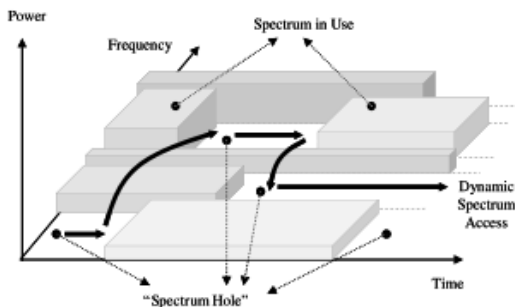


Figure 2. Illustration of Dynamic Spectrum Access. Spectrum is used without interfering with the transmission of other users [1].

Table I
CLASSIFICATION OF WHITESPACE IDENTIFICATION METHODS

	Database	Broadcasting	Sensing
<i>Infrastructure cost</i>	High	High	Low
<i>Legacy compatibility</i>			X
<i>Transceiver complexity</i>	Low	Low	High
<i>Positioning</i>	X	X	
<i>Internet connection</i>	X		
<i>Continuous monitoring</i>			X
<i>Standardized channel</i>		X	

A. Whitespace network entities and components

For the purpose of developing communication protocols, a clear description and presentation of all possible scenarios of whitespace networks is essential. The components of the whitespace network architecture are defined as follows [1]:

Primary network: An existing network infrastructure is generally referred to as primary network, which means that users of this network have an exclusive right to use the corresponding spectrum band for communication. The components of the primary network are as follows:

- *Primary User:* Primary users (or licensed users) have authorisation to operate in a certain portion of the radio spectrum. Their network access should not be affected by unlicensed users.
- *Primary base-station:* A fixed infrastructure network component, which has a spectrum license.

Whitespace network: A whitespace network does not have a license to operate in a specific spectrum band. Spectrum access for network users is allowed only in an opportunistic manner. Whitespace networks can be deployed, both, as a centralized network or a distributed (ad hoc) network. Generally, the following network components can be differentiated:

- *Secondary user:* Secondary users (or unlicensed users) have no exclusive right to use a certain frequency band. Cognitive radio functionalities are commonly required to share the licensed spectrum band without interfering with privileged primary users.
- *Secondary base-station:* A fixed infrastructure component of a whitespace network which features cognitive capabilities (III-A). Secondary base-stations provide single hop connection to users without spectrum access license.

B. Whitespace network functions

As explained before, whitespace network devices are able to adjust the spectrum band used for communication dynamically through reconfigurable radio interfaces, which implies the possibility to operate in both licensed and unlicensed bands (illustrated in Figure 3). Hence, the functionalities required vary depending on whether the radio interface operates in licensed or unlicensed spectrum bands. Thus, operations can be classified in terms of *whitespace networking on licensed band* and *whitespace networking on unlicensed band*.

Whitespace networking on licensed band: As shown in Figure 1, temporally unused portions of radio spectrum in licensed frequency bands, named whitespaces, exist in most cases, which can be exploited through cognitive communication

techniques (Section III). One main feature of whitespace networking is the coexistence with primary users and licensed networks. There are various challenges for whitespace networks operating on licensed bands due to their duty of interference avoidance with primary users. Therefore, the most important issue for operating in licensed spectrum bands consists in detecting the presence of nearby primary users. The channel capacity of spectrum holes and the channel decision mainly depends on the interference at the nearby primary users. Furthermore, whitespace devices should move to another whitespace as fast as possible if a primary user appears.

Whitespace networking on unlicensed band: First steps towards open spectrum policy were established with the industrial scientific and medical (ISM) bands (e.g., IEEE 802.11g utilizes ISM frequencies between 2,4 GHz and 2,5 GHz), which enabled an impressive variety of technologies, devices and innovative uses. However, due to the interference among multiple heterogeneous networks and an increasing amount of ISM-devices, unlicensed bands have become extremely crowded and inefficient concerning spectrum utilization, in recent years.

Whitespace networks can be designed for operation on unlicensed spectrum bands, e.g. ISM band, in order to improve the spectrum efficiency. Since there are no privileged primary users in place, all network devices have the same right to access and use a certain spectrum band. Furthermore, various whitespace networks can coexist at the same location and communicate using overlapping portions of the spectrum. Intelligent spectrum sharing techniques can improve efficiency and support high QoS requirements.

In this scenario, whitespace users focus on detecting transmissions transacted by other whitespace users. In contrast to operations on licensed bands, the spectrum hand off is not triggered by the appearance of a primary user but caused by competitive behaviour of several secondary users. Since all whitespace users have the same right to access the spectrum, all users compete with each other for the same unlicensed band. Thus, sophisticated spectrum sharing techniques are required in this architecture.

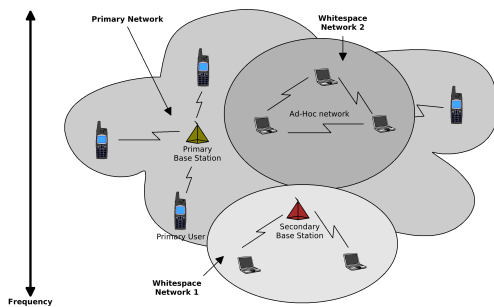


Figure 3. Whitespace networks on licensed and unlicensed spectrum bands.

C. Whitespace network applications

Raised from a high-graded flexibility and adaptability, there exist a wide-ranged field of different appliances for employing whitespace networks, which can be attached to the following applications [1]:

Leased networks: A whitespace network can be provided by allowing opportunistic access to licensed spectrum bands without sacrificing the service quality of primary users. An agreement with a third party is normally required for establishing a leased network. An example for leased networks might be the secondary access to TV spectrum for providing broadband connectivity as it is intended by the IEEE 802.22 WG.

Cognitive ad hoc network: Nowadays, wireless ad hoc networks are commonly used as a cost-effective technology for providing broadcast connectivity. Since unlicensed spectrum bands have been extremely crowded and new applications require more and more data throughput, whitespace networking can be used for establishing a multiplicity of meshed networks in dense urban areas with a high spectrum efficiency [1].

Emergency network: Public safety and emergency networks are another application field for the use of whitespace networks. For example, in the case of natural disasters, communication infrastructures may be temporarily destroyed or out of service and so called emergency networks have to be established for coordinating the work of emergency crews. Whitespace networks can enable the usage of existing spectrum without the need for an infrastructure, which can be exploited for building up an emergency network in less time.

Military network: One potential application of a whitespace network lies in its usage in military context. Due to its flexibility and reconfigurability, cognitive radios can adapt to the variable radio environment of battlefields. Furthermore, cognitive techniques could provide the opportunity of finding and alternating secure spectrum bands for military communication.

III. COGNITIVE RADIO

Cognitive radio technology enables a radio interface to use radio spectrum in a dynamic manner. Thus, we have a closer look at cognitive radio in the following. The term cognitive radio can be formally defined as follows [1]:

A 'Cognitive Radio' is a radio that can change its transmitter parameters based on interaction with the environment in which it operates.

Based on this definition, there are two basic features which characterizes a cognitive radio: *cognitive capability* and *reconfigurability* [1].

The main objective of the cognitive radio technology is to obtain the best available spectrum for transmissions through cognitive capability and reconfigurability. It is the key technology on which the whitespace network architecture is based. It enables the usage of temporally unused spectrum and simultaneously provides the ability to move to another whitespace or altering its transmission power level and/or its modulation scheme to avoid interference for the case that a primary user utilizes this channel at the same time. It is quit obvious that implementing a cognitive radio requires a new type of radio transceivers [1] and, in addition, specially adapted algorithms and protocols for realising spectrum management and spectrum sharing. In the following subsections the cognitive functions an reconfigurability capabilities of the cognitive radio technology are described in detail.

A. Cognitive capability

The ability of the radio technology to sense and process informations concerning the temporal and spatial variations of the radio environment in order to identify whitespaces and avoid interferences with other users is referred to as cognitive capability. In order to realize this capability it is not sufficient to simply monitor the power in some frequency band of interest. In the majority of cases more sophisticated techniques and procedures are required, as can be seen in the following chapters. The ability of an cognitive radio to interact with its environment to determine spectrum capacities and adapt to the dynamic radio environment is one basic difference to nowadays radio networks such as Bluetooth or IEEE 802.11, which uses predetermined spectrum channels in a very static way and normally do not possess any kind of context awareness. The task required for adaptive operation in open spectrum can be modelled by the so called cognitive cycle, which consists of the three following main steps, usually labelled as spectrum management [1]: *spectrum sensing*, *spectrum analysis* and *spectrum decision*. Issues related with spectrum management are discussed in Section IV and V.

B. Reconfigurability

Reconfigurability enables the radio interface to be dynamically programmed depending on informations about its environment provided by the cognitive capability. More specifically, the radio interface can be programmed to transmit and receive data on dynamically assigned channels of the radio spectrum, providing the capability of adjusting operating parameters for the transmission in a dynamic manner without any modifications of the hardware components. Reconfigurability enables the cognitive radio to react on modifications of the radio environment adaptively by reconfiguring the following transmission parameters:

- *Operating frequency:* Capability of changing the frequency for transmissions. Based on information about the radio environment through spectrum sensing and spectrum analysis the most suitable operating frequency can be determined and all operating functions can be dynamically shifted to this defined frequency.
- *Modulation:* A cognitive radio should adaptively reconfigure the modulation scheme to QoS parameters and channel characteristics. There are several popular modulation schemes such as, simple FSK and PSK modulation, or considerably more complex procedures like 8-PSK or various QAM schemes, which exhibit different transmission characteristics concerning spectral efficiency and error rate. For example, in the case of delay sensitive applications, the data rate is more important than the error rate. Thus, the modulation scheme that enables the higher spectral efficiency should be selected by the cognitive radio.
- *Transmission Power:* Transmission power can easily be changed during operation. If a lower level of transmission power is sufficient to maintain operation, the cognitive radio can reduce the power level accordingly. This

decreases the interference and allows more users to share the same spectrum band.

- *Communication technology:* Finally it is imaginable that a cognitive radio can also be used to provide interoperability among different wireless communication techniques.

The transmission parameters of a cognitive radio can be reconfigured not only at the beginning of a transmission but also during a transmission. In other words, a cognitive radio is able to switch to a different spectrum band at any time by changing the transmitter and receiver parameters, communication protocol parameters, and modulation scheme. Figure 4 illustrates the processes, which have to be done by a cognitive radio device for providing whitespace functionality, step-by-step.

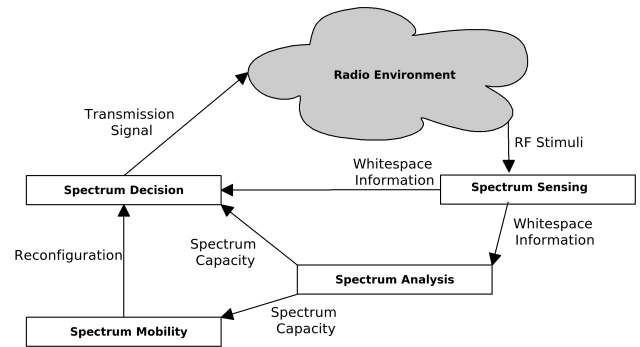


Figure 4. Cognitive radio processing, including the cognitive cycle and spectrum mobility.

IV. SPECTRUM SENSING

The most important requirement of a whitespace network is the ability to detect existing whitespaces across the radio spectrum. As explained in Chapter III, a cognitive radio is designed to be aware of its environment and changes of the radio spectrum. This kind of context awareness can either be enabled by rather static and inflexible approaches, which are commonly based on the presence of a primary network that supports secondary users by delivering whitespace informations via beacons or a central database (Section II), or by so called spectrum sensing algorithms which allow a secondary user to recognize whitespaces in a dynamic and self-contained way. Spectrum sensing as the most sophisticated and flexible approach for detecting whitespaces is discussed in this chapter. An efficient way to detect whitespaces is to sense all primary users that are receiving transmissions within the communication range of the secondary user. In reality, however, it is very difficult to retrieve such kind of direct measurement of a channel between a transmitter and receiver. Thus, recent spectrum sensing functions focus on primary transmitter detection, either based on observations of one single secondary user or computed cooperatively by several secondary users in a whitespace network.

Generally, the spectrum sensing techniques can be classified as *transmitter detection* and *cooperative detection*.

A. Transmitter detection

Having the capability to determine if a signal from another transmitter is locally present in a certain spectrum channel is one of the main tasks of a cognitive radio, the key technology of our whitespace radio architecture. Transmitter detection is based on sensing for weak signals, which normally come into existence by the presence and activity of other users. A quite simple and basic hypothesis model for transmitter detection can be defined as follows [1]:

$$x(t) = \begin{cases} n(t) & H_0, \\ hs(t) + n(t) & H_1, \end{cases} \quad (1)$$

where $x(t)$ is the signal received by the secondary user, $s(t)$ is the transmitted signal of the primary user, h is the amplitude gain of the channel, and $n(t)$ is the Additive White Gaussian Noise (AWGN), a model in which the only impairment is a linear addition of white noise with a constant spectral density and a Gaussian distribution of amplitude t . H_1 indicates the existence of a primary user signal, whereas H_0 is a null hypothesis, which indicates that there is no primary user signal in a given spectrum band.

Generally, transmitter detection can be classified into *power sensing techniques* and *feature sensing techniques*, which are quite different from each other in terms of requirements and field of application. Feature sensing techniques can be used if specific information about the primary user signal is known to the users of the whitespace network. For example, a secondary user might be aware that it is sensing in spectrum bands which are exclusively licensed for TV broadcasting and is, therefore, able to utilize specific signal features of analogue TV signals (e.g., NTSC) or digital TV signals (e.g., ATSC) for the purpose of identifying licensed users. On the other hand, power sensing techniques do not rely on special signal features and can be used if a receiver cannot gather sufficient information about the primary user signal.

In literature a multiplicity of different algorithms has been proposed for transmitter detection and evaluated. The most promising power sensing techniques are *energy detection*, *eigenvalue-based sensing* [5], *covariance-based sensing* [5], and *multi-resolution sensing* [5]. Whereas techniques like *matched filter* and *cyclostationary feature detection* are commonly used for feature sensing. Three of the most popular and best evaluated algorithms for transmitter detection are described in the following:

1) *Energy detection*: If the receiver cannot gather sufficient information about the primary user signal, one of the best detectors is an energy detector [5][4]. This detector measures the energy received in a channel during an observation interval and compares that estimate to a threshold, in order to decide whether a primary user might be active or not. In order to measure the energy of the signal, the baseband signal with bandwidth W is squared and integrated over the observation interval T . Finally, the output of the integrator, Y , is compared with a threshold λ , in order to decide whether a licensed user is present or not.

If energy detection is used in non fading environments, more specifically, in environments which are not effected by

phenomena like shadowing or multi-path fading, the probability of correct primary signal detections P_d and false alarms P_f is defined as follows [1]:

$$P_d = P\{Y > \lambda | H_1\} = Q_m(\sqrt{2\gamma}, \sqrt{\lambda}) \quad (2)$$

$$P_f = P\{Y > \lambda | H_0\} = \frac{\Gamma(m, \frac{\lambda}{2})}{\Gamma(m)} \quad (3)$$

where γ is the Signal-Noise-Ratio (SNR), $m = TW$ is the time bandwidth product, $\Gamma(\cdot)$ and $\Gamma(\cdot, \cdot)$ are complete and incomplete gamma functions and $Q_M(\cdot)$ is the generalized Marcum Q-function. While a low P_d would result in a high probability of missing the presence of a primary user in a spectrum band, a high p_f would result in low spectrum utilization since false alarms increase the number of unused but free spectrum channels.

In case of environments influenced by shadowing and multi-path fading, P_f is independent of Γ . Furthermore, the amplitude gain h varies due to shadowing and fading which results in an instantaneous SNR [1]. As it can be seen, this sensing technique relies on a priori knowledge of the noise power. However, the performance of energy detection is highly susceptible to uncertainty in noise power [5].

Compared to other sensing techniques, like matched filtering, energy detection requires a longer sensing time to achieve a desired and adequate performance level. However, its low cost and implementation simplicity render it a favourable candidate for spectrum sensing in cognitive radio systems.

2) *Matched filter*: When specific information of the primary user signal is known to the secondary user, the optimal detector for maximizing the SNR in the presence of stationary Gaussian noise is the matched filter [6]. The matched filter can detect the presence of a licensed user signal in spectrum band by correlating a known template of potential user signals with an unknown signal, in this case, the received signal within a certain spectrum band. In order to obtain the signal template, the match filter technique requires a priori knowledge of the primary user signal such as the modulation type, the pulse shape and the packet format. If this knowledge is not accurate, the matched filter performs pretty poorly. However, since most wireless network systems have pilot codes, synchronisation words, or spreading codes for spectrum spread techniques, that information can also be used for the coherent detection.

3) *Cyclostationary feature detection*: The cyclostationary feature detection is an alternative and also popular feature based detection method [1][5]. Most network systems use modulated signals for transmitting digital data, normally generated with sine wave carriers, pulse trains, repeated spreading, hopping sequences, or cyclic prefixes. This results in a built-in periodicity of the modulated user signals, which are characterized as cyclostationarity since their mean and autocorrelation exhibit periodicity. These cyclostationary features can be detected by analysing a spectral correlation function. Since the white noise within a certain spectrum channel is a wide-sense stationary signal without correlation, the cyclostationary feature detection is able to differentiate the noise energy from modulated signal energy, which exhibits spectral correlation due to the embedded

redundancy of signal periodicity. Therefore, the cyclostationary feature detection offers good robustness to the uncertainty in noise power, but, on the other hand, it is computationally complex and requires a long observation time.

B. Cooperative detection

Spectrum sensing in cognitive radio networks is challenged by several sources of uncertainty, which often cannot be solved by the transmitter detection model [1].

A low received signal strength does not necessarily imply that no primary system is located inside the secondary user's interference range, as the signal could be affected by phenomena like long term fading or being shadowed by obstacles [4]. Moreover, the transmitter detection model cannot prevent the hidden terminal problem, which occurs when a primary node is visible from a primary base station, but not from other secondary nodes. Therefore, a secondary user needs to carefully distinguish a faded or shadowed radio signal from a whitespace. Since spectrum sensing should perform robustly even under worst conditions, and because there is no possibility of interaction between a primary user and a secondary user, in most cases, the sensing information from other secondary users is required for more accurate and robust whitespace detection. In the case of non-cooperative sensing explained in subsection IV-A, the process of identifying unused spectrum bands is based on independent local observations by each secondary user. In contrast, cooperative detection refers to spectrum sensing methods where information from multiple secondary users are combined for primary signal detection. Generally, this can be implemented either in a centralized or in a distributed manner. The centralized alternative can be realized by installing a base station that gathers all sensing information from the existing secondary users in order to detect existing spectrum holes. Another possibility is to exchange all sensing observations among secondary user in a distributed manner. The latter tends to be more flexible, but also involves a higher complexity and information overhead. The exchange of sensing information among all secondary users can be provided through common control channels, which are commonly used for device coordination in wireless networks (Section VII-A).

C. Design trade-offs

Concerning the implementation of spectrum sensing functionalities in whitespace networks, system designers should balance some major trade-offs according to application specific requirements, complexity, hardware costs, and available infrastructure. Basic decisions have to be made concerning the *sensing periodicity* and *sensing sensitivity* of a cognitive radio device.

The sensing period determines the interval in which a secondary user repeats spectrum sensing while utilizing a certain whitespace. Therefore, the sensing interval represents the maximum time during which the secondary user will be unaware about changes in the radio environment, and hence may interfere with new primary users. Since short sensing periods imply higher loads for secondary users, in most cases, sensing periodicity strongly depends on application and type of the primary system.

Another trade-off has to be made concerning the sensitivity of primary signal detection. The robustness toward interference of a primary user strongly depends on the type of the primary service as well as on the spectrum band. Therefore, sensing sensitivity has to be adapted to the intended use of a whitespace network.

Finally, further trade-offs could be the number of the cooperating users which are considered and the choice between proactive and reactive sensing [4].

V. SPECTRUM MANAGEMENT

Normally, whitespaces are spread over wide frequency ranges including both unlicensed and licensed bands. Due to different and frequency-dependent propagation effects of radio transmissions, unused spectrum bands show different characteristics which have to be considered when a decision is made which band should be chosen, in order to meet given QoS requirements. Therefore, new spectrum management functions are required, which have to be specifically adapted to dynamic spectrum characteristics. As mentioned in Section III, these functions can be classified as *spectrum sensing*, *spectrum analysis* and *spectrum decision*. While spectrum sensing is primarily a PHY layer issue and is discussed in Section IV, spectrum analysis and spectrum decision are related to upper software layers.

A. Spectrum analysis

In whitespace networks, the available spectrum holes show different characteristics which vary over time. Therefore, it is important to understand the characteristics of different spectrum bands. Spectrum analysis enables the characterization of different spectrum bands, which can be exploited to select the spectrum band fitting the QoS-parameters. In order to describe the dynamic nature of whitespace networks, each spectrum hole should be characterized considering the time-variant radio environment, the primary user activity, and information such as operating frequency and bandwidth. Hence, it is essential to define various parameters which represent the quality of a whitespace as follows [1]:

- *Interference:* Some spectrum bands are more crowded compared to others. Hence, the spectrum band in use determines the interference characteristics of the channel.
- *Path loss:* The path loss increases as the operating frequency increases. Therefore, if transmission power is increased to compensate this path loss, this results in higher interference for other users.
- *Wireless link errors:* The error rate of a certain channel. This error rate depends on the modulation scheme and the interference level.
- *Link layer delay:* Subject to different path loss, wireless link error and interference, different types of link layer protocols are required for different whitespaces. This results in different link layer packet transmission delay.

Combining all characterization parameters described above can be employed in order to identify appropriate spectrum bands for a specific application, which is necessary to fulfill the subsequent spectrum decision.

B. Spectrum decision

Once all available spectrum bands are characterized, an adequate band should be selected for transmission considering the QoS requirements and the estimated channel characteristics. Based on various parameters (e.g. data rate, acceptable error rate or delay bound), the transmission mode and bandwidth can be computed, and an appropriate spectrum band can be chosen according to a certain spectrum decision rule. As described in [1], decision rules are mostly focused on a trade-off between fairness of all secondary users and communication cost.

VI. SPECTRUM MOBILITY

According to previous sections, the main target of whitespace networking is to utilize the radio spectrum in a dynamic manner by allowing the radio terminals to operate in the best currently available frequency band. *Spectrum mobility* is defined as the process when a secondary user changes its operation frequency, which arises when current channel conditions become worse or a primary and privileged user appears. This change can occur before starting transmissions as well as during a processed transmission – one major difference compared to existing traditional wireless networks. Hence, if the spectrum in use cannot be utilized any further, the communication needs to be continued in another channel by switching the transmitters' and receivers' spectrum for operation in a coordinated way. Spectrum mobility demands a new type of hand-off in cognitive radio networks that is referred to as *spectrum hand-off*. The protocols for different layers of the network stack must adapt to the channel parameters of the operating frequency and need to be transparent to the spectrum hand-off at the same time. As pointed out before, a cognitive radio is able to operate on various frequency bands. Therefore, each time the frequency of operation is being changed, the network protocols have to shift from one mode of operation to another. The purpose of spectrum mobility management is to ensure that such transitions are processed smoothly and as soon as possible. Consequently, multi-layer mobility management protocols are required to accomplish the spectrum mobility functionalities. These protocols support mobility management adaptive to different application types. For example, for data communication such as FTP, the mobility management protocols should provide techniques to store packets that are received during spectrum hand-off, whereas for real-time applications there is no need of buffering data packets.

VII. SPECTRUM SHARING

One of the main challenges in whitespace networks lies in the task of spectrum sharing, which can be regarded more or less to be similar to generic medium access control (MAC) problems in existing wireless networking systems. However, the coexistence of primary and secondary users in licensed channels, and accordingly the coexistence of multiple secondary users, as well as the wide ranged and dynamic allocated operation spectrum lead to some specific challenges for spectrum sharing in whitespace networks. In this section, the different challenges and steps concerning spectrum sharing are discussed and a short overview of spectrum sharing techniques is presented.

A. Coordination challenges

Implementing spectrum sharing functionality for cognitive radio devices necessitates sophisticated device coordination of all existing transmitter and receiver within a whitespace network. Since there may be multiple secondary users trying to access same spectrum, their access needs to be coordinated in order to prevent multiple users colliding in overlapping portions of the spectrum. Also, a transmitter-receiver handshake protocol is essential for communication in whitespace networks. Once a specific spectrum band is determined for communication, the receiver of this communication must be informed about the selected spectrum. Finally, further coordination efforts are needed for providing spectrum mobility during the transmissions of secondary users. If the spectrum in use is required by a privileged user, the communication needs to be continued in another channel by switching the spectrum for operation in a coordinated way.

Many traditional spectrum sharing solutions assume a fixed common control channel (CCC) for providing spectrum sharing functionalities, which might be also employed for whitespace networks. However, due to the fact that secondary users are regarded as visitors to the spectrum they allocate, the CCC has to be vacated without interfering. As a result, the CCC is shown as highly dependent on the topology and application and varies over time. Hence, implementation of a fixed CCC is infeasible for whitespace networking. Consequently, for protocols requiring a common control channel, a distributed coordination scheme can be used [8]. All secondary users are self-organized into local coordination groups based on similarity of available channels. Members of each group form a mini multi-hop network and use a common control channel for coordination. Since radio range changes with operating frequency due to attenuation variation [8][1], the choice of a control channel needs to be carefully decided. It would be much efficient to select control channels in the lower portions of the spectrum where the transmission range will be higher and to select data channels in the higher portions of the spectrum where a data transmission can be utilized with minimized interference and high data rates.

For the case that no CCC is used, coordination becomes a challenge. For instance, receiver driven techniques may be exploited for realizing transmitter-receiver handshakes.

B. Overview of spectrum sharing techniques

Solutions for spectrum sharing in whitespace networks can be classified by three main aspects, according to their *architecture assumption*, *spectrum allocation behaviour*, and *spectrum access techniques* [1].

The first classification for spectrum sharing techniques is based on the architecture of the whitespace network, which generally can be described as follows:

- *Centralized spectrum sharing*: In these approaches, a centralized entity controls the spectrum allocation and access procedures. Normally these procedures require distributed sensing techniques to realize dynamic spectrum access. This can be realized by forwarding all measurements of

the network nodes to the centralized entity, which finally constructs a so called spectrum allocation map.

- *Distributed spectrum sharing:* Distributed approaches are mainly proposed for cases where infrastructure deployment is not possible or preferable. In these solutions, each user of a whitespace network is responsible for its own spectrum allocation. Spectrum allocation and access is based on local policies.

The second classification for spectrum sharing techniques in whitespace networks is based on the access behaviour, which can be described in terms of *cooperative* or *non-cooperative*.

- *Cooperative spectrum sharing:* Cooperative solutions consider how one's communication affects other nodes. In other words, each node shares its interference measurements with the other nodes. The spectrum allocation algorithms also consider this information. While all centralized solutions for spectrum sharing can be considered cooperative, there also exist possibilities for realizing distributed cooperative spectrum sharing techniques.
- *Non-cooperative spectrum sharing:* In contrast to cooperative approaches for spectrum sharing, non-cooperative solutions consider only one node at hand, which can be referred to as a selfish kind of spectrum sharing. Main advantage of non-cooperative spectrum sharing techniques may be the minimal communication requirements among other nodes, for which reason this techniques offers a good trade-off for practical solutions in most cases.

Finally, the third classification for spectrum sharing in whitespace networks is based on the access technology that is used.

- *Overlay spectrum sharing:* Overlay spectrum sharing refers to a common access technique at which a node accesses the network by using a portion of the spectrum that has not been used by primary users. Since this results in minimized interference to the primary system in most cases, such approaches are popular, due to their simplicity and robustness. However, overlay spectrum sharing often means less bandwidth utilisation compared to other, more sophisticated techniques.
- *Underlay spectrum sharing:* Underlay spectrum sharing exploits various spread spectrum techniques (e.g., DSSS), which are commonly used by existing network architectures in order to minimize interference at the nodes. Once a spectrum band has been allocated, the secondary user begins transmission so that its transmit power is regarded as noise by licensed users. This technique requires sophisticated spectrum sharing techniques but can utilize increased bandwidth compared to overlay techniques.

VIII. CONCLUSION

Whitespace networks, equipped with context-aware capabilities of cognitive radio technology, are being developed to solve current network problems, which mainly result from the limited available spectrum and its inefficient usage, on the one hand, and an increasing demand for broadband wireless network services, on the other hand. Opportunistic spectrum access is envisioned to improve spectrum utilization by allowing users to

dynamically allocate and use whitespaces across the licensed radio spectrum for communication without interfering with other users.

In this paper, the main issues associated with the design and functionalities of whitespace networks are investigated by a bottom-up approach, starting from basic cognitive radio functionalities, like spectrum management and spectrum mobility, to communication protocols for dynamic spectrum sharing. Moreover, an insight into remaining problems and challenges for a practicable use of whitespace networks are provided.

In summery, whitespace networking is a multifaceted problem demanding coordinated efforts of communication technologies and protocols, on the one hand, and of spectrum regulatory, on the other hand. Although, there already exist some working prototypes and the IEEE 802.22 standard is partly completed, much more research is needed to ensure efficient spectrum-aware communication in the future.

REFERENCES

- [1] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," vol. 50. Science Direct, May 2006, pp. 2127—2159.
- [2] P. Bahl, R. Chandra†, T. Moscibroda†, and M. Welsh, "White space networking with wi-fi like connectivity," in *SIGCOMM'09*. Barcelona, Spain: ACM, Aug. 2009.
- [3] C. Cordeiro, K. Challapali, D. Birru, and S. S. N., "IEEE 802.22: The first worldwide wireless standard based on cognitive radios," vol. 1. IEEE, Apr. 2006.
- [4] A. Ghasemi and E. S. Sousa, "Spectrum sensing in cognitive radio networks: Requirements, challenges and design trade-offs." IEEE, Apr. 2008.
- [5] R. K. Gupta and S. D. Senturia, "Spectrum sensing in IEEE 802.22," in *IAPR Workshop on Cognitive Information Processing (CIP) 2008*. Santorini, Greece: IEEE, Jun. 2008.
- [6] G. L. Turin, "An introduction to matched filters," in *IRE Transactions on Information Theory*. IEEE, Jun. 1960, pp. 311–329.
- [7] Y. Yuan, P. Bahl, R. Chandra, P. A. Chou, J. I. Ferrell, T. Moscibroda, and Y. W. Srihari Narlanka, "KNOWS: Cognitive radio networks over white spaces," in *New Frontiers in Dynamic Spectrum Access Networks, 2007*. Dublin, Ireland: IEEE, Apr. 2007, pp. 416–427.
- [8] J. Zhao, H. Zheng, and G.-H. Yang,, "Distributed coordination in dynamic spectrum allocation networks," in *New Frontiers in Dynamic Spectrum Access Networks, 2005*. Dublin, Ireland: IEEE, Nov. 2005, pp. 259–268.

Cloud Computing

Benjamin Erb – {firstname.lastname}@uni-ulm.de

Abstract—This paper gives an overview over the essential concepts of Cloud Computing and the technologies it is associated with. First we have a look at various Cloud Computing definitions, the three different service types and existing technologies related to Cloud Computing. Then we elaborate technologies that enable Cloud Computing to evolve. Next, we point out the key features that cloud-based services provide and present some real-world cloud-based services. Finally we have a look at some of the remaining challenges in the context of Cloud Computing.

I. INTRODUCTION

AFTER the Service-Oriented Architectures (SOA) hype, it looks like Cloud Computing is the next big thing to hit the whole IT market. While for some it is just an obvious evolution of currently well-deployed technologies, others proclaim that Cloud Computing represents a complete paradigm shift. It could provoke a general change of information architectures and services such as the invention of timesharing or the introduction of the personal computer did before [10]. No matter if this opinion is appropriate or not, Cloud Computing is definitively the buzzword on everyone's lips. Gartner, an information technology research and advisory company, publishes their annual market prognosis in form of hype cycles in order to visualize various technological trends and their estimated impact. In 2009, Cloud Computing has reached the climax of emerging technologies in their hype cycle chart (see also Fig 2). According to Gartner, no other IT technology has drawn more attention, but also fueled more expectations in 2009.

Surprisingly, there are various approaches around trying to define what Cloud Computing actually is and mark out what it is not. The most common approaches describe Cloud Computing as the provision of various services with a focus on scalability on a pay-per-use pricing. These services range from virtualized server instances up to the offering of ready-to-use software via the web [15]. It transparently shifts the location of the IT infrastructure to the network and enables cost reductions and business advantages. Cloud Computing allows applications to elastically use resources that dynamically scale on demand. The main resources are computing power, data storage capacity and network performance. In the next sections, we have a look at various facets of Cloud Computing including features, technologies and challenges.

II. CLOUD SHAPES

Due to the hype nature, there is a lot of confusion around concerning the actual meaning of Cloud Computing. Furthermore, it incorporates various surrounding and related technologies, which complicates an exact definition additionally.

A. A Cloud Computing Definition

There is a huge number of different definitions available towards Cloud Computing. The term *cloud* itself is derived from the figurative abstraction of the internet represented as a cloud. The available definitions of Cloud Computing range from reductions it to be “*the next hype term*” [7] over pragmatic definitions focusing on special aspects to definitions, that see Cloud Computing as a general paradigm shift of information architectures [1]. Vaquero et al. [15] have studied 21 different definitions [7] and extracted common features. While a minimum common denominator was almost not detectable for them, scalability, pay-per-use utility model and virtualization have been elaborated as recurrent terms in many approaches. Vaquero et al. propose in [15] a consensus definition by their own:

“Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilization. This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the Infrastructure Provider by means of customized SLAs.”

This definition incorporates all essential characteristics of Cloud Computing. Due to the fact that Cloud Computing is still evolving, it only reflects current aspects and may be altered in the future.

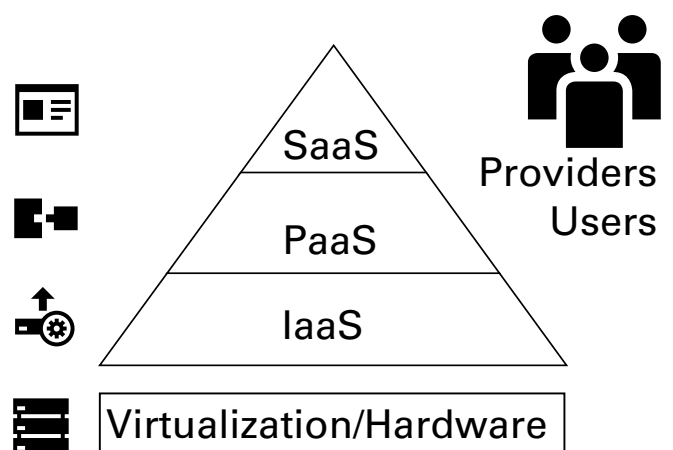


Figure 1. The Cloud Pyramid: Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). Providers offer their services which are consumed by users. Icons by <http://picol.org/>, released under Creative Commons-License BY-SA.

B. Cloud Pyramid

The key concept of Cloud Computing is the resourcing of services. These services can be generally differentiated into three vertical layers, also known as the so-called Cloud Pyramid [14] (see also Fig. 1).

1) *Infrastructure as a Service*: The provision of virtualized hardware, in most cases as virtual computing instances, is termed IaaS. The particular feature of this form of cloud service is the ability to change these instances on demand and at any time. This includes spawning new computing instances, altering existing ones by resizing or reassigning resources or shutting down unneeded instances dynamically. Basic working units for IaaS are virtual images, that are instantiated and deployed in the cloud and executed within virtual machines. They contain an operating system and additional software implementing the application. Due to complete virtualization, the physical infrastructure of the data centers of IaaS providers is totally transparent to their customers.

2) *Platform as a Service*: PaaS provides an additional level of abstraction by offering an entire runtime environment for cloud-based applications. PaaS typically supplies a software stack of dedicated components where applications can be executed on, but also tools facilitating the development process. The platform hides all scalability efforts from the developer, thus it appears as one single system to the user. Once an application is uploaded onto the platform, the system automatically deploys it within the nodes of the data center of the PaaS. When the application is running under heavy load, new nodes will be added automatically for execution. If demand declines, spare nodes will be detached from the application and returned into the pool of available resources.

3) *Software as a Service*: SaaS provides a web-centric supply of applications. Due to various technological trends, web-based applications mature and become powerful pieces of software running entirely within the browser of the user. By replacing traditional desktop applications that run locally, SaaS providers are able to publish their applications solely to the web. Cloud architectures allow them to cope with a huge number of users online. It allows users to access the applications on demand and they can be charged merely by usage without having to buy any products or licenses. This reflects the idea to consider the provision of software *as a service*.

III. RELATED TECHNOLOGIES

Interconnecting large numbers of computers in order to create powerful systems is not new—it is the key concept of distributed systems. Instead of expensive hardware upgrades of distinct machines, the integration of new nodes is here an appropriate solution for boosting overall system performance. Thus, Cloud Computing can not just lean on operational experiences of distributed systems and middleware architecture, but also reinforce available forms of network-based computing like grids and clusters [11].

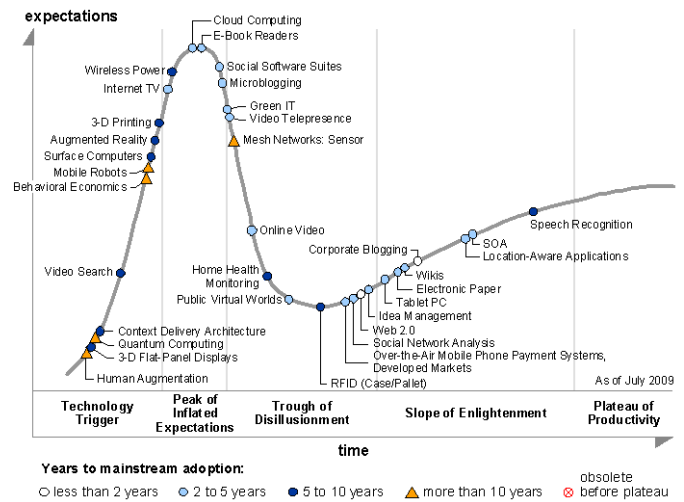


Figure 2. Cloud Computing climaxing the *Hype Cycle of Emerging Technologies*: Gartner, 2009. Source: <http://www.gartner.com/it/page.jsp?id=1124212>

A. Distributed Systems, Middleware Architectures and Service-Oriented Architectures

Distributed systems run applications by deploying components to several machines. Communication between these nodes empowers the system to integrate the different machines into the main system or application context. While basic network communication functions are generally provided by the operating systems, distributed applications often demand for a more specific way to communicate between distributed nodes, thus demanding for a middleware layer. Such an abstraction layer hides the underlying infrastructure and the complexity of heterogeneous environments. It provides generic functionalities for all nodes.

Invoking functions on remote components is a classical middleware approach, also known as Remote Procedure Call (RPC). Object-oriented programming languages introduced a more object-centric view, which led to the distribution of objects. The methods of these distributed objects can then be called remotely. In the Java world, this is done by Remote Method Invocation (RMI). Common Object Request Broker Architecture (CORBA) is a more generous architecture, that enables object distribution even between programs implemented in different languages. A new middleware paradigm is called SOA, focusing on process orientation. In SOA, software components provide services, which are defined by an external interface, but the internal implementation is hidden and transparent. Thus, SOA aims at simplifying migration and integration. Applications are created by combining different services.

Also representing a distributed system, cloud architectures must provide abilities to support some mechanisms of distributed computing. Though, cloud architectures often try to generally hide the underlying platform architecture and maintain a strict transparency by providing implicit mechanisms.

B. Grid, Cluster and Utility Computing

Cloud Computing is often confused with Grid, Cluster or Utility Computing. In fact, it combines traits from all of

these technologies, preventing clear distinctions. Computer clusters are groups of linked computers often dedicated to specific computing tasks like high performance computing. The associated machines are often strongly coupled and are located inside a shared data center. Grid Computing is similar to Cluster Computing, however the systems are coupled more loosely. Grids often consist of different computing subgroups that don't have to be collocated, so that grids nodes can have dynamical membership and are more heterogeneous. Utility Computing describes the provision of IT infrastructure including management tasks on pay-per-use basis.

IV. ENABLING TECHNOLOGIES AND CORE FEATURES

Various technological trends and economic forces are leading to the emerging presence of Cloud Computing. The more important drivers are described in this section. It also features an overview about the key features and solutions Cloud Computing offers.

A. Web-centric Computing

An important motivation for Cloud Computing is the prevalent dominance of web-based applications. Today, web applications exploit new fields of applications, that were restricted to desktop applications before. Asynchronous Javascript and XML (AJAX) as a technological evolution enables web applications to present an user experience that is almost indistinguishable from the behavior of traditional applications running locally. However, these applications reside entirely within the browser and store their data in the web and not on the harddisk of the user. For desktop applications, the operating system is the platform and execution environment, for web applications it is the browser. Traditional office suite applications like word processors and spreadsheets — which had also been responsible for the success of personal computers in the past — are already available as web applications. Such web-based services allow the user to access his data at any time and to work everywhere as long as there is an internet connection and an internet-enabled device. This also facilitates collaborative work for spatially dispersed users.

Having lightweight clients for end user access and powerful remote machines providing the services can be seen as a revival of the thin-client idea. The thin client model proposed in the '90s has proven not to represent an economical solution [16]. Due to the pay-per-use nature of cloud-based services and the negligible prices of web-enabled devices, the general conditions have changed and thus web-centric computing seems to be more promising.

B. Service-Oriented Architectures

A service-oriented view is the key characteristic of the SOA paradigm, where services represent the basic components of every architecture. SOA offers a suite for integrating, combining and orchestrating different services in order to build larger applications on top. These basic services provide common external interfaces and hide internal implementation details. These uniform interfaces, often in the form of web

services, enable a comprehensive form of interoperability and independence of languages and platforms. Components can be easily distributed and loosely coupled, as long as they use the same communication schema in order to exchange data—no matter on which system they run, how they are implemented or where they are located. A typical protocol stack for such communication using web services consists of a set of protocols. SOAP, originally defined as Simple Object Access Protocol, is used as a communication protocol between services and Web Services Description Language (WSDL) describes the services providing meta data. Universal Description, Discovery and Integration (UDDI) has been introduced as a global registry protocol for web services, however it is not used anymore in public mainly due to its complexity. The protocols for web service have a Extensible Markup Language (XML) based format in common, which facilitates interoperability but also causes a large overhead of messages. Other criticism concerning web services and SOA mention the higher overall complexity as opposed to dedicated middleware systems.

In the context of Cloud Computing, SOA is not really interesting because of the similar hypes that both have caused. Instead, it is interesting because of the common service-oriented aspects that they substantially bear on. Thus clouds represent an reasonable basis for SOA to be rested upon.

C. Infrastructural Aspects

Several companies heavily rely on the availability of their IT infrastructure for their business. Therefore, they have created powerful data centers to service their customer by all means. For instance, hardware failures for an e-commerce company leads directly to a financial loss. They also have to cope with varying service demands. Especially around Christmas, an e-commerce company has to deal with additional customers. Similar situations can be noticed for search engine providers or online news magazines in case of incidents with high public interest. Prior to Cloud Computing, the only solution was infrastructural enlargement. This has led to big data centers with low utilization at most of the time. Nowadays, such companies either can deploy their applications on external clouds or even became a cloud provider themselves by selling their unused computing power. This is exactly what Amazon did when introducing Amazon Web Services.

Another interesting development in the context of data centers involves the installed hardware. While some decades ago, mainframes have represented the core of every data center, smaller server machines have taken over this position. Today, data centers are often filled with racks hosting many servers. However, they are still designed and built for dedicated purposes, often containing specialized hardware—and again very expensive. So instead of using a small number of high performance servers, it points out to be a new trend to use commodity hardware in large numbers. Hardware failures, that are hence more probable, can be compensated by inexpensive replacements and lower maintenance costs. And an equivalent performance is maintained just by the larger number of hosts. Google, for instances, holds a network of estimated half-million servers with many of them being regular personal computers for their internal business tasks [16].

Reducing data center costs also includes external aspects of infrastructure. Data centers consume huge amounts of energy in order to run. Up to 50 percent is only used for cooling. So costs can be reduced by acquiring cheap energy and increasing cooling efficiency. Additional criteria for designing and locating data centers are environmental issues like the availability of green energy and the accessibility to Internet backbone connections.

But the most important infrastructural aspect of Cloud Computing and probably one of its the key enablers is located within the data centers. It is virtualization. Hardware virtualization enables a decoupling of physical hardware and operating systems. Thus it introduces an additional layer of abstraction by hiding the real physical hardware from the virtualized instances. The “real” system provides virtual machines that represent runtime environments for systems deployed within. While these environments are actually sandboxes, they are not distinguishable from real hardware for the operating system running inside these virtual machines. Virtualization makes systems portable. Computing resources are not bound anymore to a specific hardware to run on. They rather can be seen as virtual operating units that can be exported, moved, duplicated and deployed freely on different physical hardware thanks to virtualization. Virtualization allows consolidation of internal computing resources by reassigning virtualized instances dynamically. Utilization can be heavily increased by aggregating various instances to a few physical machines.

D. Scalability & Elasticity

Probably the most interesting feature of Cloud Computing is scalability. The usage and load of almost any application cannot be exactly determined beforehand. Thus, an optimal execution environment scales with the usage. Scalability includes different dimensions. Some applications are computing-intensive and require more instances to work on concurrently. Other applications call for dozens of additional storage units in order to challenge their data flow. Then again there are applications that demand for extremely high network traffic volumes that must be mastered by the infrastructure.

There are several ways of elastically changing the resources assigned to applications. Some systems allow to define rules, that describe the scaling behavior. The rules actually describe minimal and maximal sizes like the number of instances but also events that trigger changes. Based upon these rules, the system can automatically execute changes. Other systems only allow manual changes. However, it is still a powerful tool to acquire hundreds of computing instances within a click of a button.

Also different forms of elasticity are available implying transparency. An *internal cloud* is the resource pooling of all servers within the organization. Applications are deployed to the internal cloud and maintain inside the organizational IT infrastructure. However, applications and services can be consolidated allowing to shut down and restart servers depending on the current performance requirements. When the services are deployed to an *external cloud*, the company relies on a external service provider. This means to outsource

the infrastructure but also treads new paths of scalability opportunities. There is also an combined approach propagating a *hybrid cloud*. Essentially, it is an internal cloud which uses an external cloud for overflow handling when reaching internal capacity limits.

E. Computing

One key aspect of distributed computing is the desired additional performance when using many interconnected nodes. This is also an important goal for Cloud Computing, especially when offering computing power as a resource. However, it is important to realize, that not every application can utilize additional nodes natively. For scalability, computations or application tasks must be partitionable into smaller jobs than can be distributed and deployed onto different nodes. This is comparable to the recent situation of microprocessor development. Due to architectural issues, the performance of processors can be increased mainly by adding more cores into one chip at the moment. Such multicore systems do not automatically speed up traditional applications. According to *Amdahl's Law*, application code must be able to profit from parallelization in order to scale performance with the additional number of cores [9]. Similar properties apply to distributed computing. Applications must provide code that is parallelizable for concurrent execution on various nodes and in parallel. This is the primary reason why many large web application are written in functional programming languages like Erlang. Due to the nature of these languages, functional programs can easily be parallelized without any additional effort.

Although not strictly bound to Cloud Computing, new approaches towards parallel computing frameworks have been made. These frameworks facilitate the design and implementation of parallelizable code by hiding the intricacy of concurrency and parallelism inside the framework. At the moment, the most prominent representative of such frameworks is MapReduce [4]. It was introduced by Google and is used heavily for internal computations like page ranking. According to Google, regular MapReduce computations process many terabytes of data on thousands of machines and over thousand MapReduce jobs are executed on Google's clusters every day. The MapReduce programming model requires the user to implement a map function and a reduce function. The map function processes key/values pairs and generates intermediate sets of key/value pairs. The reduce function then merges all intermediate values and generates the computational result. The MapReduce architecture follows the master/slave schema. A master node distributes subtasks to a large number of slaves and controls the overall program. As addition to this distributed programming model, there is often a special distributed file system used, that supports data propagation between the collaborating nodes, such as the Google File System [8].

F. Data Storage

Cloud-based applications often rely on some kind of storage for persisting data entities. Other cloud services provide a dedicated, cloud-based data storage for external usage. These data storage engines combine aspects related tot the cloud

like scalability, reliability and availability with the features of traditional database management systems like consistency or transactional features. Cloud-aware database management systems evolved from distributed database management systems. A distributed database is a set of databases distributed onto multiple nodes. However, the distributed database appears as on logical database to the user. Typical distribution patterns are master-slave architectures or peer-to-peer systems. Distributed database management systems provide extended features like load balancing, fault-tolerance, data replication or enhanced reliability.

In the context of Cloud Computing, there has been a remarkable change concerning the prevailing database models used by applications. While relational or object-relational database management systems are used by the majority of web applications today, cloud-oriented applications often try to rely on non-relational database management systems. These database systems are based upon flat or hierarchical data models and don't impose a database schema. In contrast to relational data models, these data models have a weaker expressiveness. For instance, they do not support foreign keys and cannot execute query joins efficiently, whereas scalability can be heavily improved [1, 5]. Relational models demand a high organizational overhead when executing complex join queries using foreign keys or locking various tables while running in transaction contexts. Schema-less non-relational data models are kept simple and some of them even degrade to the structure of a hashtable. Instead of storing data in rows and columns, values are saved using a key, which is necessary for later retrieval. Both — keys and values — are arbitrary objects and the hashcode of the key is calculated for internal data mapping. Such key-value stores can easily be distributed by partitioning the space of possible hashes.

Due to the nature of the cloud, cloud-based database systems must cope with several different failure situations and provide necessary recovery mechanisms in order not to break classical database principles. Transactional atomicity, structural integrity, data consistency and persistence durability must be maintained even in case of massive concurrent read/write operations, node crashes or failing network connections. When provided as a service, also the enforcement of service level agreements and quality of service contracts must be adhered to. This often leads to a tradeoff between availability, consistency, cost-effectiveness and performance.

G. Economic Drivers

A significant difference between Grid Computing and Cloud Computing is the primary driving force behind. While most Grid systems are mainly based on public research projects and scientific usage, cloud architectures are mostly driven by commercial development and interest. Nowadays, the IT department of an average medium-size company is very likely to operate their own IT infrastructure in-house, including servers. Larger companies may even possess corporate data centers. Especially when the core business relies heavily on the availability of the corporate IT infrastructure, new hardware is acquired with peak service demands in mind. This leads

to powerful in-house data centers, that are hardly utilized. So expensive hardware is rarely used, although it represents a real investment. This investment is called *capital expenditure*, due to the fact that the acquirement of new hardware represents an asset. Unfortunately, these investments do not scale well with changes of the business environment. When a company buys new servers due to heavy load for a distinct time, it cannot return these new assets when demands fade. According to an economic point of view, these infrastructural costs should closely correlate with the actual resource use, thus becoming operational expenditures instead of capital expenditures.

This is exactly what Cloud Computing services provide due to their *pay-per-use* pricing. Depending on the type of service, companies can purchase virtualized hardware, traffic volumes, storage capacities, computing power or software depending on their actual demand. This not just allows to scale up in order to cope with usage peaks, it also allows companies to scale down their resources in case of declining demands. Outsourcing hardware and IT infrastructure also includes outsourcing administrative tasks, because the service providers become responsible for maintaining the infrastructure as part of their service.

When offering Cloud Computing services, providers must define detailed Service Level Agreements (SLA) and Quality of Service (QoS) guarantees [2]. Companies cannot rely on best-effort services for their businesses, thus appropriate SLA and QoS guarantees are existential together with adequate payment models. Contractual commitments typically ensure certain types of availability like infrastructure uptime, available computing power, network bandwidth and latency.

Several concerns exist about the effects of outsourcing corporate IT to the cloud. First of all, moving important business data and knowledge from the internal IT to external data centers controlled by third-party demands a high degree of trust. It also implies a distinct loss of power and control of the IT department of the company.

V. USE CASES

The following section presents some representative cloud-related services and applications.

A. IaaS: Amazon Elastic Compute Cloud

The Amazon Elastic Compute Cloud¹ is one of the most famous IaaS hosts today and was also one of the first cloud based services of Amazon that has been opened for public usage. The user can dynamically acquire instances of virtual machines and deploy his own system using the Amazon Machine Image (AMI) format. Users can choose between various instances of different size and even select special instances for CPU intensive tasks. AMI instances can be configured locally and deployed to Amazon Elastic Compute Cloud or the user can pick up pre-installed images and configure it depending on his requirements. Also the location can be chosen between different data centers worldwide. All administration tasks like booting and monitoring instances are handled via a web interface.

¹<http://aws.amazon.com/ec2/>

Amazon charges the users for the hourly usage per virtual machine and the data transfer volume. Additional features like automated scaling, external persistent storages or dedicated IP addresses are also available for a fee.

B. PaaS: Google App Engine

The Google App Engine² represents a PaaS for web applications. It is hosting also official applications like Google Docs. Applications for the Google App Engine can be implemented in Python or Java. It also provides a schema-less key/value storage for data entities. The software stack for Java supports popular technologies like the Java Servlet API and the Java Persistence API. However, there are several restrictions for the sandbox applications are executed in, for instance maximum time limits or the prevention of thread creation or file access. Google also provides a development toolkit containing tools for local execution, debugging and deployment. The Google App Engine can be used for free as long as the application sticks to several hard limits of distinct features. For instance, the CPU time, HTTP request count, data storage size or bandwidth is subjected to a daily quota. However, application deployers can pay for additional resources or quota extensions.

C. SaaS: Google Docs

Google Docs³ is typical representative for SaaS. It features an online word processor, spreadsheet, presentation, and form application that run entirely inside the browser. It is free of charge and provides several collaboration features. Google Docs supports many popular data formats for import/export and thus offers interoperability and relieve the danger of a vendor lock-in.

D. Software Frameworks: Hadoop

Hadoop⁴ is an Apache-backed, Java-based open source framework for distributed applications. It was heavily inspired by Google's MapReduce [4] and the Google File System [8]. IBM and Google are supporting the project financially for educational reasons. Hadoops hosts additional sub-projects that are cloud-oriented, like HBase and Hypertable, two distributed non-relational database systems. In 2008, Yahoo installed a Hadoop cluster that has won the terabyte sort benchmark, sorting 1 terabyte of data in 209 seconds on 910 nodes [13]. Thus a cluster completely built upon open source software defeated all proprietary systems.

E. Scalable Databases: CouchDB

CouchDB⁵ is a open source database system written in Erlang and developed by the Apache Software Foundation. It is document-oriented and neither has a schema, nor is it using rows or columns. Data entities are saved in form of JSON [3] documents. Although it does not maintain any schema, queries

can be called by using views. These views execute MapReduce-like filters and aggregation functions. CouchDB is designed for easy replication and to scale vertically. It can be accessed via HTTP using a RESTful API [6].

F. Summary

The previous presentations introduced different cloud-based services and applications. Of course, this was a very selective overview. There is already a large number of other services available. While there are probably thousands of web applications, that have been designed or redesigned as SaaS, there are just a few providers for PaaS or IaaS. Some of them are small, dedicated providers, others are big companies extending their portfolios. There are also efforts for open source PaaS and IaaS solutions. One of them is Eucalyptus⁶, an open source implementation compatible with Amazon AWS.

VI. CHALLENGES

Being a rather new technological trend and still evolving, there are still many unresolved issues for Cloud Computing. Solving them however is vital, as some of them prevent potential users to adapt Cloud Computing services for their own applications due to remaining concerns.

A. Privacy & Security

The most commonly mentioned objection to Cloud Computing is the loss of control when outsourcing internal business data to external cloud providers. This implies obvious security and privacy concerns resulting in a high demand of trust in the providers [12]. Critics point out that corporate espionage embodies a dangerous threat inside Cloud Computing systems due to the opaqueness of infrastructures for the user. On the contrary, opponents state that 75-80 percent of intellectual property breaches are performed inside the own company [1].

Also the Cloud infrastructure itself can be attacked or compromised. Such attacks can range from single virtualization jailbreaks to the defection of nodes with hypervisor rootkits and denial-of-service attacks against providers in large scale.

Another aspect focuses on the appropriate location of data in the cloud. It questions the applicability and possible enforcement of laws in multilateral scenarios. This is interesting due to different national laws in the country of the service provider, the service user, the end-user and the country the data center is located in. Security and privacy concerns are not limited to the users of infrastructural services. Also end-users can be affected when their personal data is not stored anymore in their offices and homes, but in the cloud. Private data are rendered to an external authority facing problems when there are issues with the provider or the account of the user.

B. Standardization

Another objection to Cloud Computing is the threat of vendor lock-ins due to proprietary architectures. Without standardization, user of all kind of cloud services are bound

²<http://appengine.google.com/>

³<http://docs.google.com/>

⁴<http://hadoop.apache.org/>

⁵<http://http://couchdb.apache.org/>

⁶<http://open.eucalyptus.com>

to a distinct provider they depend on. This limits competition and hinders the user to choose freely an appropriate service provider.

This not just reminisces about similar situation in the mainframe area [16], when companies were bound to the architecture of a certain vendor for decades. It also endangers the open nature of the Internet. This becomes especially dangerous, when service providers team up with application providers and hardware vendors for proprietary solutions bound to branded access devices, software systems or browsers. At least for IaaS, the AMI format is apparently gaining acceptance since it is also supported by major open source organizations and popular Linux distributions.

C. Reliability

Another challenge addresses the service providers. IaaS, PaaS and SaaS solutions are only interesting for commercial usage, when there is a high degree of reliability. The key instruments therefore are detailed QoS guarantees and SLA. Taking into account the high number of integrated system nodes, providers do not only have to bother with system faults, that are almost certain. They must also prepare for disasters and appropriate recovery strategies in case of bigger failures or force majeure risks.

VII. CONCLUSION

As we have seen in the previous chapters, Cloud Computing is able to affect almost all levels of today's IT—whether it is regarded as a complete paradigm shift or just an consistent evolution of existing technologies that are merged together. The basic entity of Cloud Computing is the provision of different resources *as a service*. Services are outsourced into the cloud, which means that the infrastructure is located externally and is accessed via network. This infrastructure is often maintained by a provider running a large data center with many machines in order to cope with the changing demands of its clients. Data centers often follow the model of traditional cluster or grid systems and heavily rely on virtualization.

Offered cloud services range from virtualized hardware instances over dedicated software stacks and platforms to complex web-based software applications. All these services can be scaled dynamically and on demand. Typical resource scalability dimensions are the number of instances, the available computing power, storage capacities and network performance. The user is accounted for his dynamical usage on a pay-per-use basis known from utility computing. Outsourcing the IT infrastructure and implementing a smooth scalability path represents an operational expenditure, which is economically more favorable than capital expenditure caused by new hardware assets. But scalability does not come for free. In order to profit from a large number of distributed computing instances, applications must be adapted. Traditional relational database systems are a good example of applications that are hard to scale, answered by a new hype of schema-less non-relational database platforms. Other results are new programming models facilitating distributed computing by hiding most of the complexity.

However, Cloud Computing is still evolving and must challenge various issues in order to succeed sustainably. Both the providers of cloud infrastructure and further research must face various objections for a long-term success. A high degree of trust is necessary to cope with the fear of missing data security and privacy. But users should also bear in mind the threat of vendor lock-ins due to missing standardization efforts and check reliability guarantees by the providers.

REFERENCES

- [1] "Cloud Computing: An Overview," *Queue*, vol. 7, no. 5, pp. 3–4, 2009, advisor-Creeger, Mache.
- [2] R. Buyya, "Market-Oriented Cloud Computing: Vision, Hype, and Reality of Delivering Computing as the 5th Utility," in *CCGRID '09: Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid*. Washington, DC, USA: IEEE Computer Society, 2009, p. 1.
- [3] D. Crockford, "The application/json Media Type for JavaScript Object Notation (JSON)," RFC 4627 (Informational), Internet Engineering Task Force, Jul. 2006. [Online]. Available: <http://www.ietf.org/rfc/rfc4627.txt>
- [4] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, 2008.
- [5] G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Voshall, and W. Vogels, "Dynamo: amazon's highly available key-value store," in *SOSP '07: Proceedings of twenty-first ACM SIGOPS symposium on Operating systems principles*. New York, NY, USA: ACM, 2007, pp. 205–220.
- [6] R. T. Fielding, "Architectural styles and the design of network-based software architectures," Ph.D. dissertation, 2000, chair-Taylor, Richard N.
- [7] J. Geelan, "Twenty-One Experts Define Cloud Computing," <http://cloudcomputing.sys-con.com/node/612375>, January 2009, visited on 15.12.2009. [Online]. Available: <http://cloudcomputing.sys-con.com/node/612375>
- [8] S. Ghemawat, H. Gobioff, and S.-T. Leung, "The Google file system," *SIGOPS Oper. Syst. Rev.*, vol. 37, no. 5, pp. 29–43, 2003.
- [9] G. Goth, "Entering a parallel universe," *Commun. ACM*, vol. 52, no. 9, pp. 15–17, 2009.
- [10] B. Hayes, "Cloud computing," *Commun. ACM*, vol. 51, no. 7, pp. 9–11, 2008.
- [11] K. Hwang, "Keynote I: Massively Distributed Systems: From Grids and P2P to Clouds," in *GPC-WORKSHOPS '08: Proceedings of the 2008 The 3rd International Conference on Grid and Pervasive Computing - Workshops*. Washington, DC, USA: IEEE Computer Society, 2008, p. xxii.
- [12] B. Michael, "In Clouds Shall We Trust?" *IEEE Security and Privacy*, vol. 7, no. 5, p. 3, 2009.
- [13] O. O'Malley, "Apache Hadoop Wins Terabyte Sort Benchmark," http://developer.yahoo.net/blogs/hadoop/2008/07/apache_hadoop_wins_terabyte_sort_benchmark.html, July 2008, visited on 21.12.2009. [Online]. Available: http://developer.yahoo.net/blogs/hadoop/2008/07/apache_hadoop_wins_terabyte_sort_benchmark.html
- [14] M. Sheehan, "Cloud Computing Expo: Introducing the Cloud Pyramid," <http://cloudcomputing.sys-con.com/node/609938>, August 2008, visited on 15.12.2009. [Online]. Available: <http://cloudcomputing.sys-con.com/node/609938>
- [15] L. M. Vaquero, L. Roderio-Merino, J. Caceres, and M. Lindner, "A break in the clouds: towards a cloud definition," *SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 1, pp. 50–55, 2009.
- [16] A. Weiss, "Computing in the clouds," *netWorker*, vol. 11, no. 4, pp. 16–25, 2007.

The Future Vehicle

Sung-Eun Kang – sung-{firstname.lastname}@uni-ulm.de

Zusammenfassung—Dieses Paper behandelt unter dem Thema *The Future Vehicle* aktuelle Forschungstrends, ohne Anspruch auf Vollständigkeit, aus dem automobilen Umfeld und der Informatik. Dabei wird gezeigt, dass unterschiedliche Disziplinen der Informatik, angefangen bei der Netzwerkkommunikation über Human Computer Interaction und Computer Vision in Fahrzeugen bis hin zur Künstlichen Intelligenz und Robotik, einen Platz in diesem Umfeld finden.

I. EINFÜHRUNG

INTELLIGENTE Systeme wie Fahrspurkontrolle, automatische Bremssysteme bei drohender Kollisionssituationen und halbautomatische Einparksysteme, haben es heute längst in die Serienproduktion moderner Fahrzeuge geschafft. Diese Onboard-Systeme mit deren komplexen Hard- und Software machen das Autofahren immer sicherer. Doch ist hier noch lange nicht das Potenzial ausgeschöpft um weiterhin die Sicherheits- und Hilfssysteme in Fahrzeugen zu verbessern. Von grundlegenden Konzepten, wie der Car-to-Car/Car-to-Infrastructure Anwendung, in der sich Fahrzeuge in Warnsituationen austauschen, über alternative Interaktionsmöglichkeiten, die weniger vom Fahren ablenken bis hin zu autonom fahrenden Fahrzeuge gibt es noch zahlreiche Forschungsaktivitäten, dessen Resultate in naher und ferner Zukunft auf dem automobilen Markt anzutreffen sein werden. Auch Aspekte, die das Autofahren und Aktivitäten, die damit verbunden sind, komfortabler zu gestalten haben in diesem Forschungsumfeld ihren Platz. In den folgenden Kapiteln werden ausgewählte Forschungsprojekte und Studien vorgestellt, in denen Informationstechnologie dazu verhilft, das Fahrzeug von morgen zu schaffen. Dabei gibt dieses Paper jeweils Einblicke in verschiedene Themengebiete, die unterschiedliche Aspekte im Fahrzeug adressieren. So wird im Kapitel II mit Car-to-X ein Gesamtkonzept vorgestellt, das Sicherheit und Effizienz im Straßenverkehr erhöhen soll. Im Kapitel III werden durch alternative HCIs verschiedene Interaktionsmöglichkeiten zwischen Fahrer und Fahrzeug untersucht, während im Kapitel IV Monitoring-Systeme vorgestellt werden, die den Zustand der Fahrer erkennen soll und somit der Sicherheit dienen. Die im Kapitel V vorgestellten autonomen Fahrzeuge bilden das futuristische Themengebiet, mit der Vision von kooperativ und selbstständig fahrenden Fahrzeugen im Straßenverkehr.

II. C2C UND C2X KOMMUNIKATION

Drahtlose Kommunikation im automobilen Umfeld haben in der wissenschaftlichen Forschung in den letzten Jahren stetig an Bedeutung gewonnen und haben so auch breites Interesse bei der Automobilindustrie geweckt, wodurch zahlreiche Forschungsprojekte und -programme mit Beteiligung von verschiedenen Automobilhersteller ins Leben gerufen wurden [5].

Dabei wird bei der Kommunikationsanwendung zwischen Car-to-Car Kommunikation (C2C) und Car-to-Infrastructure (C2I) Kommunikation unterschieden. Bei der C2C Kommunikation handelt es sich um die Kommunikation zwischen verschiedenen Fahrzeugen, während bei der C2I Kommunikation die Fahrzeuge mit fest installierten Netzknoten in Verbindung stehen. C2C und C2I können als Car-to-X (C2X) Kommunikation zusammengefasst werden. Abbildung 1 zeigt mögliche Kommunikationsszenarien im C2X Umfeld.

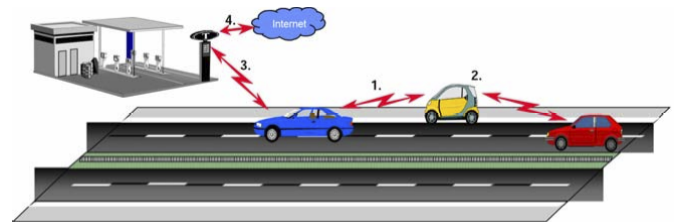


Abbildung 1. Kommunikationsszenarien: 1. Direktes C2C 2. Indirektes C2C 3. C2I 4. C2I über Internet [7]

Diese drahtlose Kommunikationstechnologie kann bestehende Sicherheitssysteme in Fahrzeugen, wie optische Radar- und LIDAR-gestützte Erkennungssysteme, erheblich erweitern und somit zur Erhöhung der Straßensicherheit beitragen. Fahrzeuge würden sich gegenseitig über Fahrbahnbeschaffenheit, Gefahren und andere Informationen austauschen und somit ein kooperatives, vernetztes System bilden. So wird die Sichtweite der Fahrer deutlich vergrößert. Dadurch könnte auch der Verkehrsfluss auf Straßen geregelt werden und somit Stausituationen vermieden werden. Des Weiteren wären Anwendungen, die der reinen Unterhaltung und Entertainment dienen möglich, wie etwa Telefonieren, Kurzmitteilungen und Chat im Fahrzeug. C2I Anwendungen umfassen Einsatzfelder wie automatisiertes Bezahlen in Parkhäusern, Kommunikation mit Ampeln, Verkehrsschildern und Schranken.

Bereits bestehende Verkehrsmeldesysteme, wie zum Beispiel Verkehrsdurchsagen im Radio oder TMC, informieren heutzutage bereits Verkehrsteilnehmer über aufkommende Staus und möglichen Ausweichrouten in annehmbarer Geschwindigkeit und Zeit. Diese Anwendungen basieren aber auf serverorientierte Architekturen, die auf eine bestehende Infrastruktur zur Datenübermittlung zugreift. Bei C2C und C2I Anwendungen basieren dagegen auf direkte und zeitnahe Kommunikation zwischen den Fahrzeugen und Straßenkomponenten ohne eine vermittelnde oder koordinierende Infrastruktur zu verwenden. Mit diesem System ist die Skalierbarkeit einer Anwendung erheblich höher und außerdem kann schneller auf Ereignisse reagiert werden.

Die Fahrzeugkommunikation bietet besondere technische Herausforderungen, da C2C und C2X Kommunikation bestimmte Anforderungen an Verfügbarkeit, Performance und Vertraulich-

keit besitzt [10]. Die Fahrzeuge bewegen sich zum einen mit relativ hohen Geschwindigkeiten, bei denen der Dopplereffekt eine Rolle spielt, die es zu kompensieren gilt. Außerdem besitzt das Fahrzeugnetz eine hochdynamische Topologie, bei der Fahrzeuge schnell in Reichweite kommen und sich wieder entfernen. Auch tages- und situationsbedingt schwankt die Anzahl der Netzknoten stark, da zum Beispiel nachts auf Landstraßen wenig Verkehr herrscht, während in Städten und insbesondere in Stausituationen die Anzahl der Teilnehmer sehr hoch ist. Würden alle Fahrzeuge in einem Stau mit voller Leistung senden, so würde auch die Kollisionswahrscheinlichkeit bei der Übertragung der Daten steigen. Des weiteren spielen Abschattung und Mehrwegeausbreitung des Funksignals wegen des dynamischen Umfelds der Fahrzeuge eine Rolle. Natürlich müssen auch die Bereiche IT Security und Privacy in einem solch komplexen System adressiert werden. Zum einen muss das System vor Angriffen gegenüber Dritten sicher sein und zum anderen müssen die Teilnehmer des Fahrzeugnetzwerks vertrauen können, dass die Daten korrekt sind und anonymisiert werden, sodass keine Rückschlüsse auf sie erfolgt werden kann. Ein Großteil an Themen dieses Forschungsgebiets war und ist in den letzten Jahren davon geprägt, einen technischen Standard für die Fahrzeugkommunikation zu entwickeln [15]. So haben sich in Europa die Firmen Audi, BMW, Daimler und Volkswagen zu einem Car-to-Car-Communication-Consortium zusammengeschlossen, um zusammen, ähnlich wie in den USA das Vehicle Safety Communications Consortium, ein Standard für die C2C und C2X Kommunikation zu schaffen.

Um EU- und weltweit einen offenen Standard für jeden Fahrzeughersteller zu gewährleisten, hat die Europäische Kommission, sowie US-Behörden in Nordamerika, bereits die Frequenzbänder im Bereich von 5,9 GHz für Verkehrssicherheitsanwendungen der C2C und C2X Kommunikation reserviert. Um den hochdynamischen Netzanforderungen der Fahrzeugkommunikation gerecht zu werden, entstand die IEEE 802.11p Arbeitsgruppe um den erweiterten W-LAN Standard *Wireless Access for the Vehicular Environment* (WAVE) speziell für die Fahrzeugkommunikation zu erarbeiten [16]. 802.11p baut auf bereits vorhandene Konzepte aus 802.11 Standards wie OFDM und CSMA/CA auf, wobei es im vorgesehenem Frequenzbereich zwischen 5,85 GHz und 5,925 GHz arbeitet. 802.11p unterstützt die drahtlose Kommunikation zwischen Fahrzeugen, die sich bis zu 200 km/h bewegen, ermöglicht Übertragungen über eine Reichweite von bis zu 1000 m und weist Datentransferraten zwischen 4 ms und 50 ms. Zwei Wellenausbreitungsmodelle (Line of Sight, None Line of Sight¹) werden eingesetzt, um Fehlermodelle durch Reflexionen und Mehrwegeausbreitung bei der Datenübertragung zu berechnen.

A. Sichere Intelligente Mobilität - Testfeld Deutschland (simTD)

Ein aktuell interessantes Vorhaben bildet das im November 2008 gestartete simTD Forschungsprojekt, das erstmals in Deutschland C2C und C2I Anwendungen in einer realen Umgebung mit einer großen Anzahl an Testfahrzeugen einsetzt [13]. Dazu sollen etwa 400 Fahrzeuge mit der benötigten Hard-

und Software ausgestattet, sowie 150 sogenannten Roadside Stations (RS) in der Frankfurter Innenstadt und Autobahnabschnitt aufgestellt werden, die als Ampeln, Verkehrsschilder und anderen Festnetzknuten Daten verarbeiten und mit den Fahrzeugen austauschen.

Durch ein Auswahlverfahren wurden in den folgenden Kategorien, die verschiedene Funktionen beinhalten, die C2C und C2X Anwendung umgesetzt:

Verkehr

- **Erfassung der Verkehrslage und Basisdienste:** Basisinformationen wie infrastrukturelle und fahrzeugseitige Datenerfassung, sowie Ermittlung und Identifikation der Verkehrslage und Verkehrseignissen
- **Verkehrs(fluss)-Informationen und Navigation:** Straßenvorausschau, Baustelleninformationssystem und erweiterte Navigation
- **Verkehrs(fluss)-Steuerung:** Beinhaltet Umleitungsmanagement und Lichtsignalanlagensteuerung

Fahren und Sicherheit

- **Lokale Gefahrenwarnung:** Umfasst die Funktionen zur Warnung anderer Fahrzeuge über Hindernisse, Stauende, Straßenwetter und Einsatzfahrzeuge
- **Fahrerassistenz:** Assistenzsysteme für Verkehrszeichen-Assistent und -Warnung, Ampelphasen und Kreuzungs-/Querverkehrsassistent

Ergänzende Dienste

- **Internetzugang und lokale Informationsdienste:** Internetbasierte Dienstnutzung und Standortinformationsdienste

Abbildung 2 zeigt den Grobaufbau und eingesetzte Technologien. Die Verkehrszentrale dient als zentrales Datenverarbeitungssystem, Steuerungs- und Datenhaltungssystem, das mit relevanten Umweltdaten gespeist wird und diese über die Versuchszentrale an die RS Units sendet. Wenn nun eine gefährliche, meldenswerte Situation eintritt, können Fahrzeuge in einem bestimmten Umkreis von den RS Units gewarnt werden. Die gewarnten Fahrzeuge könnten wiederum weitere Fahrzeuge per C2C-Kommunikation direkt oder durch Multihop benachrichtigen. Das simTD Projekt ist das erste seiner Art, dass Fahrzeug- und Übertragungstechniken zusammen mit einer Verkehrssteuerung dieser Größenordnung umfassend verbindet. Teilziel ist es somit, herauszufinden, ob RS Units und die Verkehrszentrale zusammen mit den Fahrzeugnetzknuten schnell genug reagieren können um Informationen rechtzeitig darstellen zu können. Ein weiterer Aspekt ist die Integration von weiteren Diensten wie UMTS und GPRS um Verbindungslücken im W-LAN zu überbrücken. Herauszufinden, wie man alle drei Techniken gleichzeitig gezielt ohne Verzögerung beim Umschalten innerhalb der selben Anwendung nutzen kann ist ein weiterer Forschungsaspekt. Darüber hinaus soll untersucht werden, wie andere Kommunikationstechniken, wie zum Beispiel UMTS oder Bluetooth, die Interaktion erweitern können.

Neben den technischen Umsetzungen werden aber auch andere Belange adressiert, welche die Umsetzung der Fahrzeugkommunikation in die Realität betreffen. Zum einen müssen Strategien für eine flächendeckende Einführung der C2C und C2I Anwendungen entwickelt werden, damit die ersten mit der Hard- und

¹<http://www.itwissen.info/definition/lexikon/Nicht-Sichtverbindung-NLOS-non-line-of-sight.html>

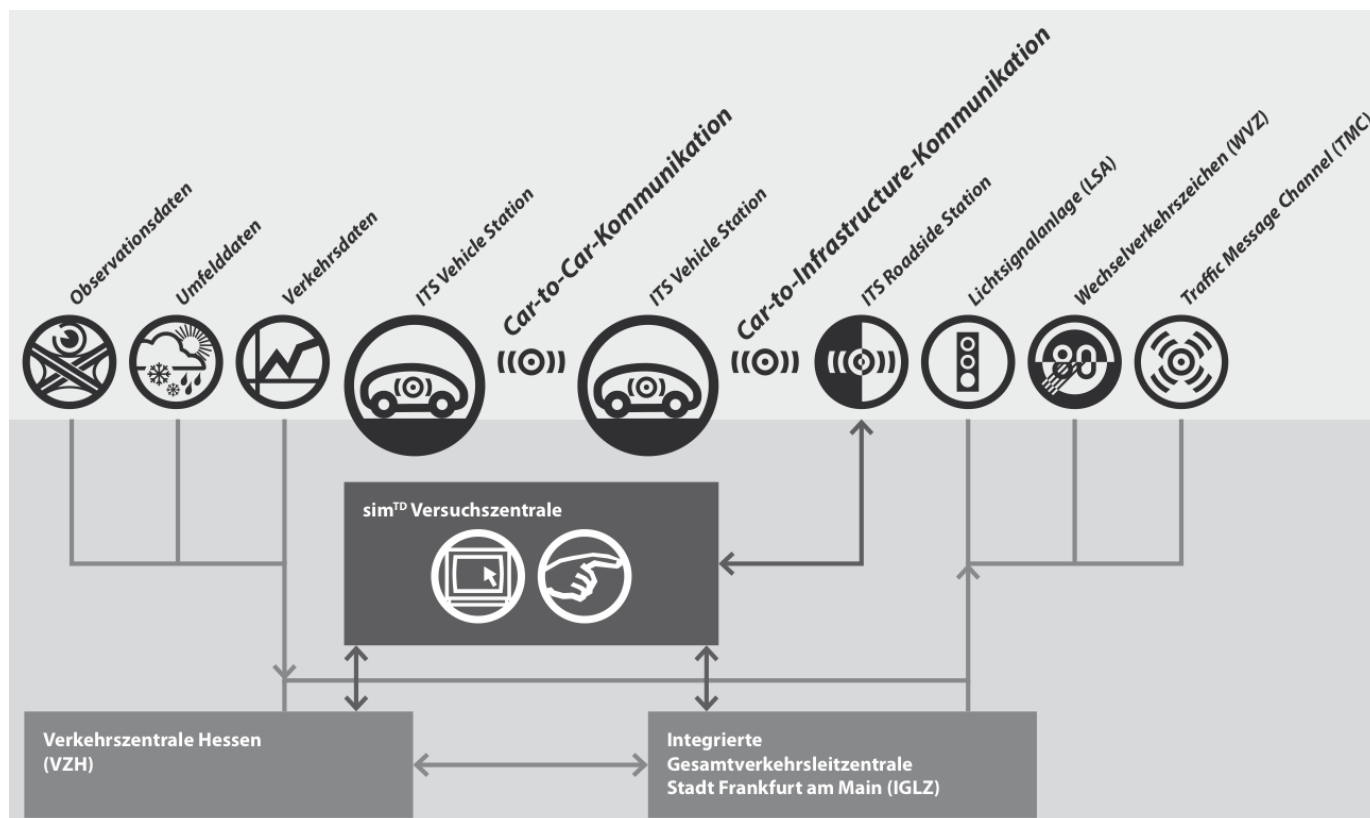


Abbildung 2. simTD Grobaufbau und Technologien [14]

Software ausgestatteten Fahrzeuge nicht ohne Kommunikationspartner da stehen. Des Weiteren sollen Mehrwertdienste, wie zum Beispiel Mautzahlungen, Anzeige von freien Parkplätzen und Parkplatzzahlungen über Funk, erforscht werden um Informationsgrundlage für Einführungsentscheidungen zu bilden.

III. ALTERNATIVE HCIS IN FAHRZEUGEN

A. Gestenbasierte Interaktionen in Fahrzeugen

Immer neuere Fahrzeugmodelle erhalten modernere Infotainmentsysteme, wie Radio, CD-Spieler, Telefon und Navigationssysteme, deren Funktionsumfang stetig steigt. Interaktionen mit dem Fahrzeug wird für den Fahrzeugführer somit eine immer komplexer werdende Aufgabe. Komfort und Nutzen im Fahrzeug steigen, aber dadurch wird der Fahrzeugführer auch von seiner Primäraufgabe, dem Fahren, abgelenkt [12]. Herkömmliche Interfaces für Infotainmentsysteme sind meistens so angelegt, dass der Fahrer sein Blickfeld für einen Moment auf dieses Interface richten muss, wenn er mit dem System interagieren möchte.

Grundlegende Voruntersuchungen [17] haben gezeigt, dass gestenbasierte Interaktionskonzepte als zusätzliche Modalität hinsichtlich bestimmter Aspekte gegenüber etablierten Eingabemethoden, wie Haptik und Sprache, Vorteile aufweisen. So lassen sich einige Eingaben durch gestische Bewegungen der Hand intuitiver und schneller durchführen. Außerdem konnte gezeigt werden, dass die Gestikinteraktionen mit geringeren Ablenkungseffekten verbunden sind.

In der Studie *GeCoM* [6] der Technischen Universität München wurde ein Bedienkonzept für Gestenbasierte Interaktionen mit Infotainmentsystemen in Fahrzeugen erstellt und erprobt um die Vor- und Nachteile dieser Eingabemethodik anhand einer realen Anwendung zu untersuchen. Dazu wurde als erstes in einer Vorstudie mit einer Reihe von Testpersonen typische Interaktionsszenarien im Fahrzeug und deren Eingabemöglichkeiten mit Handgesten identifiziert. Anhand dieser Ergebnisse wurde ein Gesteninventar entwickelt, das aus mimischen Gesten (zum Beispiel imitiertes Auf- und Abheben eines Telefons für Telefon Shortcut), deiktische² Gesten (zum Beispiel Zeigen in Richtung eines Displays entspricht Auswahl eines markierten Menüpunktes) und symbolischen Gesten (zum Beispiel Wischbewegung für den Abbruch einer Aktion). Um unbeabsichtigte Eingaben durch Gesten, die nicht für das System bestimmt sind sondern zum Beispiel durch Konversation mit Beifahrern entstehen, wird zwischen verschiedenen Handposen zur Aktivierung der Systeminteraktion unterschieden.

Anhand der vorausgegangenen Studie wurde auch ein grafisches User Interface implementiert, das durch die Handgesten gesteuert wird und im Fahrzeug auf einem Display angezeigt wird (Abbildung 4). Die Oberfläche stellt dabei alle wichtigen onboard Infotainmentanwendungen wie Radio, CD-Player und Navigationssystem zur Verfügung. Die Anwendungen wurden dabei horizontal durch Piktogramme dargestellt. Mögliche Interaktionen und Zustände werden dem Benutzer angezeigt, wobei die Displayinformation insgesamt auf ein Minimum

²<http://www.zeno.org/Meyers-1905/A/Deiktisch>

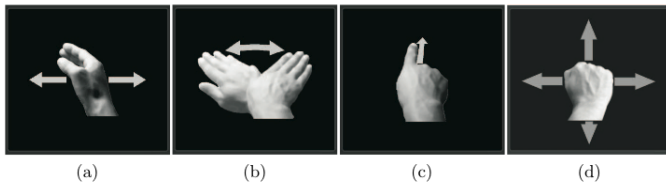


Abbildung 3. Beispiele aus dem Gesteninventar: (a) Wechsel zur nächste/vorherige Funktion (b) Wischbewegung für Abbruch einer Aktion (c) Zeigebewegung für Auswahl (d) Direkte Manipulation für z.B. Tonlautstärke [18]

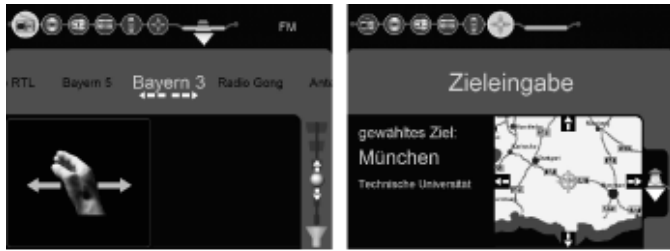


Abbildung 4. Grafische Ausgabe: Links als Radioanwendung mit aktiviertem Hilfscreen für das Durchschalten der Sender, rechts als Navigationsanwendung [18]

reduziert wurde, damit Systemzustände schnell erfasst werden können. Zusätzlich arbeitet das System mit akustischem Feedback bei Interaktionen um die Blindbedienbarkeit weiter zu fördern. Eine akustische Rückmeldung muss vor und nach einer Gesteneingabe erfolgen, da sonst üblicherweise Benutzer mit einem Kontrollblick auf das Display reagieren. Des Weiteren wurde ein audiovisuelles Hilfesystem implementiert, das dem Benutzer bei noch ungewohnten Gesteneingaben visuelle und sprachbasierte Hilfe leistet.

Für die Gestenerkennung wurde eine Kamera auf das Fahrzeugdach so ausgerichtet, damit Videoaufnahmen im Bereich der Mittelkonsole gemacht werden können. Verschiedene Filter und Algorithmen wurden aus der Computer Vision eingesetzt, sodass die Hand vom Hintergrund getrennt korrekt erkannt wird. Neben Position und Bewegungsrichtung müssen auch Handposen, wie eine Faust oder offene Hand, korrekt erfasst und interpretiert werden. In Gestenerkennungstests wurde die Wiedererkennungsrates der verschiedenen Handgesten getestet. Insgesamt erkannte das System sowohl Einzelgesten, als auch Sequenzen von einzelnen Gesten mit hoher Wahrscheinlichkeiten mit Durchschnittswerten von etwa 85%-98%.

Zur Evaluierung des gesamten GeCoM-Systems wurde abschließend eine Usabilityuntersuchung mit 25 Versuchspersonen verschiedenen Alters durchgeführt. Damit keine störenden Nebeneffekte, verursacht durch Erkennungsfehler, auftreten wurde der Test mit der Wizard-of-Oz Methodik durchgeführt. Zwei der Versuchspersonen waren in der Lage das Bedienkonzept ohne jegliche Hilfe und Anweisungen zu bedienen, die restlichen Probanden befanden das audiovisuelle Hilfesystem als ausreichend um das System bedienen zu lernen. Auch wurde das Audiofeedback positiv aufgenommen, sodass besonders im Autofahrt-Simulator häufige Blickabwendungen bei Bedienung der Infotainmentanwendungen kaum mehr auftraten.

B. Webbasiertes Infotainment im Fahrzeug: myCOMAND

Ein komplett Internetbasiertes Fahrzeug-Infotainmentsystem mit dem Namen *myCOMAND* [11] stellte Mercedes-Benz im Jahr 2008 vor, das mittels mobilen Internets Daten aus dem Web laden soll um somit Anwendungen immer aktuell halten zu können. Dabei soll sich das System bei jedem Motorstart selbst aktualisieren und Karten, Updates und andere Daten automatisch runterladen.

Folgende Anwendungen bietet *myCOMAND* unter anderem an:

- **Off-Board-Navigation:** Navigationssystem, das mit Kartenmaterial aus dem Internet arbeitet. Außerdem sollen zusätzliche Routeninformationen und Zielinformationen mit aktuellen Daten verknüpft werden können.
- **Trip Assist:** Der sogenannte Trip Assist stellt POI-Services zur Verfügung, die entlang der Route dargestellt werden können.
- **World Radio:** Radiostationen werden über das Internet gestreamt. Außerdem können Sender nach Genres durchsucht werden und eigene Musikbibliotheken erstellt werden.
- **Internet-Telefonie:** Telefonie via VoIP und Kurznachrichten über das Internet.
- **Internet-Browser:** Browser zum surfen im Internet.

Mittlerweile hat das *myCOMAND*-System seinen ersten Praxistest hinter sich gebracht. Auf einem Testgelände wurde dafür extra ein 4G LTE Mobilfunknetz errichtet, wodurch das Infotainmentsystem erstmals unter realen Bedingungen getestet werden konnte. Bisher war es nur möglich entweder über W-LAN oder 3G Mobilnetz *myCOMAND* zu benutzen. Mobilfunknetzanbieter und andere Experten gehen davon aus, dass solch hohe Bandbreiten über Mobilfunknetze, wie es LTE bietet, noch einige Zeit dauern wird und Systeme wie *myCOMAND* diese Bandbreiten für die Markteinführung voraussetzt.

IV. DRIVER MONITORING SYSTEMS

Das Forschungsfeld der aktiven und passiven Fahrsicherheitssysteme hat als so genanntes *driver's state adaptive driving support systems* ein Anwendungsgebiet hervorgebracht, das den aktuellen Zustand des Fahrers erkennen soll, während er das Fahrzeug lenkt. Motivation dafür sind die zahlreichen Unfälle, die durch Übermüdung des Fahrers verursacht werden und wie man dies durch Fahrsicherheitssysteme verhindern kann.

Es gibt zahlreiche Studien und Forschungsprojekte, überwiegend aus dem universitären Umfeld, die verschiedene Ansätze verfolgen, maschinelle Erkennung des Fahrerzustands zu verwirklichen. In den folgenden Abschnitten werden diese Projekte und ihre allgemeine Vorgehensweisen vorgestellt. Dabei ist schon im voraus zu erwähnen, dass die hier aufgelisteten Studien sich überwiegend mit der Erfassung von Körpermerkmalen und daraus schließenden Zuständen auseinandersetzen, aber eine konkrete Implementierung eines Fahrsicherheitssystem noch aussteht.

Die in den folgenden Abschnitten vorgestellten Projekte bilden bis jetzt nur die Basis für ein konkretes Monitoring-System, die sich mit der Erkennung beschäftigen, aber daraus folgende

Reaktionen nicht behandeln. Wie und wann das System in das Fahrgeschehen bei erkannter Übermüdung eingreift, zum Beispiel durch Warnmeldungen oder aktive Übernahme der Fahrerkontrolle, soll in fortführenden Studien behandelt werden.

A. Augenbasierte Erkennung

Projekte wie [8] und [4] erheben Zustand und Veränderungen der Augen als zentrales, physiologisches Merkmal, da die Augen eine offensichtliche Rolle bei Müdigkeitserscheinung spielen.

Im Paper [8] wird ein Erkennungsalgorithmus vorgestellt, der auf Umrissform der Augenlieder und der so genannten *Fuzzy Logic* basiert. Mit einer Kamera werden Aufnahmen vom Auge erstellt, die als Basismaterial für die CV-Algorithmen dient. Als erstes muss das Auge aus den Bildern korrekt und exakt segmentiert werden damit in späteren Schritten eine Analyse stattfinden kann (Verarbeitungsschritte siehe Abbildung 5). Dazu wird das Bild durch lokale und globale Schwellenwerte des Verfahren von Otsu segmentiert. Anschließend wird ein neuer Schwellenwert bestimmt, der unter dem des Otsus liegt, und nochmals segmentiert. Dann wird im zweiten Schritt die Bogenlinie des oberen Lids anhand des segmentierten Bildes bestimmt um damit die elliptische Form des geöffneten Augensegments zu errechnen. Des Weiteren wird die Position der Pupille erfasst und der Anteil des sichtbaren Bereichs der Pupille berechnet. Die Ellipsenform und der Anteil der sichtbaren Pupille werden auf Werte quantifiziert, da diese als Eingabewerte der nachfolgenden Fuzzy Logic dienen. Anhand der zwei Eingabewerte und einer einfachen Regelbasis kann dann der Zustand des Auges als Ausgabewert bestimmt werden. Ein einfaches Beispiel der Regelbasis könnte so aussehen: *If* erster Eingabewert ist x *And* zweiter Eingabewert ist y *Then* Zustand des Auge entspricht *wach*. Auf diese Weise wird der Zustand des aufgenommenen Fahrers identifiziert.

In anschließenden Tests wurden 1483 Frames nach dieser

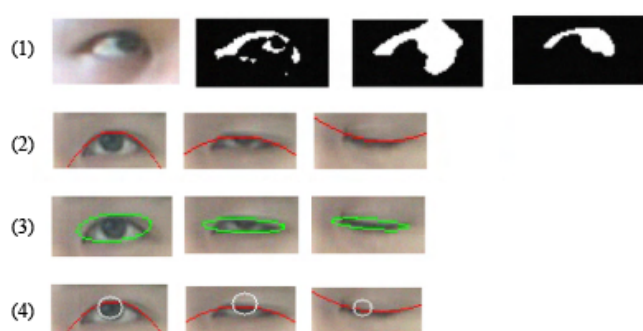


Abbildung 5. Erfassung der digitalen Aufnahmen: (1) Segmentierung des Auges (2) Erfassung der Bogenlinie des oberen Augenlids (3) Berechnung der elliptischen Form des Auges (4) Erfassung der Pupille [8]

Methode untersucht. 1409 Frames davon wurden korrekt gedeutet, was eine Rate von 95,01% entspricht.

Allgemeines Problem bei den bildgestützten Erkennungssystemen ist die Zunahme der Fehlerrate bei schlechten Lichtverhältnissen, wie zum Beispiel bei Nachtfahrten. Deswegen

wird in [4] ein Verfahren vorgestellt, das neben der Bilderfassung mit einem Infrarotbeleuchter arbeitet. Die Segmentierung der Bilder erfolgt ähnlich wie in [8] mit dem Schwellenwertverfahren von Otsu. Durch die Infrarotbeleuchtung können aber auch noch bei sehr dunkler Umgebung die Pupillen schnell und korrekt erfasst werden, wodurch das System robuster gegenüber verschiedenen Lichtverhältnissen wird.

B. Kopfpositions-basierte Erkennung

Weitere Indikatoren für Müdigkeitserscheinung am Steuer können bestimmte Kopfpositionen und -bewegungen sein. In [2] wird ein Modell für die Berechnung der Fahrerkopfbewegung mittels Kameraüberwachung der Kopfposition und Bayes'sche Netze vorgestellt. Als Referenzpose wird ein Bild von der Kamera aufgenommen und als Zustand *wach* initiiert. Die Position des Gesichts wird in einem Koordinaten-Modell abgebildet und Veränderungen der Kopfposition gegenüber dem Referenzbild werden in Winkelgrad der X- und Y-Achse gemessen. Mittels des Bayes'schen Netzes und eingespeisten Anfangswerten soll ein quantifizierter Wert bestimmt werden, der den Grad der Müdigkeit repräsentiert.

In abschließenden Testreihen wurden auch Aufnahmen mit einer Infrarotkamera gemacht um das Erkennungssystem auch geeignet für Nachtfahrten zu gestalten.

C. Mundbasierte Erkennung

In der Studie [9] werden die Charakteristiken der Mundregion als Indikatoren für den gegenwärtigen Zustand des Fahrers herangezogen. Zum einen kann anhand des Mundes erkannt werden, ob und wie müde der Fahrer ist und zum anderen soll die Ablenkung des Fahrers durch langes Reden bestimmt werden.

Bei der Erfassung der Gesichtssegmente wird hier mit einem *Color Space Model* gearbeitet, mit dem anhand von Farbräumen Kopf- und Mundpartie segmentiert wird. Die einzelnen Schritte laufen dabei wie folgt ab: Als erstes wird das Gesicht vom Hintergrund segmentiert und lokalisiert. Danach wird die Mundregion herausgefiltert und auf deren Charakteristiken, wie zum Beispiel Größe und Form der Mundöffnung, Grad der Symmetrie und sich wiederholende Muster, hin untersucht. Die Studie befasste sich nur mit der Erfassung und Erkennung der beschriebenen Bildsegmente, zieht aber aus den identifizierten Ergebnisse keine Schlüsse, wie zum Beispiel sich oft wiederholende Mundöffnungen deuten auf Müdigkeit des Fahrers hin. Dies wird auf zukünftige Untersuchungen adressiert, mit der sich die Gruppe beschäftigen will, sowie geeigneten Reaktionsmaßnahmen für das Fahrzeugsicherheitssystem.

V. AUTONOME FAHRZEUGE

Das Forschungsgebiet des *Autonomous Driving/Intelligent Vehicle Concept* geht gegenüber den intelligenten Fahrassistenten- und Sicherheitssystemen einen Schritt weiter und hat die Vision, Fahrzeuge autonom auf Straßen fahren zu lassen. Einzelne Fahrzeuge sollen dabei auch in der Lage sein, sowohl kooperativ mit anderen intelligenten Fahrzeugen, als auch mit von Menschen gesteuerten Autos auf der Straße zu koexistieren.

In den folgenden Abschnitten werden Forschungsprojekte vorgestellt, die autonome und unbemannte Fahrzeugsysteme erfolgreich umgesetzt haben.

A. Intelligent Control Systems Laboratory

Das Intelligent Control Systems Laboratory (ICSL) der Griffith University in Australien hat sich als Schwerpunkt gesetzt, ein kooperatives, intelligentes Fahrzeugsystem zu entwickeln, dass in der Lage ist sowohl auf Landstraßen, als auch in Stadtgebieten zu bestehen [1]. Dabei soll das Fahrzeug nicht auf neue Technik, die zum Beispiel zur Kommunikation am Straßenrand platziert wird, angewiesen sein, sondern durch fahrzeuginterne Sensoren mit der vorhandenen Infrastruktur der Landstraße und der Stadt zurechtkommen.

Im Dezember 2002 konnte das System zum ersten mal erfolgreich in einer Demonstration getestet werden. Demonstrationspartner waren die staatliche, französische Forschungseinrichtung INRIA und dessen Industriepartner ROBOSOFT, die drei experimentelle, computergesteuerte Fahrzeuge zur Verfügung stellten. Alle Fahrzeugmanöver fanden kooperativ mit den anderen Fahrzeugen und ohne menschliche Hilfe statt. Die Demonstration enthielt unter anderem folgende Manöver:

- Eine Überquerungsmanöver an einer Kreuzung mit mehreren Fahrzeugen
- Ein kooperatives Überholmanöver
- Ein Manöver, dass die Fahrzeuge hintereinander mit einem bestimmten Abstand und Fahrspurkontrolle fahren lässt
- Einfaches Fahren und Halten der Fahrspur

Beim Fahrmanöver an der Kreuzung war keine externe Signalisierung nötig und alle drei Fahrzeuge überquerten die Kreuzung nacheinander ohne Kollision. Die Bildsequenz in Abbildung 6 zeigt das Überholmanöver, in dem das hintere Fahrzeug zum überholen auf die andere Spur wechselt, das andere Fahrzeug überholt und anschließend wieder in die alte Spur zurück wechselt. Das zu überholende Fahrzeug verringert dabei seine Geschwindigkeit wenn es merkt, dass es überholt wird und beschleunigt anschließend wieder auf die ursprüngliche Geschwindigkeit.

Technische aufgebaut ist das System aus zwei verschiedenen Komponenten: Zum einen aus den experimentellen Fahrzeugen der INRIA und der ICSL Hard- und Software. Verbunden sind die beiden Systemkomponenten über einen implementierten CAN-Bus-Schnittstelle um Kommunikation zwischen beiden Seiten zu ermöglichen. Die Entscheidungs- und Kontrollalgorithmen, die auf der ICSL-Hardware implementiert sind, senden die Werte an den Root-Prozessor im Fahrzeug, der die korrekte Geschwindigkeit und Lenkwinkel umsetzt.

Die ICSL Hardware besteht aus einer Anzahl an Modulen, wobei jedes Modul, angesteuert durch je einen Mikrokontroller, eine Hauptfunktion repräsentiert (siehe Abbildung 7). Diese umfassen:

- **Entscheidungs- und Kontroll-Modul:** Bildet das zentrale Befehlsmodul, dass durch Input der anderen Module Entscheidungen über die Fahrmanöver trifft und als Befehl weitergibt.
- **Modul für die Motorsteuerung:** Ansteuerung der Fahrzeugmotoren.

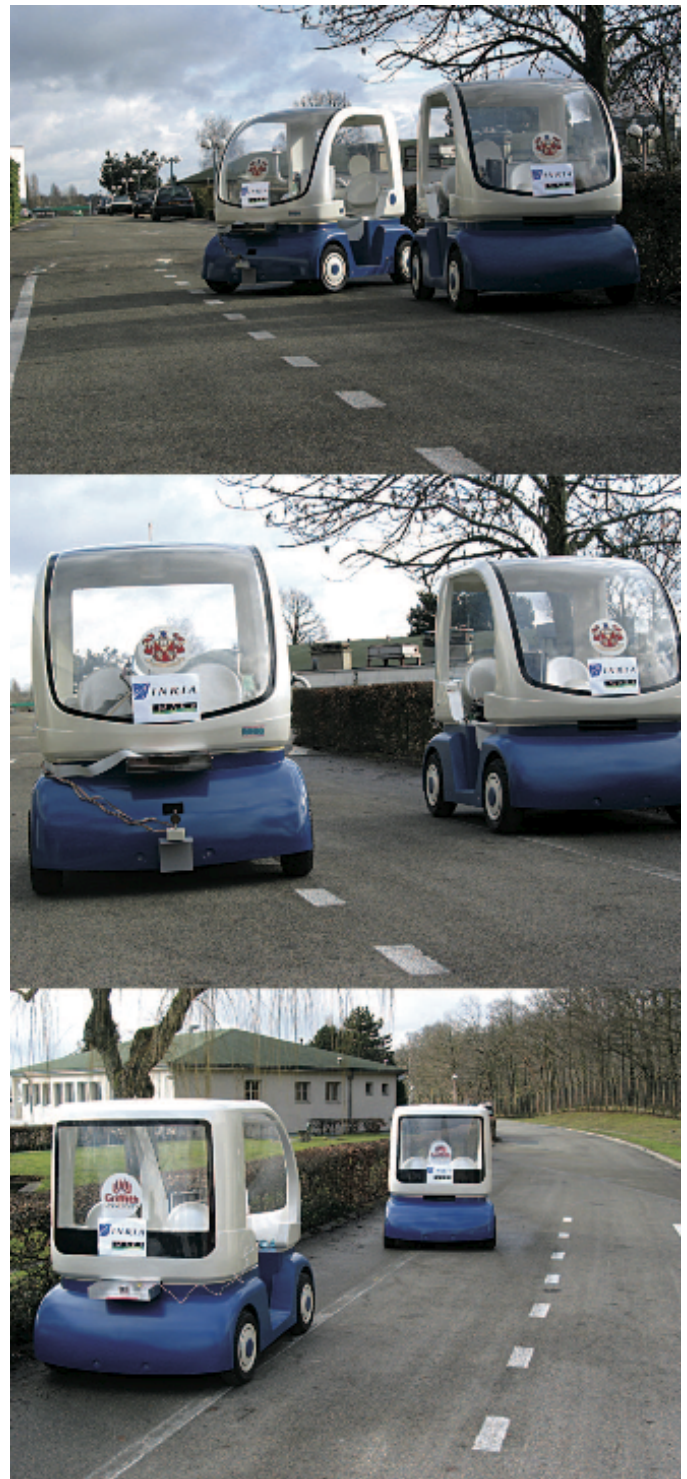


Abbildung 6. Kooperatives Überholmanöver bei der ICSL Demonstration

- **Sensorinterface-Modul:** Kontroll- und Überwachungs-Modul der verschiedenen Sensoren (wie zum Beispiel Infrarot- und Ultraschallsensoren), die Objekte wie andere Fahrzeuge, Hindernisse und andere sowohl statische als auch dynamische Objekte erkennen sollen.
- **Modul für Funkkommunikation:** Für die drahtlose Kommunikation.
- **Modul für die Energieversorgung:** Steuerungs-Modul

für die Energieversorgung der Sensoren.

- **Distanz- und Erfassungs-Modul:** Kontroll-Modul für die Erfassung von bestimmten Distanzen zwischen zwei Fahrzeugen. Dabei muss das hintere Fahrzeug immer einen Mindestabstand zum vorderen Fahrzeug erkennen und einhalten.

Jeder Reifen der INRIA Fahrzeuge besitzt eine bestimmte Anzahl an Sensoren, die durch zwei Mikrokontrollern angesteuert werden. Ein eingebetteter x86-Prozessor, der Root-Prozessor, koordiniert die zwei Controller über einen CAN-Bus.

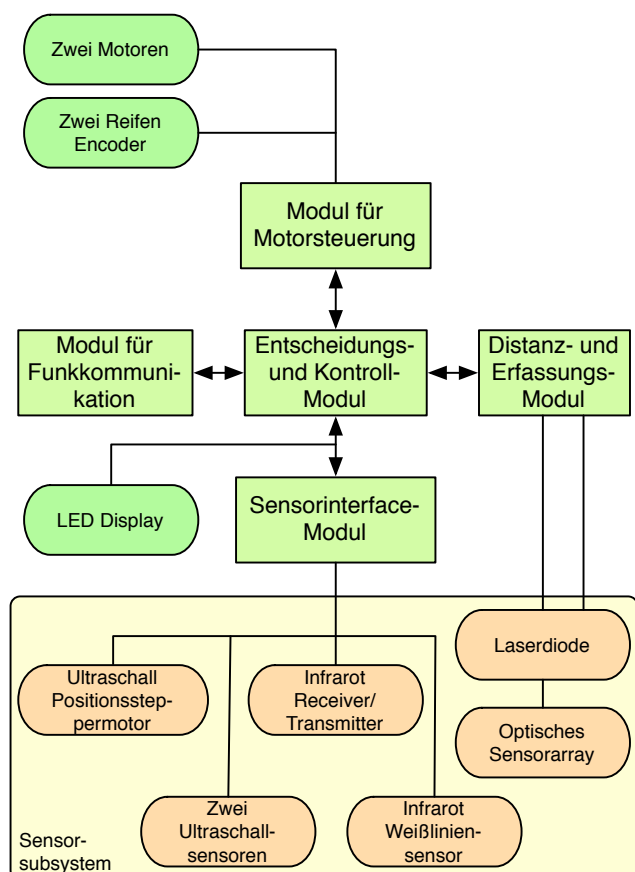


Abbildung 7. ICSL Architektur der mobilen Fahrzeuge

B. DARPA Grand Challenge

Die DARPA Grand Challenge ist ein von der *Defense Advanced Research Projects Agency* des US-Verteidigungsministeriums initiierteter Wettbewerb für autonome, intelligente Fahrzeuge [3]. Der Wettbewerb ist mit einer hohen Geldsumme ausgeschrieben und soll die Erforschung und Entwicklung von autonomen fahrenden Fahrzeugen vorantreiben. Ziel ist es, dass das entwickelte Fahrzeug selbständig eine bestimmte Strecke innerhalb einer Maximalzeit absolviert. Erstmals wurde die DARPA Grand Challenge im Jahr 2004 ausgetragen, bei der die Teilnehmerwagen eine Strecke von 214 km durch die Mojave-Wüste innerhalb von 10 Stunden absolvieren sollten. Keines der gestarteten Teams gelang es, die Strecke vollständig zu absolvieren. 2005 gelang es

dem Team der Stanford University mit ihrem Fahrzeug in der zweiten Grand Challenge als erstes, die vorgeschriebene Strecke zu bewältigen. 2007 fand der Wettbewerb, diesmal unter dem Namen *Urban Challenge*, erstmals in einer innerstädtischen Umgebung statt. Als Parcours diente ein verlassener Kasernenstützpunkt des US-Militärs. Neben der selbständigen Navigation durch die Stadt mussten die Fahrzeuge zusätzlich bestimmte Vorfahrtsregeln beachten, da neben den Fahrzeugen der Teilnehmer auch noch Fahrzeuge, die von Menschen gesteuert wurden, am Straßenverkehr teilnahmen. Die Fahrzeuge mussten eine bestimmte Strecke absolvieren, sowie bestimmte Missionen im Stadtverkehr ausführen mit so wenigen Verkehrsregelverstößen wie möglich. Sieger dieser Challenge wurde das Team der Carnegie Mellon University in Pittsburgh.

VI. FAZIT

Die in dieser Seminararbeit vorgestellten Projekte lassen einen Einblick gewähren, wie das Auto von Morgen aussehen könnte. Dabei ist es klar, dass sich einige Projekte noch in weiter Zukunft befinden und bis zur Serienproduktion noch einige Zeit vergehen wird. Akzeptanz der Benutzer, aber auch technologische Voraussetzungen spielen neben der Ausgereiftheit der Technik hierbei eine entscheidende Rolle. Mit hoher Wahrscheinlichkeit werden C2X-Anwendungen in Zukunft auf den Straßen zu finden sein, da hier bereits intensiv geforscht wurde um technologische Grundlagen und neue Standards zu schaffen. Konsortien haben sich hierfür gebildet, um Anwendungen und Technik zu definieren und zu testen. Durch C2X würde somit sowohl Verkehrssicherheit, als auch Effizienz im Straßenverkehr und Komfort für den Fahrer, erhöht werden. Durch das simTD-Projekt soll C2X erstmals in einem Großaufbau mit Fahrzeugen und Infrastruktur als Gesamtkonzept getestet werden.

Alternative HCIs versuchen dagegen durch verschiedene Interaktionskonzepte zwischen Fahrer und Fahrzeug Sicherheit und Komfort zu erhöhen. Während das myCommand-System kurz vor der Serienreife steht, ist es bei den gestenbasierten Interfaces noch fraglich, in wiefern sie es in die Serienproduktion schaffen. Usabilitytests haben aber bereits gezeigt, dass die Ablenkung bei der Bedienung des Interfaces während des Fahrens deutlich minimiert wird.

Der Forschungsstand der Driver Monitoring Systeme steht momentan noch relativ am Anfang, da bis jetzt nur Grundlagen für die Erkennung von physikalischen Merkmalen der Fahrer geschaffen wurde, aber weitere Schritte noch ausstehen. Wie das System auf die veränderten Müdigkeitserscheinungen reagieren soll, wird unter anderem Teil der fortführenden Arbeit sein.

Autonome Fahrzeuge haben bereits im Militär ihre Anwendung, im zivilen Bereich werden autonom fahrende Autos im Straßenverkehr jedoch noch länger Vision bleiben. Aktuelle Fahrzeug-Projekte zeigen bereits, dass diese sowohl kooperativ mit anderen autonomen Fahrzeugen als auch mit bemannten Fahrzeugen im Straßenverkehr zurecht kommen. Doch fehlen hier noch aussagekräftige Teststudien, die beweisen, dass die autonomen Fahrzeuge auch im realen Straßenverkehr bestehen

können.

Insgesamt sieht man, dass die heute modernen Fahrzeuge ihr Potenzial noch lange nicht ausgeschöpft haben und in Zukunft immer sicherer werden, sowie das Fahren und damit verbundene Tätigkeiten immer einfacher und komfortabler werden.

LITERATUR

- [1] J. Baber, J. Kolodko, T. Noel, M. Parent, and L. Vlacic, "Cooperative autonomous driving: intelligent vehicles sharing city roads," *IEEE Robotics and Automation Magazine*, vol. 12, no. 1, pp. 44–49, 2005.
- [2] M. Celenk, H. Eren, and M. Poyraz, "Prediction of driver head movement via bayesian learning and arma modeling," in *IEEE Intelligent Vehicles Symposium*, 2009.
- [3] DARPA. Darpa grand challenge. [Online]. Available: <http://www.darpa.mil/grandchallenge/index.asp>
- [4] W. Di, R. Wang, P. Ge, and F. Cao, "Driver eye feature extraction based on infrared illuminator," in *IEEE Intelligent Vehicles Symposium*, 2009.
- [5] W. Franz, "Car-to-car communication - anwendungen und aktuelle forschungsprogramme in europa, usa und japan," in *VDE-Kongress Fachtagungsbericht*, 2004.
- [6] M. Geiger, R. Nieschulz, M. Zobl, and M. Lang, "Bedienkonzept zur gestenbasierten interaktion mit geräten im automobil," in *VDI-Bericht 1678 ÜSEWARE2002*, 2002.
- [7] R. German and K.-S. J. Hielscher, "Fahrzeugkommunikation - überblick intra-/inter-fahrzeug-kommunikation," Slides.
- [8] J. Guo and X. Guo, "Eye state recognition on shape analysis and fuzzy logic," in *IEEE Intelligent Vehicles Symposium*, 2009.
- [9] L.-S. Jin, X. Sun, Y.-Y. Jiang, H.-J. Hou, and Y.-Q. Sun, "Study on driver's mouth segmentation and location based on color space," in *IEEE Intelligent Vehicles Symposium*, 2009.
- [10] A. Lübke, "Car-to-car communication - technologische herausforderung," in *VDE-Kongress Fachtagungsbericht*, 2004.
- [11] Mercedes-Daimler. Erster praxistest für forschungsprojekt mycomand: Infotainment ohne tempolimit: Wireless high-speed internet für das auto der zukunft. [Online]. Available: <http://media.daimler.com/dcmedia/0-921-656548-49-1243763-1-0-0-0-0-0-11700-0-0-1-0-0-0-0-0.html>
- [12] C. A. Pickering, K. J. Burnham, and M. J. Richardson, "A research study of hand gesture recognition technologies and applications for human vehicle interaction," in *3rd Institution of Engineering and Technology Conference on Automotive Electronics*, 2007.
- [13] simTD Konsortium. simtd - sichere intelligente mobilität testfeld deutschland. [Online]. Available: <http://www.simtd.org>
- [14] —. simtd technologie. [Online]. Available: http://www.simtd.org/dyn/mediaout.dhtml/264b266ea2204f57276g/mime/PDFs_Grafiken/technologie/simTD_technologie02.pdf
- [15] T. Steudten. (2007, 09) Car 2 car: Die vision kommunizierender fahrzeuge. [Online]. Available: http://www.tecchannel.de/test_technik/1728877/car_2_car_die_vision_kommunizierender_fahrzeuge/
- [16] unknown. (2009, 10) Ieee 802.11p. [Online]. Available: <http://www.itwissen.info/definition/lexikon/802-11p-IEEE-802-11p.html>
- [17] M. Zobl, M. Geiger, K. Bengler, and M. Lang, "A usability study on hand gesture controlled operation of in-car devices," in *Proc. HCI 2001*, 2001.
- [18] M. Zobl, R. Nieschulz, M. Geiger, M. Lang, and G. Rigoll, "Gesture components for natural interaction with in-car devices," in *Proc. of 5th International Gesture Workshop*, 2003.

The Crowdsourcing Phenomenon

Stefan Kaufmann – stk@uni-ulm.de

Abstract—Crowdsourcing is a relatively new, yet widely used term to describe corporations outsourcing specific problems to the general public. The purpose of this paper is to give an overview of the creation of the term and its precise definition, and the reasons why outsourcing to a crowd can be a sound method. Furthermore, different classes of algorithms based on human contribution, the differentiation between crowdsourcing and open source and criticism of the concept will be reflected on. In order to achieve this goal, various publications and related work were analyzed and contributed to this paper. Crowdsourcing was found to be a sensible method to efficiently find solutions to given problems involving material goods from a corporate point of view, while contributors rarely benefit as much from their participation. Furthermore, it was found that the term is often applied to algorithms not meeting the original definition, leading to a dilution of the term.

I. INTRODUCTION

CROWDSOURCING, a neologism coined in 2006, is one of the defining buzzwords of the past few years. A multitude of businesses have been trying to outsource decisions to *the crowd*—with varying success, ranging from multi-million-dollar revenues [21] to complete failures. In the following paper, a general overview of computer-based methods that rely on crowd participation will be given, before taking a look at what defines crowdsourcing in particular, addressing the differences between crowdsourcing and open source, and reflecting on criticism of the concept.

II. HOW SMART MASSES CAN DEFEAT SUPERCOMPUTERS

Until the 1940s, the term *computer* generally referred to a person whose profession it was to perform complex mathematical calculations, often working in parallel with many other *human computers*. Only when the existing mechanical computing devices were superseded by electromechanical and electronic *programmable* computers—capable of performing several hundred multiplication operations per second [35]—human computers were relieved of their tedious task.

While able to surpass every single human being as far as computing power is concerned, today's computers still lack typical characteristics of human thought processes. As of yet, Computers are unable to perform acts of creativity or sensible problem-solving, and while computer vision technology is constantly improving, the complete magnitude of human visual cognition is still unmatched by any machine. Therefore, as early as 1986 [13], efforts began to use informed decisions of humans to supplement computer algorithms where they met their limits.

To increase the reliability of this process, other researchers began to use the decisions of several people instead of only one person [30][10]. One might wonder why averaging the input of many participants should *improve* the overall result. After



Figure 1. Participants collectively playing “Pong” in an experiment by Loren and Rachel Carpenter at the 1994 Ars Electronica. Source: Vimeo [34]

all, averaging, for instance, the time needed to perform the 100 metres by a group of people will inevitably result in a mediocre time. Nonetheless, the idea of using the input of a large group of participants in order to yield better results is neither new nor odd. Crowdsourcing evangelists often cite Francis Galton, who found that when a crowd of nearly 800 people tried to guess the weight of an ox, even the best guesses were surpassed by the arithmetic mean of *all* the guesses—the latter being only one pound off the actual weight of 1198 lbs [15].

In “Out of Control” [25], Kelly describes a series of experiments by Loren and Rachel Carpenter, in which crowds were even able to play *Pong* and steer a (simulated) aircraft collectively. Paddles with reflective red and green ends were given to the 1991 SIGGRAPH audience, half of which controlled one *Pong* game paddle each. The ratio of red and green paddle ends for each side was recognized through video cameras connected to a computer, and determined the position of each game paddle. Simply speaking, the more players show the green side of their paddle, the higher the game paddle rises, and vice versa. A video snippet of this experiment being repeated at the 1994 Ars Electronica Festival [14] is available at Vimeo [34]. Kelly specifically remarks upon the speed with which the 5000 SIGGRAPH players adapt every time Carpenter increases the game speed, hardly missing any balls.

Kelly emphasizes the importance of such a crowd being as heterogenous as possible to perform in a robust way, even in challenging situations:

“In economic, ecological, evolutionary, and institutional models, a healthy fringe speeds adaptation, increases resilience, and is almost always the source of innovations.”

He also points out the notion of individuals representing single computers, working together to form a parallel computer,

which is inherently more powerful than regular, programmable—serial—computers due to its architecture:

“Parallelism is one of the ways around the inherent stupidity and blindness of random mutations. It is the great irony of life that a mindless act repeated in sequence can only lead to greater depths of absurdity, while a mindless act performed in parallel by a swarm of individuals can, under the proper conditions, lead to all that we find interesting.”

The necessity of a diverse group as well as the independence of the individuals’ decisions and the decentralization of the crowd are also pointed out by Surowiecki in 2004 [31]. Equally important, according to him, is the possibility to aggregate and combine all the individual contribution in a sensible way to literally “give meaning” to the data. The Internet is a very empowering medium in this context, allowing a dispersed crowd located at even the remotest locations on earth to collaborate as one superorganism—or, in other words, as individual parts of a giant, distributed human computer.

III. DISTRIBUTED HUMAN COMPUTING: A TAXONOMY

Defining the crowd to be a giant computer brings up a new question. If individuals can be a computer, of what kind might be the *programs* one can execute on it? Are there different types of algorithms incorporating networked individuals—and if yes, what are their distinguishing marks? In their work, Quinn and Bederson propose eight distinct genres to classify what they call *Distributed Human Computing (DHC)* systems—that is, systems allowing the incorporation of crowds into problem-solving efforts [27].

Games With a Purpose entice players to perform calculations in order to achieve high scores, relying on a fun factor derived from playing the game, namely, the accumulation of points. A typical example is Google’s Image Labeler [16], which makes two randomly chosen players name prominent features of a given picture during a limited timespan. Any description that has been named by both participants results in points for both players. This *forced agreement* method is not only a characteristic of the game, but at the same time assures the quality of the image tags obtained thereby.

Mechanized Labor is a way of delegating small tasks to paid contributors, most prominently showcased by Amazon’s Mechanical Turk website [2], which allows complex tasks to be broken up into small *Human Interaction Tasks (HIT)* in order to be executed by human contributors. The incentive to participate is exclusively based on monetary compensation.

Dual-Purpose Work utilizes tasks that people are already doing to harness the crowd. The ReCAPTCHA [28] system, for instance, asks users to transcribe two words. One of them is part of a regular CAPTCHA, the other one a snippet of a scanned book or newspaper that could not be recognized through OCR software.

Grand Search uses volunteers to search through large sample spaces, usually images, to find one individual outstanding sample that was being looked for. One notable example is amateurs reviewing over 560,000 images of ocean surface in the search for computer scientist Jim Gray after he went

missing at sea. Participation usually takes place due to altruistic motives.

Human-based Genetic Algorithms lets humans first contribute solutions for a given problem, after which other human participants select the best solutions, combining and mutating them. This method is based on the notion that, for instance, the text of a Wikipedia entry or the source code of an application is considered to be a genetic sequence, with edits being mutations that decide which parts of the sequence are altered or stay the same.

Another approach is **Knowledge Collection from Volunteer Contributors (KCVC)**. It tries to establish large knowledge bases through human contributors in order to build a database of “common sense” aiming to advance artificial intelligence research. One example is the now defunct “1001 Paraphrases” game construed by Chklovski [11], which was aimed to collect a database of paraphrases for a variety of given words. This approach could, however, also be classified as a Game with Purpose or even Dual-Purpose Work.

Harnessing the **Wisdom of Crowds** in Quinn’s and Bederson’s sense means inviting users to make guesses or state their opinion about something that needs to be quantified or predicted. Theoretically, the more users participate and the more diverse the participating crowd is, the results should increase in quality, even to the point of estimating election results in advance.

Last, but not least, Quinn and Bederson identify **Crowdsourcing** as a separate category, in which work is outsourced to unpaid volunteers. Since this classification of crowdsourcing being merely one subset of all possible DHC algorithms does not exactly match the definition of Howe, this special category will be reflected on in the following section.

IV. SPECIAL CASE: CROWDSOURCING

The term *crowdsourcing* was first coined by Jeff Howe in his 2006 *Wired* article [21], in which he analyzed businesses such as iStockphoto.com, TV shows based on amateur internet videos, and R&D facilities challenging hobby tinkerers and professional scientists to solve problems. All these businesses rely on the contributions of volunteers, who shoot images, upload their movies to the internet, and spend time thinking about how to get fluorite powder into toothpaste tubes without creating a mess. And, first and foremost, all those businesses are commercially oriented, creating revenue and profit with their input. Howe does not discriminate whether the products and ideas are produced collaboratively or by single individuals [19]. His only definition is the outsourcing of tasks by corporations or institutions to the crowd:

“Simply defined, crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. [...]”

I interpret crowdsourcing to be taking place any time a company makes a choice to employ the crowd to perform labor that could alternatively be performed by an assigned group of employees or

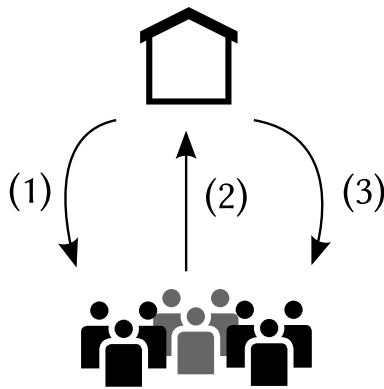


Figure 2. Basic crowdsourcing workflow: A corporation proposes a problem to the general public (1). After one or more crowd generated solutions are selected (2), the corporation manufactures products based on these solutions and sells them back to (not necessarily the same) public (3).

contractors, even if the company is just now putting up a shingle. In other words, crowdsourcing need not require an active shift from current employees (or again, contractors) to the crowd; it can start with the crowd.”

Interestingly, while Quinn and Bederson draw a distinction between mechanized labor and crowdsourcing, the Amazon Mechanical Turk [2] fits perfectly into Howe’s criteria as stated above, and is also explicitly mentioned by him later in 2006 [20]. This leads to the conclusion that Crowdsourcing is the *application of any* Distributed Human Computing (DHC) approach by a company or institution rather than a completely distinct DHC algorithm all by itself.

A. Motivation for Participants

An often cited paradigm case for crowdsourcing is Chicago-based Threadless T-Shirts. Threadless’s goal is to sell t-shirts with original, innovative and attractive designs, acquiring these designs through contests held on their website. Anybody with a valid email address can sign up with a user account and upload their original t-shirt designs, which then enter into a scoring process. Threadless’s user community assigns grades on a zero-to-five scale to those designs, optionally checking an “I’d buy it!” box. The highest scoring shirts go into print and are sold through the Threadless website, typically within the 10–15 USD range, which is very affordable for designer shirts. Users whose designs were chosen to go to print, receive 1500 USD in cash and another 500 USD to spend in the Threadless online shop.

Using this strategy, Threadless was “selling 60,000 T-Shirts a month, [had] a profit margin of 35 percent and [was] on track to gross \$18 million in 2006. This, for a company with fewer than 20 employees.” [17]

One might ask why users are willing to be content with only 2000 USD for a design they know Threadless will make way more money off. Brabham explored the motivation of Threadless’s designers in 2009 [6] and found monetary

compensation one strong, but not the sole motivator for participation. Several participants in his interviews also noted the possibility to improve their own creative skills by receiving feedback and acting upon it to improve their future designs, even leading to the opportunity to take up related freelance work professionally. However, most contributors apparently regard their own work more as a hobby than a profession.

A significant rate of the interviewees explicitly noted *The Love of Community*, i.e., the communal aspect and the personal relationships to other users as an incentive to their continued affiliation with Threadless. One of the interview partners compared this to the brand loyalty displayed by Apple Computer owners:

“i think what they [Threadless] do right is promoting the community aspect of the ‘brand’

the comparison i can give is the way owning a macbook creates a bond between people. it makes you more loyal because you feel like you’re part of something.

except that threadless has a soul”

Brabham notes that many of his interview partners used language relating to addiction when describing their relationship to the Threadless community, pointing out that this addiction is a prominent trait in online communities rather than real-life organisations or clubs.

B. Microstock Photo Agencies

Another often cited case is iStockphoto.com [24], which was the first so-called “microstock photo” agency, founded in 2000. Traditional stock photo agencies, such as Corbis [12] or Getty Images [23], cater their archive of images shot by professional photographers to news agencies or corporate clients. If, for instance, the creative department of an advertising company needs a photo of a woman in business outfit, shot against a white background, they can find a variety of matching images and license them from those agencies, starting at prices of about 100 USD and upwards.

At the time of its creation, iStockphoto’s approach was radically different to that of its old-established competitors. Whereas traditional stock photo agencies only take contributions from established, professional photographers, iStockphoto is open to pretty much everybody able to operate a camera. Any aspiring stock photographer can fill out their application form, provide a proof of identity and upload three reference images. If the photographs fulfill iStockphoto’s technical requirements regarding sharpness, image quality and color, the applicant is admitted and may upload their work to be stored in iStockphoto’s database, tagged with matching keywords. All the photos are offered for unlimited use to an exceptionally small price, starting at as little as 0.20 USD, with even the most expensive licenses being in the 20 USD range. Photographers receive 20 percent of the purchase price each time their work is downloaded, and if they choose to upload their work exclusively to iStockphoto, their share may rise to 40 percent [7].

Since the foundation of iStockphoto, several other microstock agencies have entered the playing field, increasing pressure on the former market leaders, who found themselves increasingly

unable to compete with those agencies' pricing models. Facing the growing competition of stock photography being sold at less than one one-hundredth of their usual licensing fee, Getty Images eventually acquired iStockphoto in February 2006 for 50 million USD. Howe quotes Getty co-founder Jonathan Klein [21]:

"If someone's going to cannibalize your business, better it be one of your other businesses."

The motives for participation vary between users. While a majority of users cite the opportunity to make money as a reason, the improvement of one's creative skills and the sheer fun of it also range high, according to Brabham [8]. Interestingly, compared to Threadless's participants, iStockphoto users appear to be significantly less keen on community participation.

In order to make a living from microstock agencies, the comparatively minute revenues per sold image require participants to upload significantly more photos than they would have to if they used traditional stock photo agencies. One *New York Times* article quotes one participant to have 3,500 photos posted to various microstock sites—earning him, on average, 6,500 USD per month [32].

C. Why Crowdsourcing does not equal Open Source

In [7], Brabham insists that Crowdsourcing is not to be mistaken for open source efforts. While both methods tap into the crowd as the main driving force, Brabham identifies the urge to solve a particular—often intriguing—problem experienced by oneself or somebody within the user community as a main incentive to participate in the advancement of open source software. He points out that most programmers working on the Linux source code, as an example, do this as a hobby and submit their solutions only in pursuit of a better Linux kernel and acknowledgement for their work. More importantly, there is no commercial company behind most open source projects, simply because there is no need to: Apart from the individual contributors' own computer equipment and internet hookup—which they usually own anyway—there are virtually no material costs involved. Contrastingly, most Crowdsourcing enterprises involve material goods that are invented or improved by voluntary participants, necessitating investments prior to the finished product actually being made ready for the market.

So, while open source software or projects like Wikipedia theoretically need little to no advance financing, and the finished product is usually given away digitally, for free, crowdsourced products need advance financing for actual production and are, once finished, sold for profit. Brabham questions, whether individuals striving for a physical product of their needs would be willing to publish the digital blueprints to their brain-children to the world for free, if subsequently any company could pick it up and commercialize it without compensating the original inventor. Thus, he concludes crowdsourcing to

"overcome[...] these limitations in the open source model by providing a clear format for compensating contributors, a hybrid model that blends the transparent and democratizing elements of open source into a feasible model for doing profitable business, all facilitated through the web."

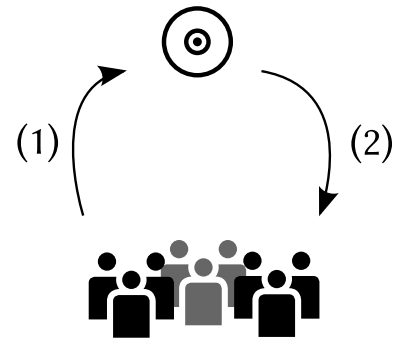


Figure 3. Open source workflow: While all open source projects have to depend on user contributions (1), no controlling corporation is necessary, and the resulting product is being given away for free (2). Also, contributors almost never receive monetary compensation for their work.

Whether, when, and how this perspective might change through increased cost efficiency in manufacturing processes and adapted licensing models remains yet to be seen.

D. Backlashes and Side Effects

Not all crowdsourcing efforts work the way they were intended by their creators. For instance, Chevrolet allowed users to create their own television ad for the new Chevrolet Tahoe by arranging video clips and title cards to their liking. Users responded by creating clips that "skewer[ed] everything from SUVs to Bush's environmental policy to, natch, the American automotive industry" [18]. Called *crowdslapping*, this effect can also be found in members of notorious image boards such as 4chan [1], whose users regularly compete in submitting entries ranging from the satirical to sheer provocation by all possible means. Brabham goes as far as suggesting rules allowing participants to crowdslap without sacrificing the original goal of a crowdsourcing contest [9].

One aspect that should not be underestimated is the composition of the crowd. As pointed out by Kelly [25] and Surowiecki [31], heterogeneity is one key factor to the actual *wisdom* of a crowd. Simply tapping into the Internet does not ensure diversity—especially since typical Internet users are between 14 and 39 years old and still predominantly male [33]. This may foster positive feedback loops where, due to its homogeneity, the crowd seeks a solution tailored to their own beliefs, instead of objectively positive results. Citing several other critical works, Roman goes as far as claiming "there is no clear difference between *the wisdom of the crowd* and *the mob that rules*." [29] Recent debates [26] over the administrative structure of the German Wikipedia seem to support this claim.

V. CRITICISM

Apart from the above mentioned side effects, other criticisms attack the fundamental properties of crowdsourcing, ranging from ethical and moral issues to the dilution of the term itself.

A. Impact on the Labor Market

Howe's term-defining article cites the case of the National Health Museum at Washington, D.C. requiring imagery for an exhibition. Instead of taking the offer of a professional stock photographer, willing to license four of his images for 600 USD, the Museum licensed 56 pictures from iStockphoto.com—for 1 Dollar each [21], with the photographers earning a share of 20 to 40 percent per licensed image [7]. Most of iStockphoto's contributors are hobbyists who don't mind being paid only 1 percent of the market rate, since they need not depend on this particular income. Professional stock photographers do, however, and many of them are unable to compete with the rise of amateurs with digital cameras. Since the advent of microstock agencies and affordable digital single-lens-reflex cameras, basically everyone able to operate a digital camera can upload their work to be licensed for as little as a few Dollars. While this enables small businesses with tight budgets to purchase art and therefore generate revenue that was not available before, it also means that stock images are no longer a scarce good, resulting in overall reduced values for stock photography.

The case is similar with Threadless T-Shirts, where designers earn a petty 2000 USD per design, and even worse when *Grand Search* methods are applied: If somebody wants their business logo designed and posts this particular challenge as an open call via the usual design market places, only one design can actually win. So, if 30 designers put time and effort into designing said logo, only one of them will win the bounty, while 29 designers will have invested their talent and time completely in vain. On the other hand, one might argue that the designers are well aware of this circumstance and should avoid entering in such competitions altogether if entering obviously rarely pays off.

Brabham also acknowledges the negative impact crowdsourcing can have on careers, but insists that, on a macro level,

“crowdsourcing is reconnecting workers with their work and taming the giants of big business by reviving the importance of the consumer in the design process.”

This conclusion sounds very optimistic when considering the case of Amazon's Mechanical Turk. Work, that was previously done by employees with regular salaries and social security, is now outsourced to anyone willing to fill out questionnaires for a few cents per Human Interaction Task. While this can be seen as a tool of empowerment for residents of emerging nations who can now benefit from globalisation, one should not forget the implications this has for the domestic labor market. Classic employer-employee relationships are in peril of being replaced by self-employing HIT performers—domestically or abroad—whose autonomy is more or less ostensible, and who regularly earn considerably less than in the traditional employment situations they superseded. Labor laws regarding minimum wage, health insurance and pension funds are effectively circumvented. Zittrain, in his 2009 *Newsweek* article [37], goes as far as calling this “A new kind of sweatshop”:

“Any jurisdiction that imposes restrictions on how crowdsourcing services operate might find itself bypassed – a firm [...] could simply disconnect all

its contractors in, say, New York, and make more work for people in Arizona.”

Moral questions also play a role. When complex tasks are broken down to the Mechanical Turk's HITs, it is hard to judge the moral implication of one's work. Whereas in traditional employment situation, individuals can—at least to a certain degree—decide whether to work, for instance, in the tobacco or arms industry, no such discrimination can be made when accepting HITs, as pointed out by Zittrain [36]: The tasks are usually heavily atomic, leaving hardly any clue as to who the original client might have been and what ultimate purpose might lie behind any individual HIT. Zittrain paints even bleaker pictures [37]:

“Iran's leaders could ask Turkers to cross-reference the faces of the nation's 72 million citizens with those of photographed demonstrators. Based on Mechanical Turk's current rates, Repression 2.0 would cost a mere \$17,000 per protester.”

B. Who Is the Profiteer?

Altogether, even Brabham acknowledges the true profiteers of crowdsourcing projects are the companies and corporations “that conceived the crowdsourcing application in the first place [7]”. This poses no problem or harm to any of the contributors as long as they are hobbyists—in *Free* [3], Anderson suggests that strictly *because* they are doing this as a hobby, contributors will not mind having to invest time and labor with little return, quoting Mark Twain:

“There are wealthy gentlemen in England who drive four-horse passenger-coaches twenty or thirty miles on a daily line, in the summer, because the privilege costs them considerable money; but if they were offered wages for the service, that would turn it into work and then they would resign.”

While this is fair and good as long as all contributors have different means of assuring their income, it seems hardly fair if crowdsourcing models were the only way to make a living for major parts of the populace, with a chosen few raking in the majority of the labor value.

C. An overburdened Term

Considering the previous implications, it is not hard to understand why blogger Josh Berkus strongly discourages the use of the very word “crowdsourcing” [5]. According to him, in addition to the exploitative nature of the principle, the term implies elitism by setting up a dichotomy between oneself and “the crowd”, and stands in stark contrast to communities where everybody has equal rights.

Berkus—quite rightly—points out the increasing use of the term to describe *any* process harnessing the combined power of a community, indiscriminately to whether it actually fits Howe's original description. Examples for this vagueness in usage can even be seen in research: While Huberman et al. do a good job identifying possible incentives for Youtube users to upload videos [22], the act of uploading one's own videos to a video sharing site can hardly be called crowdsourcing.

Berkus suggests the term *crowdsourcing* be only used when “describing using the Internet to procure cheap labor”, while for community-driven efforts more apt descriptions exist—most of them including the words *community*, *collaboration*, or *citizen*.

VI. CONCLUSION

Crowdsourcing in the original meaning is merely one way of harnessing the power of a community of people all over the world, connected to the internet and collaborating with each other. One should not assume that a solution to any problem can be achieved by tendering the challenge to the community, and even less should one assume that crowd-derived solutions are, in any sense, “free”. Quite to the contrary, finding and supporting a “crowd” large and diverse enough to suit one’s need requires time and money. Furthermore, businesses striving to use crowdsourcing methods need to open themselves to the community they want to be supported by—not least in order to react to implicit criticism of their products or their corporation, delivered through user submissions.

Very importantly, *crowdsourcing* as a means of outsourcing labor to the general public, should not be mistaken for community-based problem solving processes. While crowdsourcing relies on a dichotomy between a clearly defined goal-setter—usually the corporation tendering the challenge—and contributors, no such distinction necessarily needs to exist in communities striving for a solution to a common problem together. Instead, both the recognition of a problem and possible parts of a solution can come from any of the community’s members, who are on equal terms with each other.

Taking the idea of communities collaborating in order to achieve greater goals away from the interests of directly profit-oriented corporations yields a broader and, possibly, more interesting picture of *wise crowds* in action. While the production of community-designed material goods needs still to rely on someone willing not only to rake in the profit, but also finance expenditures in advance, such limits do not necessarily exist in other realms. Benkler, for instance, proposes the idea of *Commons-Based Peer Production* [4], which is able to summarize any effort to aggregate voluntary contributions to a unified intellectual work. Applying this broader view to the analysis of large communities working collaboratively as a massively parallel human supercomputer promises more insight into which principles work in this context, as well as in the narrower definition provided by crowdsourcing.

REFERENCES

- [1] 4chan, <http://www.4chan.org/>.
- [2] Amazon Mechanical Turk, <https://www.mturk.com/mturk/welcome>.
- [3] C. Anderson, *Free: The Future of a Radical Price*. Hyperion, 2009.
- [4] Y. Benkler and H. Nissenbaum, “Commons-based peer production and virtue,” *Journal of Political Philosophy*, vol. 14, no. 4, p. 394, 2006.
- [5] J. Berkus, “Never say crowdsourcing,” <http://it.toolbox.com/blogs-database-soup/never-say-crowdsourcing-34331>.
- [6] D. Brabham, “Moving the Crowd at Threadless: Motivations for Participation in a Crowdsourcing Application.”
- [7] —, “Crowdsourcing as a model for problem solving: An introduction and cases,” *Convergence*, vol. 14, no. 1, p. 75, 2008.
- [8] —, “Moving the crowd at iStockphoto: The composition of the crowd and motivations for participation in a crowdsourcing application,” *First Monday*, vol. 13, no. 6, pp. 1–22, 2008.
- [9] D. Brabham and D. Brabham, “Crowdslapping the government: First Amendment protections for the crowd in government crowdsourcing ventures.”
- [10] C. Caldwell and V. Johnston, “Tracking a criminal suspect through face-space with a genetic algorithm,” in *Proceedings of the Fourth International Conference on Genetic Algorithms*. Morgan Kaufmann Publishers, 1991, pp. 416–421.
- [11] T. Chklovski, “Collecting paraphrase corpora from volunteer contributors,” in *Proceedings of the 3rd international conference on Knowledge capture*. ACM New York, NY, USA, 2005, pp. 115–120.
- [12] Corbis, <http://www.corbisimages.com/>.
- [13] R. Dawkins, “The Blind Watchmaker (Harlow),” 1986.
- [14] Ars Electronica, *Katalog-Archiv*, <http://90.146.8.18/de/archives/festival-%5Farchive/festival%5Fcatalogs/festival%5Fartikel.asp?iProjectID=-8730>.
- [15] F. Galton, “Vox populi,” *Nature*, vol. 75, pp. 450–451, 1907.
- [16] Google Image Labeler, <http://images.google.com/imagelabeler/>.
- [17] J. Howe, “Crowdsourcing: A definition,” <http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing%5F.html>.
- [18] —, “Neo neologisms,” <http://crowdsourcing.typepad.com/cs/crowd-%5Fbits%5Fback/index.html>.
- [19] —, “Pure, unadulterated (and scalable) crowdsourcing,” <http://www.crowdsourcing.com/cs/2006/06/pure%5Funadultera.html>.
- [20] —, “Taking measure of mechanical turk,” <http://crowdsourcing.typepad.com/cs/2006/11/taking%5Fmeasure%5F.html>.
- [21] —, “The rise of crowdsourcing,” *Wired Magazine*, vol. 14, no. 6, pp. 1–4, 2006.
- [22] B. Huberman, D. Romero, and F. Wu, “Crowdsourcing, attention and productivity,” *Journal of Information Science*, p. 0165551509346786v1, 2009.
- [23] Getty Images, <http://www.gettyimages.com/>.
- [24] iStockphoto.com, <http://www.istockphoto.com/index.php>.
- [25] K. Kelly, *Out of control*. Addison-Wesley Reading, Mass., 1995.
- [26] Netpolitik.org, “Wikipedia: Die grosse Relevanz-Diskussion,” <http://www.netpolitik.org/2009/wikipedia-die-grosse-relevanz-diskussion/>.
- [27] A. Quinn and B. Bederson, “A Taxonomy of Distributed Human Computation.”
- [28] reCAPTCHA, <http://recaptcha.net/learnmore.html>.
- [29] D. Roman, “Crowdsourcing and the question of expertise,” *Communications of the ACM*, vol. 52, no. 12, p. 12, 2009.
- [30] K. Sims, “Artificial evolution for computer graphics,” *Computer Graphics*, vol. 25, no. 4, pp. 319–328, 1991.
- [31] J. Surowiecki, *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. Doubleday Books, 2004.
- [32] E. Taub, “When are photos like penny stocks? when they sell,” *The New York Times*, 2007, <http://www.nytimes.com/2007/06/05/technology/circuits/05syndicate.html?%5Far=1&em&ex=1181188800&en=687-225a44f80273c&ei=5087%0A>.
- [33] B. van Eimeren and B. Frees, “Der Internetnutzer 2009 – Multimedial und total vernetzt?” *ARD/ZDF-Onlinestudie*, 2009.
- [34] Vimeo, “Ars electronica 1994 - opening with loren carpenter and stadtwerkstatt,” <http://www.vimeo.com/7723467>.
- [35] M. Wilkes, *Automatic digital computers*.
- [36] J. Zittrain, “Ubiquitous human computing,” *Philosophical Transactions of the Royal Society*, pp. 3813–3821, 2008.
- [37] —, “The Internet Creates a New Kind of Sweatshop,” *Newsweek*, 2009, <http://www.newsweek.com/id/225629>.

Peer to Peer Social Networks

David Langer – {firstname.lastname}@uni-ulm.de

Abstract—Current social networks use client-server architecture. This poses questions of trust towards the service provider who has full access to all data in the network and also has commercial interest in this data. To eliminate this single entity, a peer to peer architecture is proposed. On the other hand peer to peer file sharing applications can profit from using social information. For both use-cases this paper discusses a number of implementations and presents current research.

I. INTRODUCTION

THE way we use the Internet has changed in the last years. Instead of just consuming content published mostly by companies or by some individual users, every user can publish his own content. That content can be anything from small to large. Examples are little status updates, private messages to other users, long articles and videos. This in theory was possible before, but dedicated social network services (SNSs) such as Facebook, MySpace, Xing or LinkedIn offer features and ease-of-use thus, nearly every user (and not only the tech-savvy ones) can connect with friends and colleagues.

These sites can be categorized into two use cases: Sharing professional and personal information. Networks like Facebook, MySpace or Orkut support the latter, while Xing or LinkedIn are better suited for the former. Users share information that they mostly want to keep private or at least control who can see which information. In addition to sharing information, social networks enable users to send and receive messages (both private and public) and keep in touch with friends, acquaintances and co-workers in general. Contacts, both new and old, can be found in various ways. For example a user can join interest groups, use local networks or search for people who visited the same school.

While social network websites continue to grow in user numbers, so do the problems one can see in the use of such a centralized service. Many of these problems emerge from the infrastructure: The data of a user is stored on a centralized server and thus the user has to trust the service provider to protect his personal information. The only way to fully trust the social network would be to exclude the central entity of the service provider. But this also means that one has to rethink the infrastructure, as the server-client infrastructure depends on a provider. The next logical idea would be to use a peer to peer (P2P) infrastructure. A P2P infrastructure would also be more adequate to the nature of a social network as each user profile typically represents exactly one real-life user. Thus each user could simply use his own computer to store his data as this data is fully related to him. A decentralized service could also be used in a ubiquitous way by using for example two hand-held devices on a local network.

Peer to peer systems are mostly known for their use in file sharing applications and thus face the preconception of being

illegal. But P2P systems continue to mature and have become more resistant. An example is the BitTorrent protocol where trackers, which are used to locate other peers, can be replaced by using overlay networks like distributed hash table (DHT) and peer exchange (PEX). Current P2P systems do not take social information into account and treat their users as anonymous peers. Although anonymity is in some respects exactly what some users look for when using P2P systems, one could image that these systems could benefit from recommendations based on download history. This download history could also be potentially used to speed up search queries in the network by creating interest-based groups as distinguished from random hops through peers. Real world trust could be used to create anonymized networks where the real downloader is hidden behind hops through friends and friends of friends et cetera.

The outline of this paper is as follows. Section II highlights the advantages of social networks with a peer to peer infrastructure. We also give a summary of the problems such a system faces. In section III this paper compares the needs of a peer to peer social network to regular peer to peer file sharing networks and how they can benefit from using social components. Section IV and V discuss projects currently in research and systems already in use of the former and the latter fashion. Finally, we conclude this paper in section VI.

II. WHY PEER TO PEER SOCIAL NETWORKS

As noted in section I, the main problem with current SNS is that a user has to put absolute faith in the service provider to protect his data. As people start to realize that every information they publish has an implication and has the potential to be used in a way to harm them, privacy is starting to be a major factor in the choice of social networks. There are options available on every platform to set the visibility of most of the content and information a user publishes. But the problem is not only the privacy to the outside world. The service provider itself has access to all the data and exploits it to support his most important way of making profit: advertising. As he can access every data he could potentially use it for targeted advertising or for data mining purposes. In the worst case he could sell the information to third parties. One could argue, that data on a central server can be encrypted and made unreadable to the service provider. But he could still use traffic information to create a social map of his users. The provider could map HTTP requests and updates to user profiles, as updates to this data will be done by the owning user. This information and the frequentness of requests between these IP addresses create a map which could give the provider a good impression of the social map of his users.

A. Advantages

A P2P infrastructure would support very detailed privacy settings. If one considers the use of asymmetric key algorithms, a user would encrypt his content (which resides on his computer) with his private key. He then could share his public key with specific contacts and by this means give access to his content. When a contact requests that content the user would encrypt the content with the public key with said contact. Consequently the communication can only be read by the intended contact.

A system without a central entity would make censorship impossible and the trust relationship between friends can be used to enhance the networks robustness and anonymize content requests. Anonymization can be achieved for example by using trusted friends to hide behind them. More layers between the user and the requested user make it hard to figure out the origin of the request. A peer to peer social network (P2PSN) could also utilize real-life trust in the network to replicate data on those peers. If a user trusts another one in real-life then their computers can replicate each others content as well. This way network traffic and thus, speed can be reduced as the resources do not have to be downloaded every time.

Scalability, which one would guess that peer to peer architecture has its biggest advantage over centralized online social network (OSN), is not an issue for current centralized OSN because they are all ad-supported and thus can handle the cost of hosting many caching servers. Facebook as the largest example of a online social network has no problem dealing with 300 million and counting users [11]. But for smaller social networks, especially when considering non-commercial OSN the server cost are a heavy burden.

B. Problems

A number of problems arise when one thinks about peer to peer as an infrastructure for social networks. One of them is search: You need to find people you already know or discover new people. In a normal peer to peer system with many peers this is not a problem as it does not matter to which peers you are connected to find the data you are looking for. In the case of SN the peers should ideally be based on social links and so the ring of known peers can be small – not every peer is online all the time. This will make it difficult to maintain a network of trusted peers as IP-addresses change. An alternative identification scheme is needed. A possible solution could be to use a look-up service, but one that is distributed as well. The problem of finding contacts is closely related to the problem of exposure. Current OSN suggest new contacts, as they have the complete overview over all connections between users. This way they can guess that a user which shares many friends with another user may know each other. This overview is not possible and also not desired in a distributed system as it compromises privacy.

Another problem arises when a node is offline. Should another node have the possibility to access ones content in such a case? Adequate technologies for replicating the content are needed, as well as a place to replicate it. As mentioned

before one possible solution would be to store the data on the trusted peer computers.

The process of joining the network for the first time – called bootstrapping – is a possible source for trouble as well. There is no central server which stores contact information (i.e. IP addresses) for all peers and no other initial nodes. So a new user – who has no known peers – can not do a simple search for his friends names. Existing fully distributed systems may already have a solution.

Even with the aforementioned possible problems, the use of a P2P infrastructure make it possible to create a non-commercial, self-organized, durable social network.

III. COMPARING P2P SN TO SOCIAL-BASED P2P NETWORKS

As it looks like OSN could benefit from using a peer to peer infrastructure one could try and use social information on a regular file sharing P2P application. Both use cases can be compared – the main difference is that in file sharing applications the files are usually large and are not requested frequently. Most of the time they are downloaded once by a user and then he keeps that file on his computer. In an OSN, users download many small data chunks like text or images from various sources. As speed and response times are crucial to end users, this matter has to be taken into account when designing a P2PSN

Both applications share the same research challenges. Inspired by [6], they are:

- *Privacy* As described above privacy means that a user has to have control over the information he shares. The information should also not be disclosed to any third party, including no central entity in the network.
- *Decentralization*: Central entities, for example for user authentication, tracking users or bootstrapping, are the norm in current peer to peer file sharing applications. But they can form bottlenecks, security threats and single points of failure and thus should be replaced by overlay networks. Overlay networks like a DHT can be used to replace the tracker in BitTorrent applications for example.
- *Availability* is a concern as the system depends on at least a few peers to be online at all time. People in a P2P network tend to be reluctant with their upload bandwidth which depicts a problem for the availability of the content.
- *Integrity* of the data is a concern as not every peer can be trusted. Bad peers could alter the data and attack the stability of the network.
- *Proper incentives*: This is related to the problem of availability. Peers donate bandwidth on a free basis and thus need motivation to upload enough data and to keep their file sharing application running after downloading the content they wanted.

IV. RESEARCH AND DEVELOPMENT OF P2P SN

The following sections highlights some examples of social networks, that adapt a peer to peer architecture and their attempts to mitigate the aforementioned problems.

A. PeerSoN

PeerSoN [3] is a proposal for a distributed peer-to-peer system coupled with encryption. The goals are to keep all functionality of an OSN (social links, digital personal space, communication) but also to enable the system being used locally between two devices without the need of an Internet connection.

Encryption is used to give the user control over his content. He can control who has access to his content by sharing keys. In a first approach, a public key infrastructure is assumed, where the user would encrypt his data with the public key of the intended friend to which he would share the content.

When a new user wants to share keys with a friend, they have to exchange information outside of the network to verify their credentials. This exchange can also be done when two users physically meet. Then keys can be shared between their two devices without internet access. This newly trusted friend can recommend more friends, confirming that they are trusted as well. Completely unrelated users can only be verified by the public key infrastructure. This is a problem as impersonation should ideally be impossible. PeerSoN recommends a two tier challenge response protocol to first exclude bots and then using shared memories or secrets to verify the person in question.

PeerSoN relies on a two tier architecture: Look-up happens via a DHT while the content remains on each users computer. A DHT is an overlay network which distributes its content – stored in key-value pairs – across connected nodes. It maintains stability when nodes join, disconnect and drop. The DHT is used to store information about whether a user is connected to the system, his location, IP-address, files, and where they can be found. The peers of the social network are divided into full nodes of the DHT and into nodes which connect to it only to do a look-up or to update it. This is especially important for handheld devices that lack the qualities a DHT node requires.

The information in the DHT is stored by using a special syntax for key names and for values. To get the current IP address of a user one could do a `get`-command with the user id – which is calculated by doing a hash of the users email. The response would look like “loc:home#123.45.67.89” for each location where this user is signed on. For each file a user shares, all possible download locations are stored. This way the information stays online even when the user disconnects from the network. Additionally every user has an index file which lists all the files this particular user shares.

To elaborate how the system works we will look at the sign-in process and how one can get the updates of a friend. A user can sign in to the system by first receiving his information from the DHT (sending a `get` command to the DHT using his user id as a parameter), updating it to set his location to online and including his current IP-address. By setting the correct value to offline the user can log off later. This user (A) can now get updates from his friend (B) by requesting B’s index file from the DHT (Figure 1). The response would then contain all the possible download locations for this file. If B is online, then he has the most recent version and thus, it can be downloaded directly from him.

Messages like those posted to a users personal space or private messages can be stored in the DHT while the target node

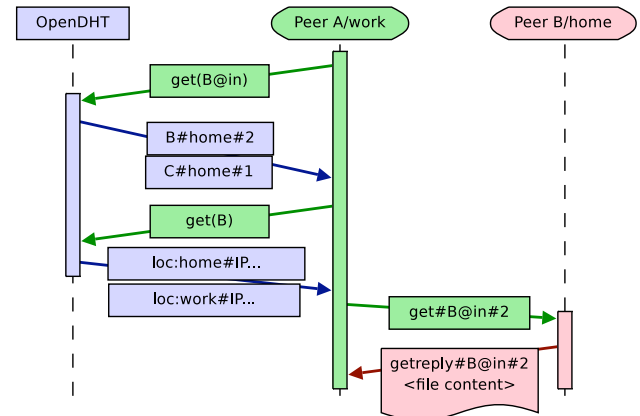


Figure 1. Peer A gets index file from peer B

is offline. Those asynchronous messages are always requested after the log-in is complete. Currently the DHT only supports messages with up to 800 characters.

The group that built PeerSoN identified three main issues:

The first issue is storage: As the shared content is small in size, the best way to speed up access is by replicating the content on multiple other nodes. This ensures also that other nodes can access the newest content, even when the creator is offline. But it also means, that a user has to host and upload more content than he consumes. That will make it hard to advertise such a OSN to the end user as an OSN can only be successful with a large user base.

The other one is how the system can ensure that the content is hosted diversely enough. Content can be replicated using friends, friends of friends, or geographic relations. Current popularity of content could also be taken into account.

Finally, the use of DHT could prove problematic as there is no authorization. Private messages, which are stored while the recipient is offline can be deleted by anyone on the network. When using OpenDHT (an implementation of DHT) this can be avoided by requiring a secret key to delete messages from the DHT. This key can be stored inside the encrypted message, so only the desired recipient can read it.

B. Privacy Preserving Social Networking Thorough Decentralization

Cutillo et al. propose an architecture for a peer to peer social network service [5]. The main two purposes of their work are eliminating the service provider as the central entity and to use real-life trust relationships between friends for anonymization and content replication purposes. They focus especially on security and how the system can be made resistant to impersonation¹, man-in-the-middle², and blackhole attacks³.

The proposed system uses three components:

¹Impersonation: Pretend to be someone the targeted user knows to commit fraud

²Man-in-the-middle attack: Attacker can access/alter information send between to nodes on a network

³Blackhole attack: Dropping packets to create a denial-of-service

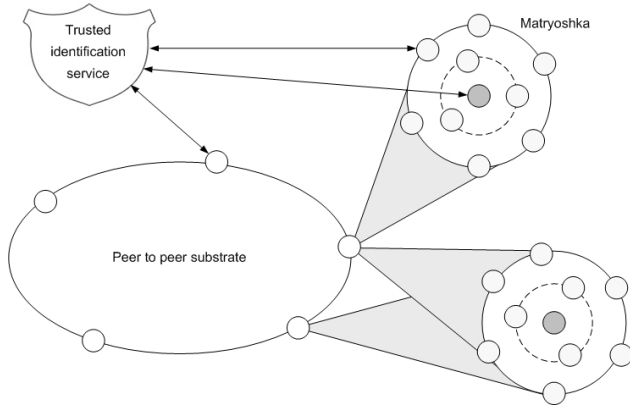


Figure 2. Main components

- several matryoshkas (social network part),
- a peer to peer substrate (e.g. DHT) (peer to peer part), and
- a trusted identification service.

One user in the network is represented by exactly one node. Each of these nodes has a virtual matryoshka build around itself for anonymization purposes (see Figure 2). These matryoshkas can be imagined just as their namesakes, the Russian nested dolls. Each layer of the matryoshka consists of a number of other user nodes. In the center is the node of the respective user. Around this, in the innermost layer, are all the trusted contacts of this node. The next ring consists of trusted contacts of the contacts from the first layer. This friend-of-a-friend method continues on every further layer. Profiles and profile data is then replicated on the first layer of trusted contacts using encryption.

A message or a request addressed to a user starts at the outermost ring of his matryoshka. The outermost ring is the one which is addressed in the DHT. From there and is transferred inwards, finally arriving at one of the nodes in the first layer (if it is a request for a user's data) or at the node itself. It gets processed and the answer is sent on the same route back to the originating node.

All communication is kept between trusted contacts, making it impossible to successfully carry out a man-in-the-middle or black-hole attack on a user, as for this the attacker has to have a node in place between two communicating nodes.

Secure communication is assured by using the trusted identification service providing certificates for a unique pseudonym and a unique identifier. Impersonation is thus made impossible.

The peer to peer substrate uses a DHT to provide contact information for nodes based on their unique identifiers. As the contact information to a specific node is not a direct pointer to the node, but instead a series of hops starting from a node on the outermost ring of the specific node, anonymity is granted.

A user can only join the when he gets invited by another user. As a first step, he creates his identity using the identification service. The contact list of the new user only contains the user that invited him. Based on this contact, the user starts the creation of his matryoshka, sending the request to do so to his contact (see Figure 3). From there the request gets passed on

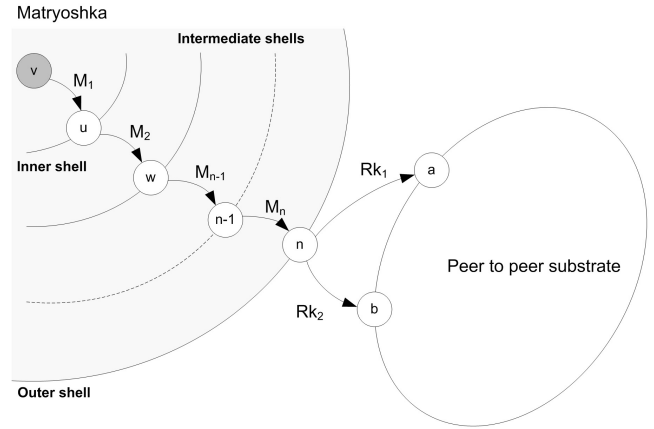


Figure 3. Key registration for node v

following trusted contacts until the desired number of shells are created. At this point the last node on the outermost shell registers the node identifier of the new user along with his own node identifier at the DHT. This chain is then utilized for all communication like profile look-up and messages exchanged between peers. The matryoshka is maintained at all times. This includes adding outer nodes (and thus creating multiple paths to a node) when a trusted contact is added to the contact list and updating paths when a node on a path leaves the network. When that happens, nodes outward to the leaving node unregister themselves from the DHT and the node on the inside starts the registration process again.

The evaluation of this proposed system shows that all the privacy and security goals are achieved, especially making impersonation attacks impossible and end-to-end confidentiality. Problems still exist with speed and availability of profiles. When a contact has a small amount of contacts, there is a high possibility that his profile data is completely offline. Speed could be a problem, as most peer to peer architectures are not designed to be used for frequent requests and small response times. The system was not implemented as a model, so no practical data is available. Additionally, asynchronous messages are not possible in this proposal.

C. imeem

An actual working peer to peer system was imeem [1]. It now is completely web-based, but imeem used to require a client that the user had to download and run locally. The client provided most of the features a regular, web-based OSN has: instant messaging, blogging and photo-sharing. But in addition it also featured a file-sharing mechanism, which enabled users to specify files and folders they wished to share with others. The content was not uploaded to a central website – instead users accessed the content directly from the publishing user's computer using peer to peer technology.

Users could create multiple hubs containing files, forums, and pictures [2] for their account. These hubs were called "meems" and could be either public or private making it possible to control privacy pretty well. The accompanying website was only used to download the client in the beginning. When the

company started to move more features from the client to the website, immeem gained significantly more users. In 2007 the client was completely abandoned. This shows that it can be hard to advertise the benefits of an application that users expect to be web-based but needs a client to be run.

D. Blogosphere

The network that is formed through blogs is a social network, and it is also comparable to an OSN. Users – or bloggers – share text, fotos or videos on their digital personal space. These posts are mostly personal [9], but it is reasonable to assume that these posts are not as personal as the ones in OSN as the user has only limited control over his privacy. Limited meaning he can only choose to operate a public blog or to password protect his posts.

Users can communicate by e-mail or by using contact forms on each users site. Interaction is also possible via comments on posts and/or TrackBacks⁴. Not every blog publisher enables comments, but when they are enabled, they provide direct feedback and room for discussion. Trackbacks are automatic notification when another blog post is related to another.

The “blogroll”, links from one blog to another (and thus people), act as some sort of friend network. These, unlike those in OSNs, do not have to be symmetric.

A more technical and standardized way of showing connections and also adding metadata to the links is by using XHTML Friends Network (XFN)⁵. XFN adds metadata like what kind the friendship is, if one has met them, if is related, and geographic information by using the `ref`-Tag. This information then can be analyzed by browsers and search engines (for example rubhub.com).

This type of social network is called the “blogosphere” and can be compared to a peer to peer approach. Every user can (in theory) host his own blog or website with no second party involved. This can be expensive, so shared or dedicated servers in a datacenter are often used. The owner of such a center could potentially access all data, but this irrelevant in this case, as this data is public (at least most of the time) anyway. In general one can say that this network is the most open one found to date.

When more metadata gets added to the web, the information can be used on one hand by search engines but on the other hand also by third parties with bad intentions such as phishing or vishing attacks or spam emails. As the information can be interpreted by computers when it is available in microformats⁶, profiles containing all known facts about a person can be created much faster. More information also means more attack vectors on a potential victim.

V. PEER TO PEER WITH SOCIAL CONNECTIONS

Privacy concerns on the Internet are increasing as users start to see how most of their actions can be traced back to them. This is especially true and easy to do in modern file sharing applications like BitTorrent. In the case of BitTorrent,

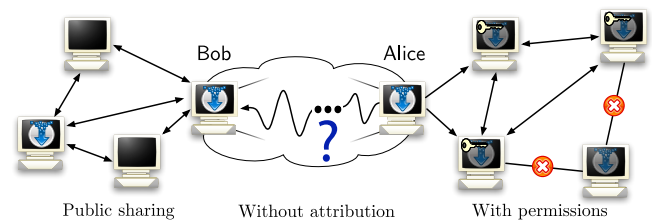


Figure 4. OneSwarm: User Bob downloads public data, while Alice downloads content from Bob using privacy preserving overlay. She offers the file to her trusted peers

it is especially easy to monitor user behavior as everyone has access to the data of the tracker. The tracker keeps track of all the connected nodes in order to enable discovery of more download sources. Privacy is a trade-off to usability, robustness and speed. In the following, we will give examples of how such systems work and some insight into research aspects.

A. OneSwarm

Isdal et al. provide a working prototype of a privacy preserving P2P data sharing application called OneSwarm [6]. OneSwarm has already been downloaded by hundreds of thousands of users, so actual usage data is available.

Figure 4 shows three privacy states in one single swarm:

- *Public distribution*: Sharing files in the BitTorrent network using backward compatibility. Content is downloaded from public sources, which means full speed but on the other hand no privacy at all.
- *With permissions*: Sharing files with trusted OneSwarm users privately. Trust on a per-file base enables users to share only the content they want with the users they want.
- *Without attribution*: Sharing files with another OneSwarm user preserving privacy. This mode is recommended when sharing data which the user wants to download anonymously. Privacy on the source and target node is assured by relaying the data through an unknown number of nodes. This is the least performant way of sharing data. Sharing without attribution can be compared to the Freenet file sharing application [4]. Freenet enables users to share files anonymously by pooling storage and storing the data near places where it is requested.

Relying on public/private key pairs, trust relationships can be created either by looking for peers in the local network, through already established contacts using Facebook or Google Talk or by e-mail invitations. Users also have the option to subscribe to community servers, which handle key exchange for a group of users.

Impersonation attacks are possible in OneSwarm as there are multiple ways to add new peers and all of them use already established trust, but, instead of real-life trust only virtual trust, is used. This work shows that by using social networks where users trust each other peer to peer file sharing can be both fast and secure enough to be usable in three sharing methods. The user has three-step control over the way he shares and downloads content and thus has also control over how much he wants to trade privacy for speed.

⁴http://www.sixapart.com/pronet/docs/trackback_spec

⁵<http://gmpg.org/xfn/>

⁶<http://microformats.org/>

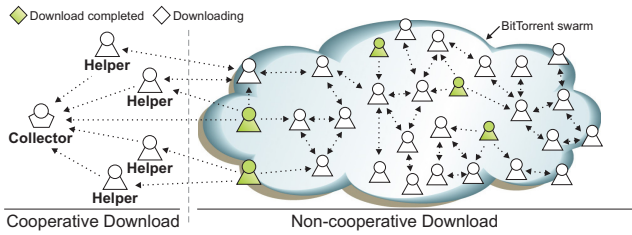


Figure 5. Collaborative downloading in TRIBLER

B. TRIBLER

TRIBLER [10], as proposed by J. A. Pouwelse et al., is a system that uses social and interest connections to provide better speed and usability for peer to peer applications. Privacy concerns are completely omitted. Pouwelse et al base their work on the opinion that users in a peer to peer file sharing system do not tend to give away their bandwidth out of pure altruism. When in a social environment as opposed to the anonymity in peer to peer systems social recognition can act as a reward. The BitTorrent protocol currently uses a tit for tat system where download bandwidth is exchanged for upload bandwidth.

TRIBLER's goals are: Using social networking to enhance download speeds and using so called taste buddies to let users find content in which they might be interested. This is done by de-anonymizing users and adding interest or social groups. TRIBLER aims to be completely de-centralized.

Social network data is stored in an overlay network while using the standard BitTorrent protocol for data sharing. Every peer holds a ranked list of content he likes – a so called taste list –, which is pre-populated with his last recent downloads. He also replicates taste lists of a number of his taste buddies along with lists of random peers, forming a mega cache. This cache gets populated and synchronized with other peers in the network using a special epidemic protocol called Buddycast algorithm. Buddycast is using both exploitation (connecting to his taste-buddies) and exploration (connecting to random peers) to find taste buddies.

While downloading content in TRIBLER the user can see how much his contacts (both friends and taste buddies) contribute and where the other peers are located on a map. The user interface also shows his friends list to give him a sense of social awareness. The user can use online friends to speed up his downloads by using them as helpers (see Figure 5) who download small chunks of data for the user (called collector). This is called collaborative downloading, and as shown by Pouwelse et al. can speed up downloads by up to six times with 30 helpers.

TRIBLER shows that a social component can improve file sharing in a way beyond just downloading content from friends. The work also shows that it is a good idea to incorporate trust and social networking capabilities into a file sharing application. Using an overlay network over the BitTorrent protocol, both speed and content discovery can be positively affected.

C. MAZE

MAZE is a peer to peer system employed in China's internal research network, for example in Chinese Education & Research Network. Its architecture is similar to Napster, but adds a friend list. Yang Zhao et al. [13] distinguish two networks in the MAZE application:

- The *friend network* is imposed by the use of a contact list. Users can add other users to their friend list. They can find them via their peer status (available bandwidth, availability), neighbors on the local network, via the contact list of other friends, or by knowing each other in real-life.
- The *download network* is constructed by the files a user downloads. Every successful download of a file to a peer results in an entry on a central log server.

It can be shown that both networks share the “small world” property like other social networks. A small-world network is a graph “in which most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of hops or steps” [12]. This is interesting especially for the download network.

Yang Zhao et al. propose to combine both networks into one network to enhance searches. They use the EigenTrust [7] reputation management algorithm for their search approach, called EigenForwarding. EigenForwarding uses the EigenTrust algorithm, but also extends it to include the private download history of each peer. All in all this algorithm makes it possible to find the best possible peer to forward a search query to.

Their research shows that by including the social information search performance is increased and better than both flooding and random forwarding. This result is backed up for example by [8].

VI. CONCLUSION

As shown in this paper, social networks have disadvantages like the lack of privacy towards the service provider. It has been shown that replacing the client-server architecture with a peer to peer architecture has many good side effects, including detailed control over the data a user shares. Simple cryptographic methods allow for a fine-grained control.

Numerous attempts to create social networks on top of a peer to peer architecture exist and most of them are very promising. These systems do not all have the same focus. Some focus solely on privacy where some trade privacy for usability.

As imeem has shown in the past, it is not easy to convince users to download software for a service that is provided as a web service in general. A large user base on other networks makes it hard for users to switch, because managing multiple networks at the same time can be tedious. Without a large amount of users switching in the beginning it will be hard for a social network to exist. Incentives like good usability and extendability are very important for users to consider the switch.

A new system should use every possible benefit of this architecture. As smart-phones are getting popular this use-case could be a big advantage for peer to peer systems as they can act on local devices and do not need a permanent Internet

connection to be usable. Challenges like search, availability, integrity have been addressed, and it was also shown that these can be overcome.

On the other hand, there is also potential to enhance existing peer to peer file sharing. Query routing and search can be more efficient when social information is available. Social information can be used to cluster peers into groups with the same interests, shortening paths for search queries. Privacy and download speed are other ways to exploit social information. Friend to friend sharing, like sharing-with-permission in the OneSwarm network, make it possible to exactly specify which content to share with which user. Other options include to share content without attribution to hide identity of the downloader. TRIBLER uses friends to speed up downloads.

Everything considered, peer to peer architectures are a very good option for social network services and file sharing applications benefit from including social information.

REFERENCES

- [1] imeem, “imeem Unveils at DEMO World’s First All-in-One Instant Communications Softwar” 2005, [Online; accessed 24-January-2010]. Available: <http://tinyurl.com/imeemunveils>
- [2] R. Broida “imeem review from PC magazine” 2005, [Online; accessed 24-January-2010]. Available: <http://www.pcmag.com/article2/0,2817,1872022,00.asp>
- [3] S. Buchegger, D. Schiöberg, L.-H. Vu, and A. Datta, “Peerson: P2P social networking: early experiences and insights,” in *SNS '09: Proceedings of the Second ACM EuroSys Workshop on Social Network Systems*. New York, NY, USA: ACM, 2009, pp. 46–52.
- [4] I. Clarke, O. Sandberg, B. Wiley, O. S. O. Wiley, and T. W. Hong, “Freenet: A distributed anonymous information storage and retrieval system,” 2000, pp. 46–66.
- [5] L. A. Cuttillo, R. Molva, and T. Strufe, “Privacy preserving social networking through decentralization,” in *Wireless On-Demand Network Systems and Services, 2009. WONS 2009. Sixth International Conference on*, 2009, pp. 145–152.
- [6] T. Isdal, M. Piatek, A. Krishnamurthy, and T. Anderson, “Privacy-preserving P2P data sharing with oneswarm,” UW-CSE, Tech. Rep., 2009 [Online]. Available: <http://oneswarm.cs.washington.edu/papers.html>
- [7] S. D. Kamvar, M. T. Schlosser, and H. Garcia-molina, “The eigentrust algorithm for reputation management in p2p networks,” in *In Proceedings of the Twelfth International World Wide Web Conference*. ACM Press, 2003, pp. 640–651.
- [8] A. Modarresi, A. Mamat, H. Ibrahim, and N. Mustapha, “How community-based peer-to-peer social networks can affect query routing?” *Networked Computing and Advanced Information Management, International Conference on*, vol. 1, pp. 432–437, 2008.
- [9] C. Neuberger, C. Nuernbergk, and M. Rischke, “Weblogs und Journalismus: Konkurrenz, Ergänzung oder Integration?” *Media Perspektiven*, vol. 2, pp. 96–112, 2007. [Online]. Available: http://www.media-perspektiven.de/uploads/tx_mppublications/02-2007_Neuberger.pdf
- [10] J. A. Pouwelse, P. Garbacki, J. Wang, A. Bakker, J. Yang, A. Iosup, D. H. J. Epema, M. Reinders, M. R. van Steen, and H. J. Sips, “Tribler: a social-based peer-to-peer system: Research articles,” *Concurr. Comput. : Pract. Exper.*, vol. 20, no. 2, pp. 127–138, 2008.
- [11] J. Rothschild, “High performance at massive scale – lessons learned at facebook”. [Online; accessed 12-Januaray-2010] Available: <http://cns.ucsd.edu/lecturearchive09.shtml#Roth>
- [12] Wikipedia, “Small-world network — Wikipedia, The Free Encyclopedia,” 2009, [Online; accessed 24-January-2010]. Available: http://en.wikipedia.org/w/index.php?title=Small-world_network&oldid=330%062877
http://en.wikipedia.org/w/index.php?title=Small-world_network&oldid=330%062877
- [13] Y. Zhao, X. Hou, M. Yang, and Y. Dai, “Measurement study and application of social network in the maze p2p file-sharing system,” in *InfoScale '06: Proceedings of the 1st international conference on Scalable information systems*. New York, NY, USA: ACM, 2006, p. 57.

Mobile Social Applications

Harry Brandenburg – {firstname.lastname}@uni-ulm.de

Abstract—This paper presents a survey of current available mobile social applications and current scientific work on that topic. Mobile social applications are becoming more and more ubiquitous in our daily life, especially since the recent development on the mobile phone market. The number of mobile social applications has recently grown bigger and bigger, supported by the increasing number of people using mobile phones with advanced functionality.

I. INTRODUCTION

The term “Mobile Social Application” can be defined in different ways. In this paper every application which was designed for mobile phones and has one or more features that have social aspects are called mobile social applications. This social aspects can differ a lot between the applications, but mostly the application tries to support social interaction between users. For Example an application that allows to tag locations and view them later isn’t called a social application, of course that changes if it enables the users to share their location with others. Today’s mobile social applications are often basic tools, bringing the known features of their online platform to the mobile platform (e.g facebook mobile application [10], studivz mobile application [40] or the MySpace application [30]). Others are trying to make use of the mobile phones specific features (GPS-location, accelerometer,...) and improve the usability. Some of these applications are presented in this paper. People developing mobile social applications face a lot of problems. The most important challenges which come with the development of mobile social applications are privacy concerns, low energy consumption and the communication between the phones and the backend infrastructure. With the growing number of sensible personal information, the people care more and more about their privacy. The mobile application has to satisfy this need by storing the personal data as safe as possible and allowing the people to define who can access their data. Another important issue is the battery life of mobile phones. The application has to implement an intelligent way to manage the resources of the phone to keep the energy consumption as low as possible while using the application, because most of the users don’t use applications when they have to recharge their phone every few hours. How this problems could be solved and what methods are used by modern mobile social applications is shown in this paper.

The first section categorizes the mobile social applications based on their main focus and gives some examples on every category. Some possible future approaches and other ways to use mobile social applications are shown at the end of this overview. The second section presents some applications in detail and shows what features they have and how they have been implemented. The third section is about middleware which was developed to implement mobile social applications on top

Table I
APPLICATIONS AND RELATED CATEGORIES

Category	CenceMe	Micro-Blog	Connecto	PeopleFinder
location	+	+	+	+
blogging	-	+	0	-
data sharing	+	0	0	-
networking	+	0	+	+
meeting	+	0	0	0
coordination	0	-	0	-
voting/rating	-	+	-	-

of them. The fourth and last part of this paper is a conclusion which summarizes the insights gained in this paper.

II. OVERVIEW

There are mobile applications for nearly every purpose available. These can be categorized because of their main aspects as follow:

- location based applications
- mobile blogging applications
- social data sharing applications (videosharing...)
- social networking
- dating and meeting applications
- applications that help to coordinate things
- voting/rating applications

This isn’t a definite categorization, because there are applications that apply on more than one category or even not really on either one of them. The overlapping categorization of applications is shown in table I for the applications which are discussed in detail later in this paper.

A lot of applications are made for sharing your current location and looking up where other users are. These location based applications are quite common today. One typical location based program is Friendticker [12]. It’s available as a Java application and as an iPhone application. It uses Near Field Communication (NFC) chips to identify the current location. The localisation data is saved on a small chip which can be detached nearly everywhere. Friendticker implemented the chip on a sticker which has also a QR-Code and a SMS hint on it for mobile devices which can’t read NFC chips. The data gained from the sticker is uploaded and from now on the other users can view the actual position of the mobile device. A few other features are sharing your location and sending a notification to all your friends. You can also ticker messages similar to Twitter. Another location based application is Belysio [5]. Like most applications of this category, this program uses GPS to share your location. Besides finding out where your friends

are, this application offers the possibility to communicate with your friends by sending instant messages. It is also possible to create a “Geo Journal” which is similar to a blog entry with your current location.

Other applications are focused on real-time sharing of your location. A few examples are ipoki [18], mologogo [27], footprint history [11], mysonar [29] or bliin [6], which have features like detecting your current location via GPS and sharing it with others (e.g. on a map which can be viewed online). These applications are often called location tagging applications. Other common features of these applications are sending messages between their community members. MySonar also adds some flirting features.

The following applications use the GPS sensor of mobile devices for tracking your location. These programs are often used to publish own training tracks online. MapMyTracks [21] is one of these applications, which allow the user to record their GPS data while doing sports like cycling, running or sailing and publishes their route online afterward. Other users can rate and comment on these routes then. This tool can also be used to analyze your training, because it keeps track of additional information like your current or average speed and so on. The Nokia Sports Tracker [41] is another similar application to MapMyTracks. It offers different features to analyze and improve sportive activities. Users are also able to compare their results or find new tracks for their own training program.

Applications which are focused on streaming video data are another class of mobile social applications. They belong to the social data sharing category. One of them is called Okud [33], a program which allows the user to upload videos and comment on other videos. Public videos are shown on the website, similar to Twitter, except that the messages contain videos instead of small text messages. Another video sharing application is MobiSNA [14], a mobile social video sharing service. With MobiSNA the user can share his videos with friends and create video content online. It is also possible to share your video in real-time. Other features of this application are video blogging, building of video interest groups, or viewing video story lines. Qik [37] is another mobile application of this category, it allows you to upload your videos or stream live on their website.

Programs which are designed to create and update a multimedia blog are another sort of mobile social applications. These applications are part of the category of “mobile blogging applications”. Most of them have also location tagging or sharing features, but their focus is set on sharing experiences and other things that happen in life with others. Examples for this category are programs like OvalPath [34] or Micro-Blog [13]. Ovalpath is a so called “lifeblogging” application. It is designed to combine mobile multimedia blogging, social networking and location based services. Micro-Blog differs a little bit from that, and will be shown in detail in section III-B.

A further category is about applications which allow the user to vote or rate something. A lot of programs are focused on rating locations, for example restaurants, discos etc. These applications allow the users to tag their locations, and if they want the users can give a feedback about that location. They

can also add a rating and a report of their experience they have made there. Buzzd [8] or MyCityMate [28] are two examples of this type of application. An example for applications for voting things is Jukola [32]. Jukola is an application which is designed like an interactive mp3 jukebox, where people can vote which music should be played next on a democratically basis. The votes are submitted via mobile phones or other wireless devices, while the actual standings can be watched on a public display. Jukola offers also the possibility for local bands to upload their music into the Jukola system and get them some publicity by this.

A lot of applications are primarily focused on communicating with others, the so called mobile social networking applications. Most of these applications haven’t been developed on a mobile platform at first, but the developers added mobile clients for them in the last few years. Facebook [10], MySpace [30], StudiVz [40], JuiceCaster [19], Twitter [42], Mobimii [25], MocoSpace [26], Seismic [39] or qeep [36] are just a few examples. A few applications of this type use bluetooth to find or detect others around you, for example aka-aki [1], MobiLuck [24], or Next2Friends [31]. Other social networking applications try to enhance the mobile networking by improving communication with your contacts. Cluestr [15] shows in a survey that enhanced group communication is useful and can save a lot of time. Features of Cluestr are a thread-like billboard, a poll feature for voting and a ToDo list which each can be viewed and changed by the whole group. The next application is called Friendlee [2], a program which was made to analyze your contacts and your call and messaging history and detect your closest social contacts automatically. It is also possible to share your location and status with your contacts.

The last category is about mobile dating applications. These applications allow the user to find new people and flirt with others with their mobile phone. Examples are Icebrkr [17] or MeetMoi [22]. Like online dating sites these applications can be used to find singles living near you. Most of the applications of this category are also able to suggest possible partners for the users based on their profile data, interests or other social aspects. Social networking features are also quite common in applications of this category.

There are also a lot of ideas what future mobile applications could look like. Two people from the DUB Group of the university of Washington published a paper which mentions how volunteer activities could be supported by mobile social applications [20]. Because of the fact that a lot of people want to do volunteer work, but don’t know where to start they propose a program which intervenes the user at the right moment based on his location, activity or routines. The software should also be able to give the user positive feedback about the impact of their work if they volunteered somewhere to keep them motivated. Another group had the idea to enhance the use of mobile applications on transitional spaces [4]. On the way to work in the subway or on the way to shopping in the bus when people don’t have to do something in general are meant by transitional spaces.

III. APPLICATIONS IN DETAIL

A. CenceMe



Figure 1. Buddy Status of the CenceMe Application [23]

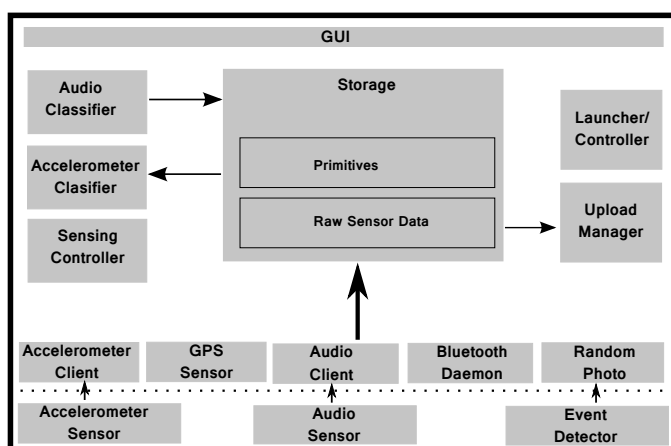


Figure 2. Architecture of CenceMe

1) *Overview:* CenceMe[23] is a “people-centric sensing” application, that means it is build to automatically detect the users’ status by analyzing the input of the mobile phones internal sensors. So the user doesn’t need to update his status by himself all the time. The data which is needed to detect the users’ current status is gained by using accelerometers, proximity sensors, microphones, the camera and GPS sensors of the phone. The users status is then automatically updated at facebook. Some exemplary statuses could be “dancing at a party”, “in conversation”, or “at work”. The users are able to set some privacy rules and they can view the status of their friends as shown in Figure 1. If a friend isn’t online at the moment, his last known status will be displayed.

2) *Technological details:* The CenceMe application was developed for the Nokia N95 mobile phone which uses the Symbian operating system. The application was written with Java Mobile Edition. The backend infrastructure is hosted on server machines. The following features are provided by the CenceMe software on the phone:

- sensing of information
- classification of the sensed data
- showing the users actual presence on the phone

- uploading the data to the backend servers

Because of the Java mobile platform and the Symbian operating system there have been some programming restrictions. The symbian OS is able to interrupt the third party programs at any time, which has to be regarded while designing the software. It is also possible that the programs resource requests could be denied because of low resource availability. There have also been API and operational limitations, for example there is no API on the N95 phone for accessing the internal accelerometer. Another difficult problem is the high energy consumption of the internal sensors like GPS or bluetooth which result in a short battery life. Because the used version of Symbian with JME didn’t support power cycles with the bluetooth or GPS sensor an power aware duty-cycling has been implemented. The designer wanted to implement an portable software, so they tried to implement as many parts as possible of the software in JME, but the API limitations mentioned before made it inevitable to implement some features directly, using the native Symbian API. Finally the software was implemented using a software suite running on the N95 mobile phone and a backend infrastructure. The communication with the backend is done by the upload manager. The upload manager connects and disconnects to the backend servers accessing the data from the local storage and uploading it.

One important technical feature of the CenceMe software is the classification of gathered sensor data, while complex classification algorithms are done by the backend machines, relatively easy classification is done directly on the phone. The architecture of the phone software is shown in Figure 2.

The accelerometer sensor, the audio sensor and the event detector are C++ modules of the symbian platform. The sensed data is accessed by the JME methods through a socket connection. The Bluetooth Daemon is used to retrieve the MAC addresses of other devices and find out if there are more phones using the CenceME application around. The accelerometer client is written in JME and it is used to get the actual accelerometer data. The audio classifier gets the data from the audio sensor and uses a discrete fourier transformation to analyze the audio data on the phone and checks if the audio sample represents human voice. This classified data is called a “primitive” and it is stored in the storage.

The sensing controller manages the sensors. It is responsible for starting, stopping or monitoring the sensor clients. The activity classifier uses the data which has been stored in the local storage by the accelerometer. The activity classifier is able to differ between sitting, standing, walking or running. As described here, the classifiers on the phone use different methods to generate different types of “primitives”.

The GUI which presents the available sensing data to the user is called ClickStatus, the design is shown in Figure 1. The privacy settings can be changed using the privacy settings GUI on the phone. ClickStatus uses the privacy settings provided by the privacy settings GUI. After starting the ClickStatus and logging in with the CenceMe credentials the program downloads a list of CenceMe buddies from the server. This list can be refreshed by a menu command option every time. The latest known sensing information of a buddy can be accessed by selecting it from the buddy list. Buddys are Facebook friends

which have installed the CenceMe program on their phone too.

The backend software of the CenceME application is written in Java. The communication between the phone and the backend servers is implemented by using remote procedure calls with the apache XML-RPC library. The data is stored in a MySQL database on the server side. The data exchange is initiated by the phone and the so called "primitives" are uploaded and saved in the database then. The backend supports "push" and "pull" methods to make the presence data available to other platforms. The push-based data publishing is done by the push connectors.

3) *Conclusion:* The user study which involved 22 people showed that they enjoyed using the CenceMe Application. The CenceMe Application fortified the users curiosity which leads to an increased use of the application to look up what the other people are doing while on the move. CenceMe can also help people to learn more about their own activity patterns and social status. The location, activity, the historical log, the random images and the social context were the most popular features. There is also a version of CenceMe software for the iPhone available today [9].

B. Micro-Blog

1) *Overview:* Micro-Blog[13] is a tool for sharing, browsing and querying global information. Like CenceMe it was developed for the N95 phone. The idea of Micro-Blog is that the users generate multimedia blogs, while they are on the move and then enrich this information with data from the mobile phone, like positioning data. These blogs are uploaded on a backend server then, which means that they are available for the others from now on. The users are able to view a world-wide map and the blogs within the application.

They are also able to query selected regions, which means that the query (e.g. are there any restaurants near the train station in ulm ?) is uploaded into the system, and Micro-Blog sends the query to all the connected mobile phones in the desired area. These queries have to be answered by other users then. How the GUI looks like when answering a query can be viewed in Figure 3. All users in this region will get the query, if they have enabled this option. Other features of Micro-Blog are a news service, where the users are journalists, who can post audio-visual news.

2) *Technological details:* The Micro-Blog client was programmed with Nokia's Carbide C++ Version 1.2. To access the mobile phone's location or network service and file system the Nokia S60 3rd SDK was used. Micro-Blog was designed to interrupt the normal operations of the phone as little as possible. The default state of the client is the idle state. Whenever the user starts a new blog, the state is changed to "Blog". In the "Blog state" the content which is generated by the user is saved. If the user is finished the content is tagged with the location provided by the phone. Users are able to change this location later to enable offline blogging. While sending the content to the backend, the client switches to the "Sync" state and transports the content to the server via a TCP connection. After sending the content the software returns to the idle state.

Mico-Blog uses an periodic polling mechanism to read the data from the sensors. The application state is changed to



Figure 3. The query system of Micro-Blog[13]

"Log", whenever a new result is returned or if an alarm has been activated by an asynchronous sensor. The sensor data is then stored on the phones file system. After that the system returns back to the Idle state. If a user activates the feature, his phone is always visible on the map and a persistent TCP connection to the remote server is established. New location coordinates are uploaded then periodically. The "Query" state can be achieved if the phone is qualified for some pending queries from the server. In the "Query" state these queries are forwarded to the phone. The phone displays the queries on the GUI and waits for user action. After the user has finished with the query the created results are transported back to the server and the program returns into the Idle state. The backend of the Micro-Blog application consists of server machines with an MySQL 5.0 database. To view the data which has been generated with Micro-Blog worldwide, an Apache Webserver, JavaScript and PHP 5.0 are used. The communication is done with the HTTP protocol. A user generated query is routed to all the phones which have an active TCP connection at this time. When a new phone connects to the server, Micro-Blog checks if some queries fit the phones position and settings and pushes the relevant queries on that phone. The response of a query is directly send back to the originator if the HTTP connection is still open and it is also saved in the database so that the users are able to view the responses later.

3) *conclusion:* A study with 12 students showed that the idea of Micro-Blog in this form has still some problems, most students mentioned that the application had an unfriendly user interface, which lowered their motivation to use the program. Because of privacy concerns, most blogs generated in the study



Figure 4. The GUI of Connecto[3]

have been tagged as private. Most people liked the idea of replying to queries. Only three students disabled this feature because they were afraid of getting queries which don't match their interests.

C. Connecto

1) *Overview:* The idea of Connecto[3] is an "always on" tool, for sharing and tagging your location which runs on a windows mobile phone. It uses the GSM cell towers for positioning. The target of Connecto is to enhance the users social communication by sharing their status and location in an unobtrusive manner including automatically location sharing and status updates. The GUI of the application is shown in Figure 4

2) *Technological details:* The application shows context and localization information amongst small groups of friends on the display. Three types of data are shared by the program: the location information of the mobile phone, the time how long the telephone has been at the current location and the ringing profile of the phone. The location has to be defined by the user and it is shown as a label at the GUI which can be re-defined later.

Connecto has been developed for windows mobile smart-phones. It was programmed in C-Sharp and it uses GSM technology to locate the phone. This technology is adequate enough, because the locations are entered by the user itself and these are often locations like "work" or "home" which don't have to be very precise to be detected later. Connecto uses a backend with servers on which an SQL database is running. The communication with these servers is done over an GPRS connection with XML datasets. The data stored on the phone

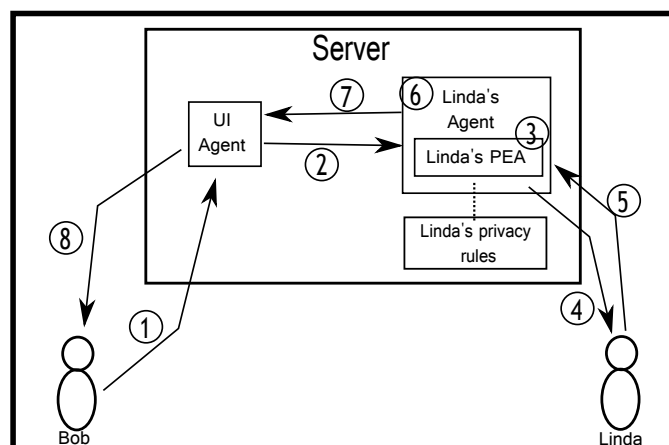


Figure 5. The query system of PeopleFinder

is uploaded to the server every seven minutes to save network traffic and keep the energy consumption low.

3) *Conclusion:* Connecto was used in a study to explore the practices around microblogging, location and status sharing within a social group. The participants used Connecto as an application which supported the ongoing "story" of conversations between single users and whole groups. The study showed that Connecto became an program for social exchange, enjoyment and friendship. The function to look up the current ringtone of others allowed the user to coordinate their calls. For example some users didn't call their friends, when they saw that they had activated the "silent" profile of their phone, because they mentioned that their friends were busy at the moment. Some couples used the localization information of their partner to ask their partner to pick something up on their way home from university when they saw that their partner had just left his working place. This shows that the use of this application improved the communication between individuals and lead to an more efficient use of time.

D. PeopleFinder

1) *Overview:* PeopleFinder [38] seems to be a regular location based application, which enables you to find out where other peoples are located and share your own location with them, but in contrast to most applications, PeopleFinder offers efficient and extensive security options. This is done by a security system that allows a easy creation and modification of policy based rules by the users. This system has been implemented because more and more people are concerned about their privacy while using mobile social applications. The feeling that everyone could know where you are at the moment makes a lot of people feeling insecure. So they tried to develop PeopleFinder to fit the peoples' growing need for privacy.

2) *Technological details:* The PeopleFinder application was developed for mobile phones using the windows mobile operating system. The application uses the GPS technology to locate the phone, or a GSM triangulation if GPS isn't available. The user of PeopleFinder are divided into two groups. One Group of users is called target users, these are users who share their location with others. The other group is called requesting

users, these are the users who are able to submit queries about the location of another user. A user can be a target and an requesting user at the same time. An example how the system works if a requesting user sends an query to find an target user can be viewed in Figure 5. The request submitted by the request user (called Bob here) is forwarded by the UI Agent of the target users (called Linda here) mobile device to her PeopleFinder Agent. The Agent checks whether the query matches the privacy rules which have been declared by Linda. If the query is accepted a notification is send on Linda's phone which then returns her current location. After that the location is processed by Linda's agent and then send to Bob's phone. At the end the forwarded result is shown on Bob's phone. If the query is denied by the Agent or if the target user's device couldn't be found an ambivalent message is shown on the requesting users phone. So the requesting user doesn't know whether his query has been denied or the device is offline.

The query processing may be complex under some circumstances, for example an user could define privacy rules which allow queries only at certain locations on weekdays. So the location will be received by the Agent also on the weekend, but won't be forwarded to the requesting user. The policies of the PeopleFinder application are defined by using a rule extension of the OWL language. This allows the policies to be restricted in many different ways. For example to locations, social relationships or calender activities.

3) *conclusion*: A study with 19 participants and another one with 60 participants showed that people are really concerned about their security and they have to think hard about articulating effective privacy preferences. The time the user needs to define accurate policies can be shortened by providing wizards or other detailed information on how the rules have to be generated. The studies also showed that people tend to soften their policies over time.

IV. MIDDLEWARE

While most people only think about the application and its design, some people had the idea to develop middleware on which mobile applications could be developed in an easy way by providing an API for third party programmers. These middlewares introduced in the following are two examples of them. They are focused on the development of the application itself, that means that the developers don't have to care about problems like communication with the sensors or the communication with the backend, because these problems are solved by the middleware.

A. MobiSoc

1) *Overview*: MobiSoc [7][16] is a middleware which is specialized on Mobile Social applications. It tries to meet the specific requirements of mobile social applications like privacy, low energy usage, the collection and management of social data and the collection and management of localization data. The interaction of the components can be viewed in Figure 6. MobiSoC is a central part of the architecture which manages the social states and provides a service API for programmers to communicate with the middleware. MobiSoc includes different

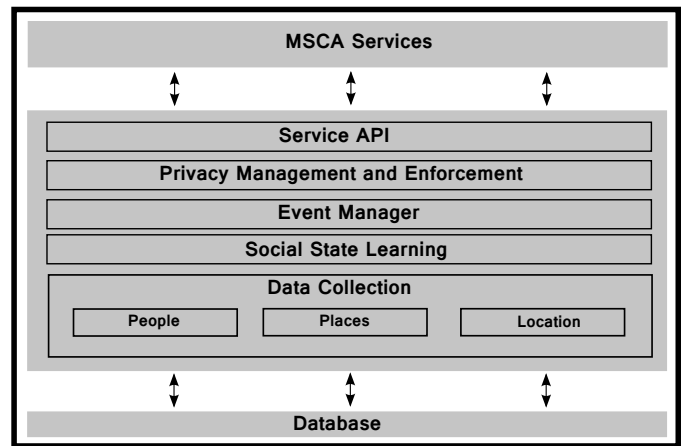


Figure 6. The modules of MobiSoc

submodules which are shown in Figure 6. The data collection module is splitted into three parts. The people module is designed to collect, store and update user profiles. The places module was made for collecting geographical data and maps and the location module controls the localization data of the applications.

The social state learning module are about gathering the information about the social state of a user. It is divided into submodules which offer specified algorithms and functions to analyze the data and collect information out of it. The event manager module is responsible for the communication between the middleware and the phones. This is done by a pull approach, so the mobile phones have to ask the middleware if there are new events available for them. The privacy module manages the privacy rules. Constraints can be made about location, events, profile data and social network data.

Each application developed on top of MobiSoc is splitted into two parts: a service party which is running on the server and communicates with the MobiSoc API and a thin client which communicates with the service over the internet. This means that the applications can't interact directly with the middleware, they have to communicate with their service.

2) *Applications*: Two exemplary applications are Tranzact and Clarissa to show the possibilities of the MobiSoc middleware. Tranzact is an application which allows the user to send place-based queries to others. For example a worker could be interested about the special offers in the cafeteria and sends a query which is received by the phones of the social contacts which are located at the cafeteria at the moment.

The other application is called Clarissa, it is a mobile social matching application. People can define when they have time and what interests they have, and then an matching score with other users who are available at the entered time is generated based on their affinities and interests which is then shown on the phone. The GUI of Clarissa is shown in Figure 7

B. MobiClique

The MobiClique [35] middleware differs from most applications and other middlewares because it builds an ad hoc network using bluetooth to communicate with other devices.



Figure 7. The GUI of Clariza[7]

But the data can be synchronized with the Facebook network too, whenever the phone is online. The bluetooth network of MobiClique is build using opportunistic connections between neighboring devices. That means that every phone makes an bluetooth scan periodically to find other nearby phones. Everytime another phone is found the device tries to start an interaction with the other phone. What sort of interaction is started depends on how the two users profiles share interests and if they have some previously defined relationships , for example a friendship. The users are alerted then and they can choose to make an data exchange if they want. In this case an exchange can be a lot of things, for example a friendship introduction or an content dissemination. MobiClique uses the Facebook [10] social network and the Facebook API. The data can be synchronized with Facebook whenever the phone is online.

1) *Applications:* Three prototypes of applications have been implemented on top of the MobiClique middleware. One application for mobile social networking shows the current set of phones in range with their social profiles. It also provides a GUI to add/remove friends and an option panel where alarms can be enabled/disabled when friends, friends of friends or members of the interest group are in range. The second application enables the user to send messages between a pair of users. A file attachment can be sent with the message. The third application is a newsgroup application similar to the traditional Usenet where multiple users can have a discussion. As on the messaging application file attachments can be attached to a message.

V. CONCLUSION

The paper showed that there are a lot of social applications on the market and in development, and that an definite categorization of this applications isn't easy because most applications have features which belong to different categories. There are also applications which try to connect categories,

for example blogging and social networking are such features which are offered in a single application.

The applications and studies also showed that most of the peoples concerns regarding privacy or the battery life can be met to a certain degree. People are willing to share their data if they feel secure and the use of the program appears to be safe. Because of that mobile social applications will be used more and more in the next years which leads to a growing impact of social applications in our daily life. The target is to save time and communicate in a more efficient way. For example finding new contacts or an adequate social occupation will be done more and more with applications, without having to call your friends. All things you have to know are accessible through the application at once.

REFERENCES

- [1] Aka-Aki, "Aka-aki." [Online]. Available: <http://www.aka-aki.com/>
- [2] A. Ankolekar, G. Szabo, Y. Luon, B. A. Huberman, D. Wilkinson, and F. Wu, "Friendlee: a mobile application for your social life," in *Friendlee*. Bonn, Germany: ACM, 2009, pp. 1–4. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1613858.1613893&coll=ACM&dl=ACM&CFID=58502448&CFTOKEN=39389252>
- [3] L. Barkhuus, B. Brown, M. Bell, S. Sherwood, M. Hall, and M. Chalmers, "From awareness to repartee: sharing location within social groups," in *Connecto*. Florence, Italy: ACM, 2008, pp. 497–506. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1357054.1357134&coll=ACM&dl=ACM&CFID=58502448&CFTOKEN=39389252>
- [4] N. Belloni, L. E. Holmquist, and J. Tholander, "See you on the subway: exploring mobile social software," in *See you on the subway: exploring mobile social software*. Boston, MA, USA: ACM, 2009, pp. 4543–4548. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1520340.1520697&coll=ACM&dl=ACM&CFID=58502448&CFTOKEN=39389252>
- [5] Belysio, "Belysio." [Online]. Available: <http://www.belysio.com/>
- [6] Bliin, "Bliin." [Online]. Available: <http://bliin.com/>
- [7] C. Borcea, A. Gupta, A. Kalra, Q. Jones, and L. Iftode, "The MobiSoC middleware for mobile social computing: challenges, design, and early experiences," in *The MobiSoC middleware for mobile social computing: challenges, design, and early experiences*. Innsbruck, Austria: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007, pp. 1–8. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1361492.1361526&coll=ACM&dl=ACM&CFID=58502448&CFTOKEN=39389252>
- [8] Buzzd, "Buzzd." [Online]. Available: <http://buzzd.com/>
- [9] CenceMe, "Cenceme." [Online]. Available: <http://cenceme.org/>
- [10] Facebook, "Facebook." [Online]. Available: <http://www.facebook.com>
- [11] FootprintHistory, "Footprint history." [Online]. Available: <http://www.footprinthistory.com/>
- [12] Friendticker, "Friendticker." [Online]. Available: <http://www.friendticker.de/>
- [13] S. Gaonkar, J. Li, R. R. Choudhury, L. Cox, and A. Schmidt, "Micro-Blog: sharing and querying content through mobile phones and social participation," in *Micro-Blog: sharing and querying content through mobile phones and social participation*. Breckenridge, CO, USA: ACM, 2008, pp. 174–186. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1378600.1378620&/coll=ACM&dl=ACM&CFID=58502448&CFTOKEN=39389252>
- [14] L. Gou, J. Kim, H. Chen, J. Collins, M. Goodman, X. L. Zhang, and C. L. Giles, "MobiSNA: a mobile video social network application," in *MobiSNA: a mobile video social network application*. Providence, Rhode Island: ACM, 2009, pp. 53–56. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1594139.1594155&coll=ACM&dl=ACM&CFID=58502448&CFTOKEN=39389252>
- [15] R. Grob, M. Kuhn, R. Wattenhofer, and M. Wirz, "Cluestr: mobile social networking for enhanced group communication," in *Cluestr*. Sanibel Island, Florida, USA: ACM, 2009, pp. 81–90. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1531674.1531686&coll=ACM&dl=ACM&CFID=58502448&CFTOKEN=39389252>

- [16] A. Gupta, A. Kalra, D. Boston, and C. Borcea, "MobiSoC: a middleware for mobile social computing applications," *Mob. Netw. Appl.*, vol. 14, no. 1, pp. 35–52, 2009. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1503496.1503500&coll=ACM&dl=ACM&CFID=58502448&CFTOKEN=39389252>
- [17] Icebrkr, "Icebrkr." [Online]. Available: <http://www.icebrkr.com/>
- [18] Ipoki, "Ipoki." [Online]. Available: <http://www.ipoki.com/>
- [19] Juicecaster, "Juicecaster." [Online]. Available: <http://www.juicecaster.com/>
- [20] S. K. Kane and P. V. Klasnja, "Supporting volunteer activities with mobile social software," in *Volunteer activities*. Boston, MA, USA: ACM, 2009, pp. 4567–4572. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1520340.1520701&coll=ACM&dl=ACM&CFID=58502448&CFTOKEN=39389252>
- [21] MapMyTracks, "Mapmytracks." [Online]. Available: <http://www.mapmytracks.com/>
- [22] MeetMoi, "Meetmoi." [Online]. Available: <http://www.meetmoi.com/>
- [23] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell, "Sensing meets mobile social networks: the design, implementation and evaluation of the CenceMe application," in *CenceMe*. Raleigh, NC, USA: ACM, 2008, pp. 337–350. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1460412.1460445&coll=ACM&dl=ACM&CFID=58502448&CFTOKEN=39389252>
- [24] MobiLuck, "Mobiluck." [Online]. Available: <http://www.mobiluck.com/>
- [25] Mobimii, "Mobimii." [Online]. Available: <http://mobimii.com/>
- [26] MocoSpace. [Online]. Available: <http://www.mocospace.com/>
- [27] Mologogo, "Mologogo." [Online]. Available: <http://www.mologogo.com/>
- [28] MyCityMate, "Mycitymate." [Online]. Available: <http://www.mycitymate.com/>
- [29] MySonar, "Mysonar." [Online]. Available: <http://www.mysonar.de/>
- [30] MySpace. [Online]. Available: <http://www.myspace.com>
- [31] Next2Friends, "Next2friends." [Online]. Available: <http://www.next2friends.com/>
- [32] K. O'Hara, M. Lipson, M. Jansen, A. Unger, H. Jeffries, and P. Macer, "Jukola: democratic music choice in a public space," in *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*. Cambridge, MA, USA: ACM, 2004, pp. 145–154. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=1013115.1013136>
- [33] Okud, "Okud." [Online]. Available: <http://www.okud.de/>
- [34] Ovalpath, "Ovalpath." [Online]. Available: <http://www.ovalpath.com/>
- [35] A. Pietiläinen, E. Oliver, J. LeBrun, G. Varghese, and C. Diot, "MobiClique: middleware for mobile social networking," in *MobiClique: middleware for mobile social networking*. Barcelona, Spain: ACM, 2009, pp. 49–54. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1592665.1592678&coll=ACM&dl=ACM&CFID=58502448&CFTOKEN=39389252>
- [36] Qeep, "Qeep." [Online]. Available: <http://www.qeep.de/>
- [37] Qik, "Qik." [Online]. Available: <http://www.qik.com/>
- [38] N. Sadeh, J. Hong, L. Cranor, I. Fette, P. Kelley, M. Prabaker, and J. Rao, "Understanding and capturing people's privacy policies in a mobile social networking application," *Personal Ubiquitous Comput.*, vol. 13, no. 6, pp. 401–412, 2009. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1569346.1569366&coll=ACM&dl=ACM&CFID=58502448&CFTOKEN=39389252>
- [39] Seesmic, "Seesmic." [Online]. Available: <http://seesmic.com/>
- [40] StudiVz, "Studivz." [Online]. Available: <http://www.studivz.de>
- [41] N. S. Tracker, "Nokia sports tracker." [Online]. Available: <http://sportstracker.nokia.com/>
- [42] Twitter, "Twitter." [Online]. Available: <http://www.twitter.com/>

Social Computing: an Overview

Manuel Deil – {firstname.lastname}@uni-ulm.de

Abstract—As mankind is reaching out to the aera of ubiquitous computing, computers influence not only our work but more and more dominate and support our whole life. We carry our cell phone to every place, twitter each wooly thought and become virtually friends with everybody proposed by facebook. This new trend in social life provides enormous possibilities as well as risks, no one so far eliminated or even thought about. Due to this fact, social computing unites a variety of different scientific disciplines trying to benefit from the new trend and simultaneously easing the risks.

This paper provides an overview on Social Computing as an emerging and interdisciplinary research field combining information technologies and social studies. Thus, Social Computing can be seen as bidirectional, mutual assistance between using IT to support social sciences and applying social knowledge to the world of IT. Investigating the behavior of groups, for example, helps to predict the spreading of pandemics, may reveal criminal groups, and can be used to obtain a healthy member structure in online communities. Furthermore, social interactions and collaborations raise the question about trust in E-commerce, wiki content, and safety of private information. Moreover, computer models can lead to more efficient economic activity and a better integration of social aspects. Recapitulatory, a conclusion measuring the advances and risks of the presented state-of-the-art techniques will be given, as well as an open-minded foresight for the future.

I. INTRODUCTION

SPEAKING of social computing, most people think of social networks like myspace and twitter, but this new research trend reaches far beyond social networks.

In a sense, it is a new research field with long history. Basically, social computing is a symbiosis of two schools of thoughts (Fig. 1) [19]:

- The first one centers on information-technology, speaking of social computing as social software and focussing on applying knowledge from social studies to conceptualize and improve the efficiency and performance of software.
- The second school, however, is more focused on social sciences and relatively new. Close to the new emerging web sciences, it emphasizes designing and evaluating cyber-physical systems, to study and manage social behaviors. The roots for this research trend can be found in the launch of TV and fax as first *on line* and *real time* systems.

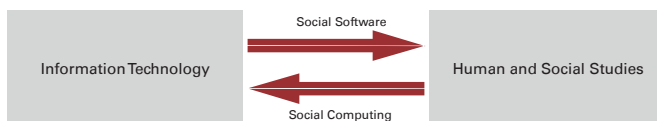


Figure 1. Two Schools of Thoughts: Social software and Social Computing [9]

Hence this year's SocialCom¹ was opened up by the steering chair with the words [9]: “*Social computing can be broadly defined as computational facilitation of social studies and human social dynamics as well as design and use of information and communication technologies that consider social context.*”

As one can see, it is not a single-sided media hype but rather a co-operation of, in various aspects totally different, disciplines trying to meet the ravages of time.

The paper will present several main research fields of social computing. Those are structured into Behavior of Groups (II-A), Social Collaboration and Trust (II-B), Integration of Social Aspects and Computer Models (II-C), and Further Research Fields (II-D). In the end a conclusion (III) measuring the advances and risks of the presented state-of-the-art techniques as well as an open-minded foresight for the future will be given.

II. RESEARCH FIELDS

It is generally accepted, that the Internet is the original motivation for and the natural consequence of Social Computing. While we still consider cyberspace being virtual today, we soon will have to change that perspective, as our real world is more and more held together by IT. Due to this change, a new understanding of human interaction is necessary, where traditional models fail today. The truly innovative aspect of Social Computing hereby is the collaboration between physicists, psychologists, computer scientists and much more, together developing new methods and models. Due to that large-scale interdisciplinarity, it would go beyond the scope of this work to deeply introduce all research fields in Social Computing. For this reason it is limited to some of the main research fields, which will be explained consecutively.

A. Behavior of Groups

Nowadays, investigating and predicting the behavior of groups is much easier and more precise than just 10 years ago. With most people using cell phones and internet everywhere, unimaginable data volumes on human mobility are recorded every minute and can be used for different objectives.

1) *Prediction of Epidemics:* Since mankind is threatened by pandemics like avian flu (H5N1) and swine flue (H1N1), the call for reliable prediction models is getting more and more intense. Such models on the one hand can help to contain such diseases before they can rise to pandemics and on the other hand aid to organize the development of adequate vaccines. At the Max-Planck-Institute for Dynamics and Self-Organization, a new analysis model on human mobility has been developed to

¹SocialCom: IEEE International Conference on Social Computing

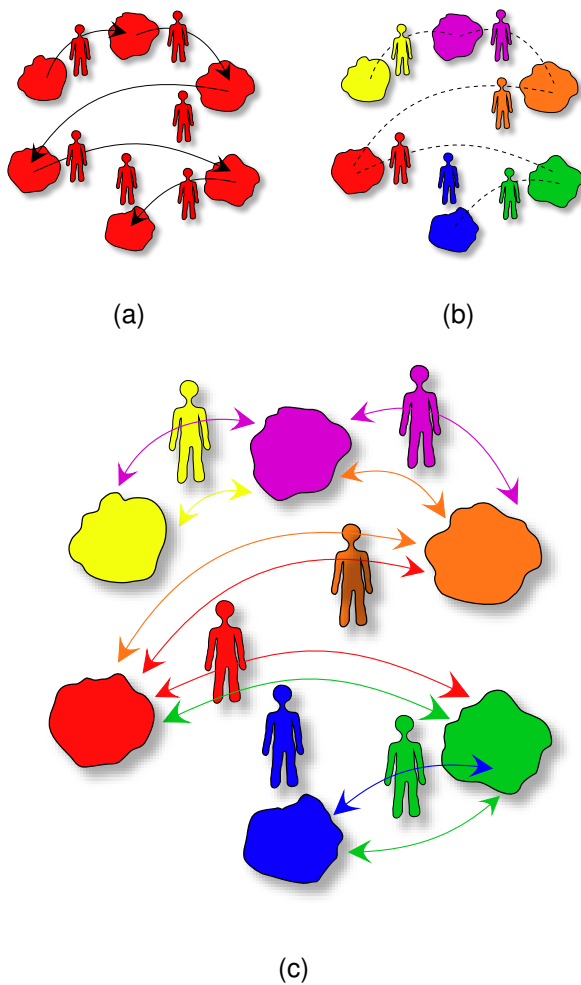


Figure 2. Mobility in spatial epidemics models [1]: (a) diffusive dispersal: unbounded random movements; (b) direct coupling: effective force of infection between populations (c) explicit incorporation of bidirectional host movements. Color code symbolizes bidirectional links due to hosts from different nodes.

predict how such spatial disease dynamics are spreading around the world [1]. The new approach is based on two prominent theories:

- *diffusive dispersal* (Fig. 2(a)) emanates human hosts move randomly and chaotically around, which results in a reaction-diffusion model.
- *direct coupling* (Fig. 2(b)) relies on the fact that humans have places they often return to (e.g. home, work) and therefore the effective force of infection between distinct populations is emulated.

Both theories exclude important aspects of human mobility, fundamentally the bidirectional movements between home and dislocated locations. Therefore, taking differences and similarities in other topologies like networks or lattices into account, much more precise forecasts can be developed. The model in [1] is designed to count on bidirectional host movements, especially the movements from center nodes (*home*) to distant nodes (*accessible destinations*). Such movements almost always include a return to the center node before moving to another distant node (Fig. 2(c)). Taking this fact on human mobility

Table I
PEOPLE ATTENDING EVENTS e_i [8]

event	persons	color-code
e_0	H C B E A	red
e_1	C B D H F	yellow
e_2	H C A B G F	blue
e_3	G A F B E	black

behavior into account, much more precise predictions can be made. With more data on human mobility becoming available for research in the future, the work of Belik et al could be used as a framework to manage and organize such data, and develop more and more exact models for future spatial disease dynamics.

However, computer models may help to avoid future pandemics, but yet are not precise enough to curtail all epidemics – but they promote a direction for the future.

2) *Detection of Criminal Networks*: A completely different kind of research on the behavior of groups is the *ISI* (*IEEE Intelligence and Security Informatics Conference*)². The *ISI* was founded in response to the terrorist attacks on September 11, 2001. Remarkably, these attacks were basically organized using social networks [8]. Thus, *ISI* employs collaborative mining in multiple social networks to facilitate obtaining data for criminal group discovery – so far, this stays predominantly a manual process. First of all, to solve that problem the problem itself has to be defined accurately:

At the Simon Fraser University in Canada such an adequate definition meeting the demands was developed [6]: seeing a social network as a multi-graph $S=(P,R)$ (Fig. 3) with P being a set of persons and R a set of relationships, a structural environment is provided. Additionally, an event e_i may consist of several such relationships, indicating that the linked persons took part in the event (Table I). In the problem definition, it is negligible what an event exactly is, hence, an *event* can constitute from chat history, a tagged photo, an email, or even a police criminal record. As a whole, the problem consists of a set of n social networks $S_i=(P_i,R_i)$ and a set Q of i persons ($Q = q_1, q_2, \dots, q_i$), who the police could take into custody. Now, people associated to the ones in Q , ordered by cooperation distance, are considered suspects in the investigated crime. Based on this formal definition and several advanced algorithms (e.g., multi-agent models taken from game theory) a framework emerged, providing automated network data analysis and extracting multiple social network data to transaction datasets.

3) *Weighting to Identify Member Roles in Online Communities*: Going back to online communities, a different challenge can be revealed: identifying key members and superficial individuals. With more and more people joining online communities, it has become increasingly hard work to manually review user comments in order to provide a healthy community environment. Computer-based user categorization can simplify this task, but requires high-level knowledge on human-human and human-information relationships [15]. The

²ISI: <http://www.isiconference.org/>

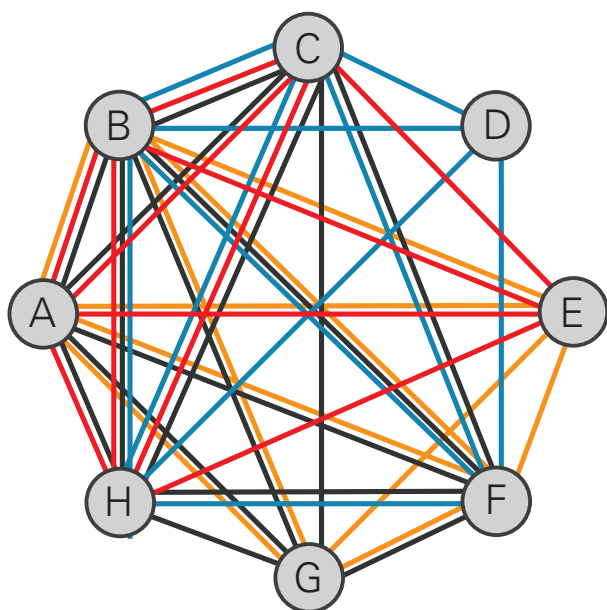


Figure 3. Social network as multi-graph for persons and events in Table I [8]

strength of such relationships can vary from *very weak* to *very strong* – the intricacy increases exponentially with the number of members. Moreover, members can act indirectly through other members, which also has to be taken into account to judge a member's value as a whole. Such indirect actions, for example, can be information a member shared with another, who shared the same information with someone else. Depending on the community, different roles are required to keep the community process alive and maintain quality at a high level. A software bulletin board may require software experts, while a health support group depends on a combination of information providers and empathetic listeners. Acquiring the needed user roles helps to rate the quality of posts and to eliminate those undermining the community's reputation. One can categorize three kinds of roles:

- *Relationship-based roles* (e.g. gatekeepers)
- *Behavior-based roles* (e.g. pollinators³, spammers⁴, conversationalists⁵)
- Combination of the two

Focussing on an open discussion bulletin board, one can identify two roles keeping up the community sense:

- *Leaders*: members featuring information input, consistency and cohesiveness
- *Motivators*: members supporting the conversation

At first glance, members with high contribution rates seem to be equivalent to the aforementioned *leaders* and *motivators*.

³pollinator: a user with a high number of days active and low posts per thread ratio [7]

⁴spammer: a user with a moderate to high number of days active and almost entirely initiating threads which then receive no follow-up messages from this author [7]

⁵conversationalist: a user who initiates about as many threads as he/she replies to and shows about the same posts per thread ratio on both initiated and non-initiated threads [7]

However, a closer look reveals, a criterion is necessary to filter out so-called *chatters*. According to Nolker [15], chatters are characterized by the following:

Two community members contribute a lot in a protracted discussion, but hardly submit postings on other discussions. According to their postings count they would inherit key member qualities, but as they are not contributing to the community conversation as a whole, they would be considered chatters in the common sense – regardless of the quality of their posts. To rate posts and posters, a deeper understanding of behavior-based measures is necessary:

- *Degree*: the number of conversations that a member is engaged in or the number of members a member has conversed with.
- *Betweenness*: the number of pairs of other members who can converse with each other indirectly through a member with shortest relay.
- *Closeness*: average conversation distance between a member and all the others in the community.

The importance of a poster compared to the thread and other posts can be determined using a well-known method (*TF*IDF*: Term Frequency Inverse Document Frequency) for finding key terms *T* in text documents *D*. The basic idea of *TF*IDF* is that the importance of *T* to *D* is proportional to the frequency of *T* in *D* and inversely proportionally to the total number of *D* that *T* appears in. The resulting quality of keywords can be applied to rate posts of a user, and summing up the posts rating leads to the implicit importance of the member for the community. The presented approach from the *IEEE/ACM International Conference on Web Intelligence* [15] was successfully used to determine and filter high volume users in a Usenet Bulletin Board in 2005 – but still awaits its great aperture.

Considered as a whole, one has to admit, that no matter which quality the models on human behavior in groups have, they still stay hypothetical models. They will help understand various mysteries of human group behavior, but do not suffice to make accurate predictions without manual engagement.

B. Social Collaboration and Trust

Web 2.0 has connected people in a way no one ever expected. According to Alexa⁶, as of December 2009, Facebook and Youtube rank above Wikipedia in the “*top 500 sites on the web*”, based on daily visitors and pageviews. This trend brings deviant needs with it, no one thought about in detail up to now. For example, facebook with more than 350 million active users⁷ has a high number of already deceased people with still available profiles. This morbid fact alone would be unremarkable – but for a short time, a facebook account now can be marked as *memorial*, allowing users to become “*virtually friends*” with deceased people [5]. Examples like the aforementioned, point out the necessity of in-depth studies on how to deal with such problems. Speaking of Social Collaboration, Kittur et al [10] set up a definition on Social Computing in three words:

⁶Alexa – the web information company: <http://www.alexa.com>

⁷Facebook Statistics: <http://www.facebook.com/press/info.php?statistics>

- *Connectivity*: the medium how people are connected to each other plays an enormous factor transforming and transferring the information. SMS⁸ and twitter⁹ limit the message-length. This leads to the formation of new languages of abbreviations and acronyms¹⁰. The same applies to phone, email, instant messaging, chats, blogs and much more.
- *Collaboration*: 300 BC Aristoteles already said: “*The whole is greater than the sum of its parts.*”[17] According to that quote, collaborative filtering, trust, knowledge sharing, and competitive collaborative learning fall into this category.
- *Community*: grouping and clustering people represents the relations between them. These networks can evolve out of spatial closeness or functional means.

1) *Trust in Wiki Content*: A well-known example for community collaboration is *Wikipedia*¹¹, an online open source encyclopedia. It is a set of tools for collaborative authoring of tagged hypertext content, enriched with version control and user feedback [12]. Users can contribute their knowledge by building a structured hypertext article on a topic from scratch. Quality is implicitly ensured through user feedback and the version control system which allows undoing changes and reverting to previous versions, if necessary. Despite its efforts, quality and credibility, Wikipedia is frequently criticised when erroneous entries get exposed and lead to protests. A prominent German example for such an erroneous entry is the story of an anonym journalist, adding an additional first name to the name of German Secretary of Commerce *Freiherr von Guttenberg*¹². As widely accepted newspapers cited the wikipedia article, the newspaper articles were referenced in wikipedia – substantiating false facts in detail. A comparison between *Encyclopedia Britannica* and *Wikipedia*, performed in 2005 due to such scandals, revealed, that *Wikipedia* contained about 30 percent more errors [12]. Hence, according to Kittur et al [11], the risks associated with the usage of Wikipedia among other things are:

1. *Accuracy*: Not knowing which content is accurate; often exacerbated by lack of references.
2. *Motives*: Not knowing the motives of editors, who may be biased for various reasons.
3. *Expertise*: Not knowing the expertise of editors.
4. *Stability*: Not knowing the stability of an article and how much it has changed since the last viewing.
5. *Coverage*: Partially factional coverage of topics.
6. *Sources*: Cited information may come from hidden or non-independent source

Despite all risks, for most people *Wikipedia* stays the first work of reference – although in case of scientific research it is sufficient at most for getting a quick overview to start on a topic.

2) *Trust Is in the Eye of the Beholder*: The digression on *Wikipedia*’s credibility gap shows very clearly the need for

trust. But “*Trust Is in the Eye of the Beholder*”[4] – the title of the corresponding article from DeFigueiredo describes best why the abstract term *trust* has not got a common meaning in general linguistic usage. So at first, when do we need trust? Trust helps us to know how to react in interactions with other people. eBay’s rating system is a good example for trust using reputation functions to quantify trust. There are two kinds of such reputations:

- *Consensus-based reputation*: all users have to agree on the trustworthiness of another party.
- *Personalized reputation*: each person can control how much she trusts the other party, independently from others.

Up to now, the common way of valuing trust uses reputation functions mapping trust to numbers like school grades. This approach leads to problems like how much one should trust the friend of a friend and the comparison of interpersonal trust ratings.

Almost all people tend to confuse *trust* and *trustworthiness*: they think trust is a characteristic of a person’s trustworthiness and everyone would agree with their classification if they had the same level of knowledge on that person. But this is a false conclusion – person A may trust person B, even though she is not trustworthy and person A may distrust person C even though she *is* trustworthy. Trust incorporates two roles – a *trusting* and a *trusted* role. Normally these roles are performed by distinct parties. Furthermore trust is *domain-specific*: you can trust a person to take care of your finances, but not to be on time for a meeting. Emanating from trust as a quantifiable value on a single scale, one can totally order his trust for the same domain and clear out inconsistencies. For example if Alice claims (all for the same domain):

- Alice trusts Bob more than Charlie;
- Alice trusts Charlie more than Derek; and,
- Alice trusts Derek more than Bob.

Pointing out the inconsistency to Alice, one assumes she will revise her trust scale, whereas it is of no concern whether trust is represented by a binary value or a number in a given intervall. Getting more formal, trust can be incorporated by an annotated direct graph $G = (V, E)$, where each vertex V is an agent and each directed edge in E has an associated trust value. This construct is also known as a *reputation graph* [4].

Such a reputation graph (Fig. 4) does not have to be complete – if there is no edge between two vertices, the value of trust between these two agents is unspecified. Combining several reputation graphs to a squence leads to a *world*. The first graph hereby is the *direct experience graph* whereas the following graphs are the *1-indirect graph*, *2-indirect graph*, and so on. The indirect graphs indicate the level of reliance on the other person’s judgement.

Before finally defining *trust*, we need to establish the personal *trust threshold* of an agent: all other agents whose trust values from the current agent are above the threshold, are **trusted** – otherwise they are **untrusted**. This intuitive understanding of trust is used in every-day interactions with others. Many websites and peer-to-peer networks attempt to clone that

⁸SMS: Short Message Service

⁹twitter: <http://www.twitter.com>

¹⁰http://simple.wikipedia.org/wiki/List_of_Internet_slang_words/

¹¹Wikipedia: <http://www.wikipedia.org>

¹²wikipedia scandal: <http://www.sueddeutsche.de/kultur/278/457933/text/>

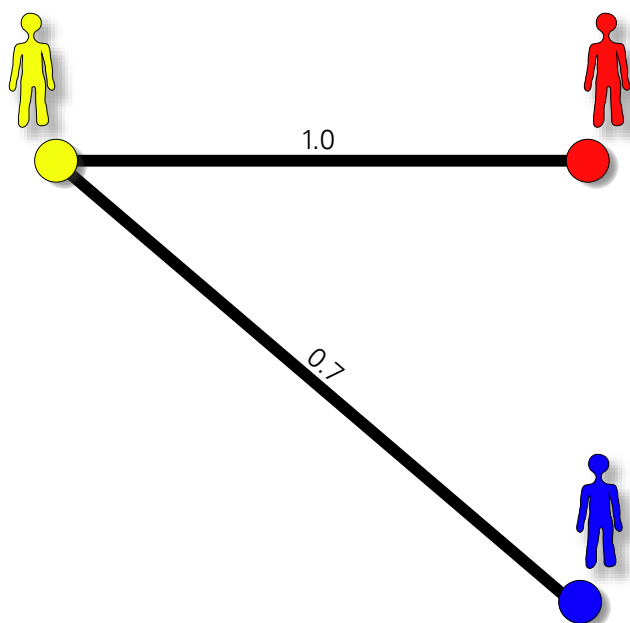


Figure 4. Reputation Graph: $G = (V, E)$

concept – ignoring the fact, that a single universal trust rating, not protected from manipulation, is not as useful to the end user as it could be. Such a trust rating system needs to change the assigned ratings to an individual, depending on the domain and on who is asking the question in order to resist manipulation. Going back to the establishment of trust in E-Commerce, at the moment two network effects solve such trust problems:

- obtaining the peer's earlier transactions information
- warning other nodes when a node is abusing trust

The level of trust in a node increases with the amount of information available on its behavior. Therefore, network density and measurement directly relate to trust [20].

3) *Privacy in Social Computing*: Another aspect of trust is *privacy in Social Computing*, because the more a user trusts a service providing sufficient trust mechanisms, the more he will use the service / share information. To provide users' privacy, system designers would ideally be aware of two aspects [14]:

- *users' perceptions*
- *real risks*

Research on the field of privacy revealed many of the more obvious threats associated with user access control, especially in the domains of ubiquitous and context-aware computing. Private data might be concealed (e.g. providing the users location, but not the identity). In a compatible context, in the sense of having the right background information, this might be enough to reveal the full identity using backtracking. Threats on user privacy in Social Computing Systems can be consolidated in the following categories:

1. *Inappropriate use by administrators*: For example, the system admin sells personal data without permission.
2. *Legal Obligations*: The system admin is forced by an organization such as the police to reveal personal data.

3. *Inadequate Security*: The system uses weak security tools which can be by-passed by unauthorized others.
4. *Designed Invasion (Poor Features)*: For example, a cell phone application that reveals location to friends, but does this without informing the user or providing control of this feature.
5. *Social Inference through lack of Entropy*: Provided concealed data can be used to conclude private data (See backtracking example above).
6. *Social Inference through Persistent User Observation*: For example, Bob is often in Alice's office. Therefore, their relationship must be romantic.
7. *Social Leveraging of Privileged Data*: For example, David cannot access my location, but Jane can. David asks Jane for my location.

At the moment, it is difficult to technically provide an appropriate level of privacy protection. Especially, there is no way to control what data is transferred outside the Social Computing system and to limit the associated abilities or to predict potential threats. A study by the New York University Polytechnic Institute on community-specific privacy concerns figured out, that users not only adopt more *restrictive information sharing settings* but also reduce the amount of shared information in case of a possible threat to their personal data [16]. Such possible threats frequently get revealed by system users and are broadcasted to every linked user using the *FOAF (Friend Of A Friend) ontology*¹³.

Speaking of Social Collaboration and Trust in Social Computing Systems, today various techniques try to provide trust to users. Their credibility depends on the used security features. Digitally signed email and tax returns, for example, using certificates from firms like VeriSign¹⁴, are attributed a high credibility.

C. Integration of Social Aspects and Computer Models

1) *From Pervasive to Social Computing*: In the early days of *Pervasive Computing* a change from classic desktop-HMI¹⁵ towards many crosslinked computing and networking resources, embedded in the environment, was envisioned. These visions by Mark Weiser [21] and his colleagues nowadays have partly become reality with wireless network devices (e.g. mobile phones, smart phones, PDAs, laptops, netbooks, etc.) and the rise of Web 2.0. The use of social network services currently takes place in the internet – whether mobile or stationary. *Pervasive Social Computing (PSC)* claims to be the paradigm when social services detach from the internet to become pervasive themselves. In contrast to *Pervasive Computing*, aiming at individual interaction integrated in the physical environment, PSC concentrates on providing assistance for tasks depending on social relations between users [2].

A short scenario to demonstrate what PSC is:

After a day of work, you want to go for a tennis game, but instead of directly contacting one of your friends, you use an application on your PDA¹⁶ to advertise your task, expecting

¹³FOAF Project: <http://www.foaf-project.org/>

¹⁴VeriSign Inc.: <http://www.verisign.com/>

¹⁵HMI: Human-Machine-Interaction

¹⁶PDA: Personal Digital Assistant

the system to find a good match for it. Based on your social network, and on the other requests for playing tennis advertised in the system, the application will answer your request, by suggesting the opponent being available and satisfying the stated requirements (e.g., location).

To fulfill these requests, the following tasks have to be dealt with:

- *Semantic Specification and Reasoning on User Tasks:*
A mobile user submits his interest on an activity (e.g., playing tennis tomorrow) expecting the system finding the *best matching user* for it. This can include selection on other common interests.
- *Social-based Matching of User Tasks:*
The transaction of User Tasks requires a mix of different self-evolving strategies for creating new plans and resolving conflicts on social preferences and similarities.
- *Dynmaic User Task Discovery:*
In contrast to internet-services with access to world-wide user-databases, pervasive environments cannot benefit from a huge user group as their connectivity and other parameters may be limited, and thus have to use different methods reaching enough people to determine a *good match*, e.g. the *FOAF ontology*.

2) *Computer models:* Despite supporting the accomplishment of tasks in network-like architectures, Social Computing can aid Social Studies with computer models: supplementing the traditional means of passive observations and statistical methods with models of artificial societies. In such a controlled environment disruptive factors can be avoided in order to gain optimal results. Wang [18] proposed a way to realize such models, in 2007. Orientated at the well-tried three-stage approach (modeling, analysis, control), he developed the ACP-Approach:

- *Artificial societies for modeling*
- *Computational experiments for analysis*
- *Parallel execution for control*

With approximation at high accuracy and the parallel execution of several artificial systems and the real system, the ACP-Approach provides techniques to sue the artificial systems to emulate the real system in a way, to use their behavior to ameliorate and increase performance back in the real system. This constantly repeating process of crossover mirroring can be used in any kind of workflow in economics as well as many other problems.

D. Further Research Fields

With Social Computing aiming at non-traditional informatics, the interdisciplinarity in research fields reaches physics, psychology, pedagogy, health, economics, and much more. In this section a brief overview will be given, how pedagogy, health and economics can benefit from Social Computing:

1) *Pedagogy:* The use of Social Computing to enhance lifelong learning is not clarified yet, but in particular young people entering higher education have integrated web 2.0 applications seamlessly in their everyday life and demand the educational institutions to support their digital learning

needs [3]. There are at least three different ways to deploy Social Computing tools in higher education:

1. *institutional tool:* replacing or extending the virtual learning environment to improve administrative procedures and representing the education institutions.
2. *communication tool:* among students and between students and teachers, supporting also the exchange of knowledge and material, but mainly creating an environment of understanding and assistance.
3. *methodological or didactic tool:* virtual courses and classes, within certain disciplines, with a focus on improving, facilitating and enhancing knowledge acquisition.

Social Computing tools provide a huge potential for higher education institutions to meet the requirements of changing learning contexts and to support lifelong learning. Furthermore, these tools can help assisting current and prospective students. Concretely speaking of *learning*, Social Computing applications inherit a massive potential to enhance the learning process – both single-learners and lively, productive communities of learners and researchers for the creation and exchange of knowledge. The biggest threat to effective learning today are inadequate solutions, like childish 3D-Worlds with animated swings intended to provide IT students a better understanding of abstract problems.

2) *Health:* With the rise of web 2.0 applications several on-line services emerged to provide help on health and healthcare:

- Google health¹⁷
- Microsoft HealthVault¹⁸
- PatientsLikeMe¹⁹

These services provide users the possibility to save all his medical data (diseases, allergies, currently occupied medications and other personal data). The user can search for doctors and hospitals according to his medical background and location, and share his profile data with the hospital staff. As the service is not limited to ask doctors for advice, one can release his data for specific sites and organizations. The *Share ePillBox*²⁰, for example, automatically sends new pills when the former ones should be consumed. Furthermore, one can also calculate his personal risk of a heart attack. Whether one would like to share his whole medical life with Google or Microsoft, everyone has to decide for himself, as the security and privacy is not comparable to the official medical records. At Google all this data is accessible with the normal email-password – trackable with a simple key-logger²¹ in an internet cafe. A completely different approach in terms of Social Computing aided health is the support of older people's social interactions, where social networks can be used as *health feedback displays* [13]. Social networks in this context are models on interactions between people in the real world and not to be mixed up with online social communities like facebook or MySpace. In 2005, at the Intel Corporation²² prototypes were developed and tested in the

¹⁷Google health: <https://www.google.com/health>

¹⁸Microsoft HealthVault: <http://www.healthvault.com/>

¹⁹PatientsLikeMe: <http://www.patientslikeme.com/>

²⁰Share ePillBox: <http://epillbox.info/>

²¹key-logger: a tiny programm recording every typed character

²²Intel Corporation: <http://www.intel.com/>

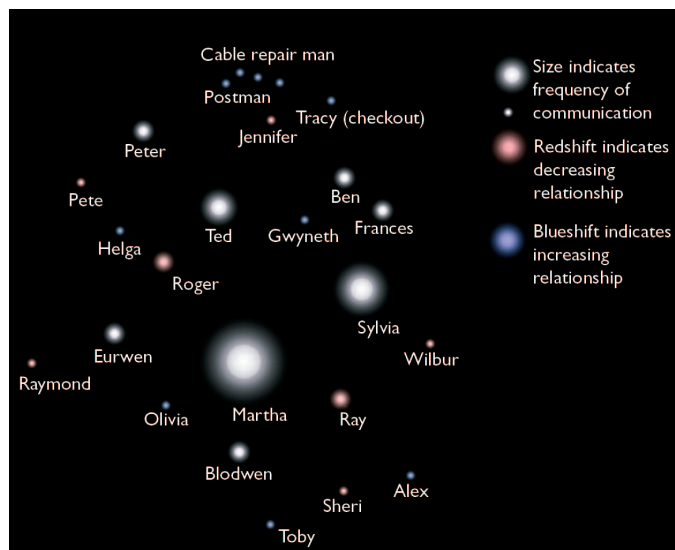


Figure 5. Conceptual model depicting trends of social interaction. Inward movement reflects greater interaction with the elder [13].

homes of older adults and their caregivers. An example for such an prototype is a solar display like shown in Fig. 5. It provides a conceptual model on applying such social networks to real life. The egocentric social network relies on sensor (tracked by phone or infrared badges) and interview data, to demonstrate interaction trends with friends and family. Capturing trends in changing interaction patterns can be used to reveal downward trends early and to intervene isolation and depression before they are manifested. The main challenge for such systems is not the data analysis as one may assume, but the design for people who are not used to IT systems.

3) *Economics*: The new market of web 2.0 applications and online communities for some people is a self-filling gold mine. The facebook application farmville²³, for example, makes a lot of money everyday with virtual goods. Users can pay by cash to get bonus points for special offers. With farmville only having fix costs of their webserver and developers, it has a high profit margin. But Social Computing is not limited to such new business fields and can also be applied to more traditional business fields, like the financial sector. At the Tianjin University of Finance & Economics an agent-based model for interactions between irrational and rational investors was developed, that may possibly drive irrational investors out of the market, in the near future [22]. Their ABC method is based on five steps:

1. *Build the conceptual model*:
a standard behavioral finance model with the freedom to use heterogeneous and time-varied investor behaviors.
2. *Design the artificial stock market's architecture*:
clear definition of the behavior of heterogeneous investors, tradable assets' risk and return characteristics.
3. *Code and run the ASM program*:
choosing a proper simulation platform, coding and debugging the ASM²⁴ architecture to generate simulated stock market

²³farmville: <http://farmville.com/>

²⁴ASM: Artificial Stock Market

data.

4. *Examine the ASM data*:
process raw data using financial econometrical methods
5. *Presenting new behavioral finance theories*:
extension or development of behavioral finance theories

The introduced ABC method is intended to gain insights how investors interact under given circumstances and not to simulate real stock markets. That is where the limits of such models become clear. they can be used for better understanding and surely for some predictions, but do not achieve the complexity a die-hard broker has in mind.

III. CONCLUSION

Social Computing opened up possibilities to a new dimension of computer-based interaction. Creating and sharing information has never been easier, as well as reaching potential customers. The new, more user-friendly tools of Web 2.0 boost the number of users being able to create and share content and social applications. The decentralization leads to new innovations, as people all over the world can get involved in such projects. Furthermore, Social Computing holds a considerably potential of opportunities for both research and business related technology.

At the moment, no one can predict the complete extend Social Computing contains. The most potential, as well as risk in my opinion, lies in analyzing and predicting the behavior of groups. Organizations and governments manipulate information as well as terrorist networks use social communities to plan world wide assaults and endanger the world we currently enjoy. On the other hand, gained data on human mobility can help to fight epidemics and terrorist actions in sufficient time. Additionally, self-evolving social communities may get along without much administration – expunging the associated risk of abuse. Our private data (name, birthday, gender, preferences and purchase behavior) becomes more and more equivalent to cash, enabling the companies “paid” in that currency to harvest our data for personalized ads and a possibly manipulation of our perception. Currently, it is not clear how the balance sheet will turn out in the end. Conventional media like newspapers are threatened with extinction, due to the “all-for-free”-opensource thought of the web, and with them investigated news. Meanwhile, state borders fade out in the web and the whole world is getting one great network-linked community collaboratively creating Wikipedia articles and sharing flickr-photos.

In my opinion, the most remarkable fact of Social Computing beyond all risks and possibilities, is the new trend away from **limited experts** towards **multidisciplinary scientists**, upgrading the digital world to the next level, which is in complete contrast to the human evolution of the last 300 years.

REFERENCES

- [1] V. Belik, T. Geisel, and D. Brockmann, “The impact of human mobility on spatial disease dynamics,” *Computational Science and Engineering, IEEE International Conference on*, vol. 4, pp. 932–935, 2009.
- [2] S. Ben Mokhtar and L. Capra, “From pervasive to social computing: algorithms and deployments,” in *ICPS'09: Proceedings of the 2009 international conference on Pervasive services*. New York, NY, USA: ACM, 2009, p. 169–178.

- [3] Y. P. Christine Redecker, Kirsti AlaMutka, "Learning 2.0—the use of social computing to enhance lifelong learning," *EADTU Annual Conference Poitiers (France)*, 2008. [Online]. Available: <http://is.jrc.ec.europa.eu/pages/EADTUpaperFINAL.pdf.pdf>
- [4] D. DeFigueiredo, E. Barr, and S. F. Wu, "Trust is in the eye of the beholder," *Computational Science and Engineering, IEEE International Conference on*, vol. 3, pp. 100–108, 2009.
- [5] J. M. DeGroot, *Reconnecting with the Dead via Facebook: Examining Transcorporeal Communication as a Way to Maintain Relationships*. Athens Gen. Collections, 2009. [Online]. Available: http://rave.ohiolink.edu/etdc/view?acc_num=ohiou1241547393
- [6] A. M. Fard and M. Ester, "Collaborative mining in multiple social networks data for criminal group discovery," *IEEE International Conference on Computational Science and Engineering*, vol. 4, pp. 582–587, 2009.
- [7] M. S. Fernanda B. Viegas, "Newsgroup crowds and authorlines: Visualizing the activity of individuals in conversational cyberspaces," *HICSS*, 2004. [Online]. Available: <http://www.connectedaction.net/wp-content/uploads/2009/08/>
- [8] D. Z. Hsinchun Chen, Fei-Yue Wang, "Intelligence and security informatics for homeland security information communication transportation," *IEEE Transactions on intelligent Transportation Systems*, vol. 5, pp. 329–341, 2004. [Online]. Available: <http://tiny.cc/2q8Ec>
- [9] L. T. Y. Justin Zhan, "Message from the socialcom 2009 steering chairs-volume 4," *IEEE International Conference on Computational Science and Engineering*, vol. 4, p. xxiii, 2009.
- [10] I. King, J. Li, and K. T. Chan, "A brief survey of computational approaches in social computing," in *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, June 2009, pp. 1625–1632.
- [11] A. Kittur, B. Suh, and E. H. Chi, "Can you ever trust a wiki? impacting perceived trustworthiness in wikipedia," in *CSCW'08: Proceedings of the ACM 2008 conference on Computer supported cooperative work*. New York, NY, USA: ACM, 2008, p. 477–480.
- [12] A. B. W. Manoj Parameswaran, "Social computing: An overview," *Communications of the Association for Information Systems*, vol. 19, p. 20, 2007. [Online]. Available: http://cism.mcombs.utexas.edu/attachments/105_Social_Computing_An_Overview.pdf
- [13] M. Morris, "Social networks as health feedback displays," *Internet Computing, IEEE*, vol. 9, no. 5, pp. 29–37, Sept.-Oct. 2005.
- [14] S. Motahari, C. Manikopoulos, R. Hiltz, and Q. Jones, "Seven privacy worries in ubiquitous social computing," in *SOUPS'07: Proceedings of the 3rd symposium on Usable privacy and security*. New York, NY, USA: ACM, 2007, p. 171–172.
- [15] R. Nölker and L. Zhou, "Social computing and weighting to identify member roles in online communities," in *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on*, Sept. 2005, pp. 87–93.
- [16] O. Nov and S. Wattal, "Social computing privacy concerns: antecedents and effects," in *CHI'09: Proceedings of the 27th international conference on Human factors in computing systems*. New York, NY, USA: ACM, 2009, p. 333–336.
- [17] T. Scaltsas, *Is a Whole Identical to its Parts?* SCALTSAS Mind., 1990.
- [18] F.-Y. Wang, "Toward a paradigm shift in social computing: The acp approach," *Intelligent Systems, IEEE*, vol. 22, pp. 65–67, Sept.-Oct. 2007.
- [19] —, "Social computing: Fundamentals and applications," *Intelligence and Security Informatics, 2008. ISI 2008. IEEE International Conference on 17-20 June 2008*, pp. xxxv–xxxviii, June 2008.
- [20] X. Wang, F. Liu, R. Yang, and F. Xie, "Social computing-based trust establishment in e-commerce," in *Internet Computing in Science and Engineering, 2008. ICICSE'08. International Conference on*, Jan. 2008, pp. 246–249.
- [21] M. WEISER, "The computer for the twenty-first century," *Scientific American*, vol. 265, no. 3, pp. 115–124, 1991. [Online]. Available: <http://ci.nii.ac.jp/naid/10016933747/en/>
- [22] Y. Zhang and W. Zhang, "Can irrational investors survive? a social-computing perspective," *Intelligent Systems, IEEE*, vol. 22, no. 5, pp. 58–64, Sept.-Oct. 2007.