

UNIVERSITY OF TWENTE.

---

# **Towards automated truancy detection**

Master of Science Thesis

Robbert H.J. Ottenhof

University of Twente  
Department of Electronic Engineering, Mathematics and Computer Science  
Distributed and Embedded Security Group

*Enschede – The Netherlands*

August 29<sup>th</sup>, 2011

## Abstract

Truancy and school dropout are important problems in The Netherlands. It is government policy to reduce the number of dropouts in the next five years. According to the Dutch Ministry of Education, truancy is an indicator for possible future dropouts. Decreasing truancy would therefore be one of the possible actions that might reduce the number of dropouts.

However, in order to tackle truancy, it needs to be registered first. Due to the increasing digitalization of education, class attendance registration procedures have become more cumbersome for a lot of teachers, which results in a decrease in truancy registration. This does not support the goal of the Government.

From this problem, the main research question follows: "How to automate the process of truancy registration?" From this main research question, two sub research questions have been formulated. The first question is: "How do students use modern media at school?" The second question is: "How can this behavior be registered automatically?"

To answer the first sub question, a survey has been conducted among a convenience sample of schools for secondary education in The Netherlands. To answer the second sub question, a software tool has been developed and a practical experiment has been performed with this tool at a secondary school in The Netherlands.

The figures from the survey show that 99% of the students in secondary education have a mobile phone, so that almost everyone has a personal device capable of generating data traffic. Already a third of the students uses internet with their mobile phone. Furthermore, social network sites are popular among secondary school youth; 96% has an account with a social network site like Hyves or Facebook. The practical experiment with the software tool shows that it is possible to detect the presence of students based on their online behavior.

Our proposed system registers presence, which can be used to register absence if all students use mobile phones all the time. However at present "only" a third of the students use internet on their mobile phones at school. This percentage is expected to increase for two reasons. Firstly, the market for smart phones is growing exponentially. Secondly, schools will make increasing use of mobile phones to access school services. Therefore we expect that our proposed method will indeed be able to register absence in the near future. In the meantime it is important to verify the acceptability of the system as proposed in this thesis, as well as some technical assumptions.

## Preface

Security, in the broadest sense, has always been an interesting field to me. As a matter of fact and not really a coincidence, computer science and telematics contain a large security component which is emerging in line with the increasing integration of computers in our daily life.

Looking for a final Master project, the Distributed and Embedded Security (DIES) group became quite an obvious group to discuss ideas with about final projects. The head of the group has close contact with a Professor in social safety studies at the faculty of Management & Governance. This Professor in turn has good contacts with the Dutch HALT office (*Dutch: Bureau HALT*) which takes care of prevention of, and fights against, youth delinquency. A specific form of youth delinquency is truancy, an issue which has high priority at the Dutch Ministry of Education, and so at the HALT office.

The HALT office knows that the previously mentioned faculty of Management & Governance is part of a technical university. With this idea in mind, HALT asked whether any technological research would be possible in the field of truancy recognition. This is still very broadly defined but the research proposal ended up at the faculty of Computer Science and was discussed with me during the first meeting with the head of the DIES group.

However, I indicated that I would like to do my final project externally, at a company. I also selected a company yet, which is called "Topicus," a software producer which amongst other develops software for educational purposes. This seems a good triangle so I discussed the research proposal at Topicus and the cooperation was born.

The initial research subject was stated as follows: recognition of students based on the data traffic coming from modern communication devices. After some fine-tuning the final name of the project became "Towards automated truancy detection."

## Acknowledgements

Neither this thesis nor my study would have been completed without the support of specific persons. Technical, educational and social support or one of the possible combinations out of these three, were important sources of ideas, positive criticism and motivation.

I would like to say a word of thanks to my first supervisor Professor Pieter Hartel, who gave me the chance to do this final project at the Distributed and Embedded Security (DIES) group, and who was always available to answer questions and came up with ideas of any kind.

I would like to say a word of thanks to Professor Marianne Junger (Social risks and safety studies, faculty Management and Governance) who taught me amongst other things how to do a more decent analysis of my research results and who never seemed to become tired of my ongoing flow of questions.

I would like to say a word of thanks to Pieter-Tjerk de Boer, who on behalf of the Design and Analysis of Communication Systems (DACS) group was willing to take care of the assessment of my work, even on such a relative short notice.

I would like to say a word of thanks to Marco van der Niet, one of my supervisors at Topicus B.V., for monitoring my progress while being at Topicus and always coming up with creative ideas and comments if the right track seemed to be abandoned.

I would like to say a word of thanks to Wouter van der Veer, one of my supervisors at Topicus B.V., for thinking with me in the technical field and always having an alternative solution to compare or compete with mine.

I would like to thank my parents for all those years they have supported and motivated me to do my best and get the most out of it. I am proud to present them the result by means of this thesis.

This list would not be complete without stating the daily support of my girlfriend, Bibi, with whom I live together for almost three years (about which I am very happy) and who supported me with her critical view and being my personal psychologist.

## Contents

<b>Abstract.....</b>	<b>2</b>
<b>Preface .....</b>	<b>3</b>
<b>Acknowledgements .....</b>	<b>4</b>
<b>Chapter 1 Introduction .....</b>	<b>7</b>
1.1 Background .....	7
1.2 Goal .....	7
1.3 Approach .....	7
<b>Chapter 2 State of the art .....</b>	<b>9</b>
2.1 Routine activities of students going to school .....	9
2.1.1 Social .....	9
2.1.2 Technical .....	10
2.2 Routine activities of people playing truant.....	13
2.2.1 Social .....	13
2.2.2 Technical .....	14
2.3 Possible approaches for truancy monitoring .....	16
2.3.1 Electronic monitoring .....	16
2.3.2 Biometrics .....	22
2.3.3 Data traffic analysis .....	25
2.3.4 Accelerometer .....	28
2.3.5 GPS .....	30
2.3.6 WiFi triangulation .....	33
2.3.7 GSM cells .....	35
2.3.8 Software based (Topicus).....	36
2.4 Discussion .....	37
<b>Chapter 3 Classroom observations .....</b>	<b>39</b>
3.1 Research methodology .....	39
3.2 Phones and usage .....	39
3.3 Bluetooth.....	40
3.4 School policies regarding mobile phones and adherence .....	41
3.5 Music.....	41
3.6 Other interesting facts .....	42
3.7 Conclusions.....	42
<b>Chapter 4 Statistical analysis.....</b>	<b>43</b>
4.1 Introduction.....	43
4.2 Method .....	44
4.2.1 Sample .....	44
4.2.2 Data collection.....	44
4.2.3 Concepts.....	45
4.2.4 Analysis .....	46
4.3 Results .....	46

4.3.1 Sample .....	46
4.3.2 Descriptives.....	47
4.3.3 Mobile telephone use .....	48
4.3.4 Wireless networks at school .....	49
4.3.5 Internet use and online social networks .....	50
4.3.6 Attending lectures .....	51
4.3.7 Multiple regression analysis .....	52
4.4 Discussion .....	55
4.4.1 Mobile telephone use .....	55
4.4.2 Other communication devices.....	57
4.4.3 Wireless networks at school .....	58
4.4.4 Social networks .....	59
4.4.5 Attending lectures .....	59
4.5 Conclusions.....	60
Literature .....	61
<b>Chapter 5 Prototype.....</b>	<b>62</b>
5.1 Architecture .....	62
5.2 User interface .....	63
5.3 Handlers .....	65
5.4 Data storage.....	66
<b>Chapter 6 Network traffic analysis .....</b>	<b>68</b>
6.1 Background .....	68
6.2 Test environment .....	68
6.3 Results .....	70
6.4 Conclusions.....	77
6.5 Recommendations.....	78
<b>Chapter 7 Conclusions and future work.....</b>	<b>79</b>
7.1 Answers to research questions .....	79
7.1.1 Sub research question 1 .....	79
7.1.2 Sub research question 2 .....	80
7.1.3 Main research question .....	80
7.2 Future work .....	81
7.2.1 Research .....	81
7.2.2 Prototype .....	82
<b>Literature .....</b>	<b>84</b>
<b>Appendices.....</b>	<b>87</b>
Appendix A Classroom observation form.....	87
Appendix B Questionnaire.....	88
Appendix C Documentation "Personal Information Grabber" .....	91

## Chapter 1 Introduction

### 1.1 Background

Truancy is a well-known problem in The Netherlands and, according to the Dutch Ministry of Education, a decent indicator for possible future dropouts. Decreasing truancy would therefore one of the counter actions to reduce the number of dropouts. At least two important facts are going on in the Netherlands. First, Dutch government explicitly stated that they want to decrease the number of dropouts from 40.000 in the year 2010 to 25.000 in the year 2015. Second, digitalization of education rapidly changes the way of communication and administration at schools across the country. Research showed that this digitalization causes a sometimes dramatically decrease of truancy registration, due to lack of human cooperation and the lack of knowledge of IT systems among teachers. This does not help to achieve the goal of the Dutch Ministry of Education.

### 1.2 Goal

The goal of this project is to research how the detection of truancy could be automated. To achieve this goal, the research has been divided into two parts. For the first part a sub research question has been formulated: "What does the behavior of students with respect to modern media look like?" The second part of the research attempts to find an answer to the question is: "How can this behavior be registered automatically?"

### 1.3 Approach

In order to achieve decent research results, the research has been conducted following a specific plan, defined prior to the start of the actual research. This plan is basically divided into two main parts. First, field research has been performed different schools for secondary education in The Netherlands. Data were collected by means of a convenience sample of schools for secondary education in The Netherlands. However, care was taken to collect data in schools at different locations, to cover the entire age-range of high school students, and to include all Dutch school-types which include VMBO, HAVO and VWO, in increasing level order. In total 49 schools were approached, in 14 cities. Seven schools participated in the study, divided over four cities. In each school, data were collected in different classes with respect to education level and study year. Data were available from two schools in Gorinchem, two schools in Oldenzaal, one in Oss and two in Raalte.

The field research consists of attending lectures and writing down all observations that could be relevant for developing a truancy detection system. Besides that, all students in the

lectures that have been attended were asked to fill in a questionnaire. 701 students took part in the survey.

Second part of the research plan is the development of a software tool that is able to monitor a data network and filter and store specific events occurring in the data traffic. A fully automatic presence detection system has been designed and a prototype has been implemented. This prototype has been used in a practical experiment at a secondary school in The Netherlands. In the context of this experiment the wireless network of the school has been monitored for one week and data traffic has been captured. Interesting events in the data traffic have been filtered and stored in a database.

This thesis is organized as follows. Chapter 2 reports about a literature study performed prior to the start of the research in order to investigate the state of the art of automated presence detection. Chapter 3 describes all findings which have been noticed during the field research at secondary schools. In chapter 4 the outcomes of the survey are explained and analyzed. Chapter 5 summarizes the explanation of the working of the software prototype while Chapter 6 reports about the outcomes of a practical experiment in which the prototype has been used. Finally in chapter 7 the research questions are answered and form the conclusions of this research. Chapter 7 also proposes several directions for future work.



## Chapter 2 State of the art

Prior to the start of the research that has to be performed to answer the research questions in this project, the state of the art of presence detection has been studied. The result of this literature study is reported in this chapter.

### 2.1 Routine activities of students going to school

In order to be able to detect and conclude about people that play truant, it is good to have some psychological background knowledge. To be able to recognize exceptional behavior, it is necessary to be familiar with the normal behavior. For this reason, routine activities of pupils and students in primary and secondary schools, respectively, are investigated based on available literature. The special attention for the use of communication devices and multimedia divides this section into two parts, a social and a technical part.

#### 2.1.1 Social

In this section the focus is on the social life that the youth lives nowadays. Two main activities of youths are going to school and socializing with friends. It seems that both main activities often happen synchronously.

More general, researchers of the London Police Bland and Read state that young people will congregate in public and that this is both inevitable and socially necessary [1]. For example an experiment in the year 2000 among a group of 2,272 students between 7 and 17 years old showed that more than 90% enjoy it to be with friends at school [2]. However, the opportunities offered by the school for being with friends are not always that good [2]. This might be a reason to use modern communication to keep in touch and exchange messages during the day at school.

This congregating is part of the process of personal development from childhood to adulthood. However, group size may influence individual behavior - teenagers often behave in front of a group of peers in ways they would not if they were alone or in pairs [1].

The after-school activities that are part of the behavior of the youth are several and include: (1) playing school or community sports or participating in school clubs, (2) watching TV alone, (3) doing homework or reading alone, (4) hanging out with friends in a public place, (5) hanging out with friends at someone's house, (6) exercising, jogging, working out, other forms of

exercise or leisure sports, (7) spending time alone, and (8) participating in community organizations. The study in [3] shows that adolescents from the four countries Hungary, Switzerland, Netherlands and USA spent their time in remarkably similar ways; most of their time was spent in solitary activities, followed by peer, family, and community/sports activities.

Another study, performed in England in 2007, provides besides the activities of adolescents also the corresponding percentages of time used for these activities during the day [4]. The collected data is summarized in Table 1. It reflects the percentages of time use relative to other activities but cannot be used to derive exact times from, since it was not investigated how long the youth is at school and stays in bed. Media use means watching TV, listening music and use of computer playing games and surfing the internet. The subject "reading" covers all reading activities not associated with school work. Several more activities were encountered, like shopping and being at work in a job, but due to low frequencies these activities were not added to the list.

**Table 1: Adolescents activities and percentages [4]**

<b><i>Activity</i></b>	<b><i>Percentage</i></b>
Schoolwork	36.31 %
Hang out / media use	34.78 %
Resting	10.20 %
Phone	4.16 %
Extracurricular activity	3.93 %
Meal	3.87 %
Physical Exercise	3.01 %
Reading	2.27 %
Party	1.47 %

### 2.1.2 Technical

After having investigated the social characteristics of students who go to school in the previous section, this section has the focus on the behavior of students at school with respect to modern media. Almost every student already has his own mobile phone. Most secondary schools however adhere a policy which states that mobile phones in school are prohibited. The currently popular smartphones have much more functionalities than only SMS and voice services. Despite the relatively small display size and low processing power, the key advantage which is all time availability, makes smartphones suitable as tools for supporting learning, in and outside classrooms [5].

There is already ongoing research to find efficient ways to integrate smart phones in everyday learning and to encourage teachers to use the options mobile devices offer to enhance their lessons [5]. The results concentrate on teaching and learning scenarios in which students work in virtual teams - teams whose members communicate and collaborate (partly) over the Internet. Schmiedl et.al. expect that in about two years of time, their tested scenarios can be part of everyday lessons.

Today's adolescents grew up using the Internet, and in turn they are extremely familiar with the large amount of services available online. Youth are especially involved in online socialization with various methods of computer mediated communication (CMC), such as e-mail, chat rooms, instant messaging, and social networking websites. Moreover, not only are more adolescents using the Internet to socialize, they are also spending more time online [6].

During the past six years the fast development of mobile devices, especially cell phones, has presented an opportunity to develop new interactive classroom systems. Classroom Feedback Systems (CFS) provide one possible technological mechanism that can efficiently enable interaction in classes. Advanced CFSs provide the ability to answer a range of question types, from simple yes/no through to detailed responses, free-form questions and role-playing. Current platforms range from small infra-red units, through radio units, to the use of Web systems accessed by wireless personal digital assistants (PDAs) or laptops [7].

Even more integrated under students are mobile phones. These devices have the advantages of being familiar, permanently configured to work correctly, and battery lives generally measured in days rather than hours. Thus, as a platform for a classroom feedback application, the mobile phone is highly preferred over other types of devices.

Attempting to combine the social and technical considerations in Section 2.1 we propose the use of the script as in [8]. First, the social aspect motivation is addressed. Several factors can be the reason to get motivated to go to school. The most important ones are listed below:

- Sleep well
- Pack your bag
- Make your lunch
- Study your exams
- Finished your homework

To investigate the motivation in more detail, the script has been set up similar to scripts that describe crimes and their causes as in [8]. This script is depicted in Table 2. This table also

shows the responses on the steps, that is, how these steps could be monitored using modern communication technology.

**Table 2: Script: youth motivated to go to school**

STAGES	STEPS	RESPONSES
<b>Preparation</b>	Pack school bag	
	Make lunch	
	Finish homework	Log homework hand-in session
<b>Entering setting</b>	Enter school building	Detect data traffic, identify the generating user
	Meet classmates	Check-in (RFID)
<b>Enabling conditions</b>	Classroom available	Analyze devices connected to wireless AP
	Teacher present	Responsible for students checking-in (RFID)
<b>Completing the education</b>	Listen to teacher	
	Make exercises	Track activity during online exercises
	Ask questions	
<b>Exiting the setting</b>	Leave school building	Detect absence of data traffic
		Check-out (RFID)
<b>Aftermath</b>	Meet classmates	Data from appointments made via electronics
	Make homework	Check if school database is queried

## 2.2 Routine activities of people playing truant

This section focuses on pupils and students that play truant. Like in the previous chapter, this chapter gives special attention to the use of communication devices and other digital electronics used by the pupils playing truant. The structure of this chapter is also similar to the previous chapter, there exists a division between the social part; what activities are performed while playing truant, and the technical part; how can these activities be monitored/logged/traced by reading a data traffic log or use another way of modern technology.

### 2.2.1 Social

In Section 1.1 ordinary students and their possible behavior characteristics are described. This section addresses the behavior of a more or less special kind of students, namely truants.

Truancy is a violation of State law. Underlying issues that have been identified include family poverty, less education, substance abuse, cultural variation in the valuing of public education, and pressures on the youth to work and provide childcare for younger siblings [9]. In fact, truancy can be traced to four main causes [10-12]: 1.: unsupportive school environments, 2.: lack of community support, 3.: chaotic family life, and 4.: personal academic or social deficits.

Most of the students who played truant loved wasting their time by going to entertainment places. Next most important activities while playing truant are “doing a part-time job” and “mingling.” Further, helping the family and taking part in criminal activities and joining bad groups are activities reported instead of being at school as is supposed [13]. The last two indicate a risk factor for delinquent behavior in youth. It has been found that truancy is related to substance abuse, gang activity, and involvement in other criminal activities such as burglary, auto theft, and vandalism [9]. For those reasons it is important to identify strategies to intervene with chronic truants [9], particularly because truancy enforcement can be effective in reducing youth disorder occurring during school hours [1]. After all, truancy is not just a social problem, leading to unruly behavior among the young people who are not present at school. What is maybe more emerging from an educational perspective is that truancy also has a clear effect on school performance. At the individual level (poor grades encourage the cycle of poor attendance to continue) as well as at the school level [10, 14].

In the hours after school, where truancy does not exist, there can still be derived that pupils who spend more evenings outside their homes in general have a less school-oriented attitude

and, therefore, tend to score lower on tests. It is striking to note, however, that despite all the criticism with regard to the alleged effects of television, there is no significant relation between the number of hours spent on watching television and the score obtained on the test [14].

Investigations over a longer term shows that adults who were frequently truant as teenagers are much more likely than those who were not to have poorer health and mental health, lower paying jobs, and increased chance of living in poverty [9].

Attwood & Croll presented evidence that shows how truancy increases through the years of secondary education until, at the end of their first term in Year 11, about 1 in 10 young people reported that they had truanted at least 'several times' over the past year. There is a degree of continuity year to year in truancy and most year 11 truants had reported truancy earlier in their school careers. But the sharp increase in truancy apparent towards the end of compulsory schooling means that there are necessarily many new truants in Years 10 and 11. The five years of compulsory secondary school in the UK, covering the ages of 11 to 16 years, start at Year 7 and finish at Year 11 [15].

An investigation in The Netherlands showed that the percentage non-attendance with unknown reason in Dutch MBO schools is 24,4% which is six times the non-attendance at Dutch secondary schools (4,1%). However, since it is possible to give a legal reason of absence like for example illness, even if someone does not have a legal reason for being absent, it is hard to get insight in illegal absence. Absents from which it is unknown why they are absent can have either a legal or an illegal reason for their non-attendance. Presence on the other hand is something that allows for more proper results. Presence at Dutch MBO (69%) is way lower than at secondary schools (89%) [16].

### 2.2.2 Technical

After having investigated the social characteristics of students who go to school and play truant, this section has the focus on the behavior of students at school with respect to modern media. Among all respondents (3319 sixth grade students attending 24 middle schools in Southern California), 37% reported that they prefer Internet (surfing the Internet, using chat rooms, emailing or instant-messaging) as a favorite way to spend their time. Internet use was the 6th favorite thing to do in a multiple choice question listing 12 common activities for adolescents, following that of watching TV/Videos, listening to music/radio, hanging out with friends, video or computer games, and team sports [17].

At school, 90% of the respondents could access the Internet, 40% used it at least occasionally, and 13% used daily. At home, 79% could access the Internet, 62% used it at least once in their lifetime, and 37% used it at least 1 h per day. Nearly all (99%) of the respondents reported that they could access the Internet from either school or home [17].

There are several steps and various reasons that can cause school refusal. The most common are listed below [18].

- Sleep bad / Overslept / Necessity to wake up early
- Homework not finished
- Did not study exams
- Dislike particular lessons

To indicate how truancy possibly occurs, a script has been set up, similar to scripts that describe crimes and their causes as in [8]. This script is depicted in Table 3. This table also shows the responses on the steps, that is, how these steps could be monitored using modern communication technology.

**Table 3: Script possibly adhered by truants**

STAGES	STEPS	RESPONSES
<b>Preparation</b>	Claim spurious attendance	Authentication during attendance registration
	Ask classmates to check-in truant	
	Finish homework	
<b>Entering setting</b>	Go to meeting point	Analyze data traffic at local hot spot
	Meet other truants	
<b>Enabling conditions</b>	Parents at work	
	No supervision	
<b>Completing the truancy</b>	Play games	Derive client IP address from game server
	Watch movies	
	Commit crimes	
<b>Exiting the setting</b>	Leave friends	Analyze online mail/chat activity
	Go home	
<b>Aftermath</b>	Punishments	

## 2.3 Possible approaches for truancy monitoring

In order to monitor the attendance of students at school, several different approaches in the field of localization and positioning and in the field of identification are possible. The main purpose of each individual approach is stated between braces:

- Electronic monitoring using wrist- or ankle band (localization & identification)
- Biometric registration (identification)
- Data traffic analysis (identification)
- Accelerometer (identification)
- Global Positioning System (GPS) (localization)
- WiFi triangulation (localization)
- GSM cell (localization)
- Software based (identification)

Identification happens based on biometrics or by means of a dedicated electronic device like a wrist band or chip card. Localization happens by enabling communication of the client device with a (fixed) registration device.

### 2.3.1 Electronic monitoring

Using tracking systems, criminal justice agencies can monitor an individual's location and be alerted to any unauthorized movements. Technology, thus, can be useful in detention, restriction and surveillance. However, constant surveillance of people, particularly through the use of devices fixed to their body, or even implanted beneath the skin, raises serious civil liberty and ethical concerns [19].

Justice agencies claim three main rationales behind the use of electronic monitoring:

1. Detention:

Electronic monitoring can be used to ensure that the individual remains in a designated place. For example, home detention schemes typically require offenders to be at home during established curfew hours. This was one the first uses of electronic monitoring and remains the most popular.

2. Restriction:

Alternatively, electronic monitoring can be used to ensure that an individual does not enter proscribed areas, or approach particular people, such as complainants, potential victims or even co-offenders



### 3. Surveillance:

Finally, electronic monitoring may be used so that authorities can continuously track a person, without actually restricting their movements.

There are a number of technologies available that can aid with the detention, restriction or surveillance of individuals within the criminal justice system. Most involve some kind of device that is locked onto the subject's wrist or ankle with tamper-proof elements to prevent removal. There exist **passive** systems in which wearers are periodically contacted by telephone to ensure that they are where they are supposed to be. The individual's identity may be verified by a password, a device that the subject wears or a biometric. Passive systems are only effective for detention purposes.

Furthermore there exist **active** systems which utilize a device worn by the individual that continuously emits a signal. A corresponding device in the person's home relays the signal to a monitoring station. If the wearer strays too far from home or breaks the device, the authorities are alerted [20].

A variation of this system utilizes mobile equipment that can detect the presence of the individual's device. Active systems primarily seek to enforce detention, although they may be extended to achieve some restriction and surveillance as well.

In the context of an educational institution, this "detention" could be seen as the compulsory attendance of the lectures at school. So by using such an active system, the attendance of students during the lectures could be monitored.

The surveillance purpose in criminal context can be achieved to some degree by placing monitoring devices at bus stops and train stations so that the individual can be tracked to and from work. This is not directly necessary for the student attendance monitor.

UniNanny® is an electronic attendance monitoring system designed at the University of Glamorgan to minimize the problems associated with traditional paper-based monitoring. The system omits the need for students to pass registers around the classroom, and makes cheating the system more difficult as students must entrust a fellow student with their individual microchip. UniNanny® eradicates the need for administrators to manually enter data into a database, enabling staff to spend more time supporting students. Academic staff are also spared the task of passing on registers for analysis, thereby reducing time delays in identifying absenteeism. The procedure is as follows. A lecturer attends a learning event with a small electronic baton (stick), and a 'fob' (microchip) for that learning event. At the beginning of the learning event, he/she places the fob near the baton – a process which takes seconds - to indicate the lecture for which attendance will be registered. The baton is then passed around

the learning event. Each student introduces their fob to the baton, indicating that he or she is present. It is estimated that the process is at least four times quicker than students signing their name on paper. The baton has finally returned to the lecturer in much the same way as a piece of paper would be handed around all the students in the class and returned at the end. The lecturer might use the same baton for any number of learning events that day or that week but lecturers are encouraged to take the batons back to the central administration area on a daily basis, where the information can be downloaded to the database. The data that is downloaded is held on a local machine and also automatically updated to a central web server. Administrators can now monitor students' attendance, using information on their local machine, or they can use more complicated search facilities, which is web-based [21].

Computerized swipe card systems are available in many schools for absence registration, these are however costly. Many areas rely on lengthy manual systems that require hours of work to elicit even the most basic of information [22].

Touching with a mobile terminal has been found to be an intuitive, natural and non-ambiguous interaction technique that does not incur much cognitive load for users [23].

In the attendance supervision trial performed in Oulu, Finland pupils were given contactless smart cards named "Robo" containing the pupil ID. Upon arriving at school pupils in the first grade class 'logged in' by touching with an NFC (Near Field Communication) smart card an active card reader device and pupils in the special-need class logged in by touching a NFC-enabled mobile phone. The reader devices recorded the card ID (the child's name), the direction (arrival in school) and a time stamp in the backend system. At the end of the school day pupils touched the reader devices again to mark their departure.

The log of arrivals and departures was automatically compiled by a backend system, and could be read by a teacher in a classroom in real-time. If a login did not occur, the pupil was marked absent by default. If a pupil logged in late, the backend system recorded the lateness. Parents were able to get information of their children's attendance details via an online 'citizen's portal' and through text messages sent to their mobile phones. The system prevented truancy by informing tutors, administrators, and parents of absences in real time, enabling instant intervention.

Children, as well as their teachers, became very fast familiar with the login process, and the attendance supervision was soon integrated into their everyday school routines, mainly due to the intuitiveness and effortlessness of the NFC touch-based interaction technique [24].

The attendance control proposal in [25] is based on RFID over Ethernet. It further comprises converting the existing student debit cards for student cards tagged with an RFID tag. In addition, by installing one RFID reader per classroom and having them all connected to the Institute's Local Area Network (LAN). Finally, a computer system will be responsible for recording the attendance of students and providing web services. Once a student enters a classroom he will place his card near the reader if he intends to register his presence in a particular class. The reader should emit a sound when the card is successfully read. After a successful reading, the RFID reader establishes a communication with the RFID server in order to send the student card identification. Successful tests have been done using five RFID readers simultaneously in the system. The operating frequency of the system is 125 kHz and the detection range is approximately 10 centimeters.

Another system that makes use of RFID is described in [26]. If a person wearing an RFID tag passes through the electromagnetic field generated by the RFID reader, the person is registered based on the unique identification number on the RFID tag. Important difference with the other RFID based systems mentioned in this report is that the user does not have to show his RFID tag in front of the reader. Instead, the tag is read from a greater distance so it can remain in the user's pocket. Basically a server program takes the Reader ID and then receives Tag ID's from the client. On the basis of a flag-value it maintains real time data in the log in the form of 'enter/ exits'. Attendance of students against every course is marked on the basis of calculated Stay-In time from the 'Enter/ Exits' in the log. If there Stay-In time, matches with the required time, then attendance is marked as 'Present'. Attendance of those students not attending/coming to class/institute is also marked as 'Absent'. Before marking attendance, duplicate record is checked in order to avoid redundancy.

A third system that makes use of RFID is introduced in [27]. The main function of the RFID Based Attendance System designed in this project is to scan and verify a RFID tag. Then, attendance will be taken based on the ID scanned. The system is compact and light weight and can run using power adapter or battery power. Therefore, it is very portable and can be carried to the class for taking the attendance. The system has a high identification and verification speed. The operating frequency of the system is 125 kHz and the detection range is approximately 5 centimeters.

The system in [28] makes use of the IEEE 802.15.4 "ZigBee" technology to monitor attendance of employees at the workplace. The system works as follows. Each employee has an IP-Link Tag associated with a unique ID as the employee identification. The IP Link Tag sends employee attendance information directly to a ZigBee dongle through the ZigBee network. The

dongle receives the employee attendance information and then sends them by Bluetooth to a PDA that serves as mobile transfer. The PDA parses the employee attendance information to filter the available information and sends an XML-based [29] document to a pc in the role of middleware web server. This pc can be queried to browse the employee attendance records.

**Table 4** shows all information presented in this section in a summarized and schematic way. The most important features are listed to make an easy comparison possible. In the first column the name of the system (if the system has a name) is stated. The second column shows the "identification base" which describes the physical object that is used to identify a person. The third column "human interface" shows the used technology that is used at the user end of the system, so how the identification base communicates with the system/infrastructure. Column four indicates whether the registration device (reader) is directly or indirectly connected to the underlying infrastructure and database or that it has to be connected manually using (USB) cables or cradles. The infrastructure column shows how the underlying infrastructure is built. In the column "data reviewing" is stated how the collected attendance data can be accessed, viewed and processed afterwards.

**Table 4: Electronic Monitoring Applications**

Name	Identification token	Human interface	DB connection	Infra structure	Data reviewing	Source
<b>UniNanny</b>	Fob (microchip)	- -	Indirect	- -	Web interface	[21]
<b>Robo</b>	Smartcard	NFC	Direct	- -	Web interface	[24]
-	Student card with RFID tag	RFID	Direct	TCP / IP via existing LAN	Web interface	[25]
-	RFID tag	RFID	Direct	- -	Web interface	[26]
-	RFID tag	RFID	Indirect (via USB)	- -	Hyper Terminal	[27]
<b>Vendor: Nedap</b>	RFID card	RFID	Direct	- -	PeopleSoft	[30]
<b>MUEAMS</b>	IP-Link Tag	ZigBee	Direct	Bluetooth & WiFi/3.5G	Any mobile device	[28]

Most of the systems try to tackle the time consuming problem of pen and paper registration of students to recognize absenteeism. Although this goal is reached successfully, the solutions still

suffer from cheating possibilities and/or ethical issues. For example the possibility to take a friend's tag to class and register it while the owner is physically absent. Or the association with crime while wearing an ankle band. Furthermore, the registration still takes place manually, and students need to present their smartcard or fob or tag. This is something to work around in a next generation attendance monitoring system, or truancy detection system. Registration and processing should both take place "behind the scenes" or "under water."

**Advantages:**

- Easy integration of (RFID) chip into student card.

**Disadvantages:**

- Cheating by taking fellow student's card and check him in pretending being present.
- Need to show chip card at every check-in and check-out event.

Table 5 shows the effectiveness of the different technologies that are presented in this section, with respect to the ability of the systems to reduce truancy. It turns out that most of the papers only describe the used techniques but do not focus on results. The papers that do focus on results show a reasonable increase in the registration of the attendance. This however does not reflect a direct decrease in truancy but a more decent registration will discover more truants so at least there can be taken measures in order to attempt truancy.

**Table 5: Effectiveness of the different electronic monitoring techniques**

<b>Electronic Monitoring</b>		
<b>Problem</b>	<b>Approach</b>	<b>Effectiveness</b>
Lack of prison space [20]	Equip perpetrator with ankle band for home detention	Too easy to refuse and run away without permanent tracing abilities.
Lack of prison space [19]	Passive-/Active-/Global Positioning-Systems	Correctional programs have potential. However, lack of physical restrictions is a risk factor.
Pen-and-paper attendance registration [21]	UniNanny®: portable attendance registration device	Increase of attendance registration between 30 and 90%, case specific.
Evaluation of Attendance Mon. system [22]	Self-declaration using web service	Attendance registration was less than reliable in these tests, only 40%.

Save time for teaching [24]	Attendance registration using personal NFC cards	Quick familiarization and integration in everyday school activities.
Automation of attendance registration [25]	Provide RFID cards and equip building with RFID readers	System satisfies but no focus on the increase of school attendance related to the use of the system
Automation of attendance registration [26]	Provide RFID cards and create EM fields at entrance/exit	Experiments proved that the processing by the Windows Service was 100% exact and timely.
Pen-and-paper attendance registration [27]	Provide RFID cards and equip building with RFID readers	System satisfies but no focus on the increase of school attendance related to the use of the system
Attendance monitoring [28]	Integrate wireless sensor networks with mobile ubiquitous technologies	n/a
Drop-out at Dutch ROC schools [30]	Several: RFID / fingerprint attendance system; RFID lockers	Increased the attendance registration from 30% before introduction to 90% after.

### 2.3.2 Biometrics

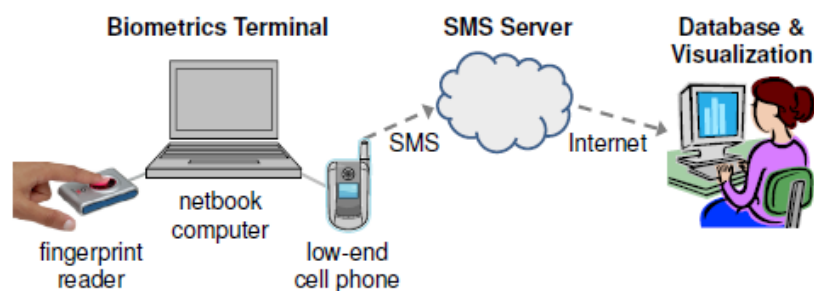
Biometric technology is becoming increasingly prevalent in today's society. Fingerprint technology is by far the most common used modality, followed by iris recognition [31].

Network75 offers a biometric reader suitable for monitoring the attendance of international students at UK Universities [32]. The biometric reader satisfies UK home office regulations that state that there is a requirement for universities to record the attendance of international students, notifying the UK Border Agency if an International student holding a student visa is absent from the University for more than 10 working days without authorization.

The UniNanny® attendance monitoring software produces a range of reports on student attendance that can be used to track international student absence and are suitable for presenting to the UK Border Agency to prove the absence of international students therefore satisfying the requirements associated with the proof of attendance of international students.

At another experiment in England, at Balby Carr Community Sports College in Doncaster, video cameras have been installed recently, as well as a CCTV controlled biometric access control system, requiring students to use their fingerprint to gain entry to the college. The site manager did not report any negative feedback, and recognized that all students were happy to have their fingerprints recorded, and were educated on the implications beforehand [33].

The biometric terminal as presented in [34] and shown in Figure 1 is used in health care attendance/visits monitoring in India. The system consists of three main components: a low-cost netbook, a commodity fingerprint reader, and a low-end mobile phone that is connected via USB to the netbook. Messages sent from the terminal are received by an SMS server and made available over the Internet to administrative users, who can download the messages from any location and automatically import them into a database for further analysis and visualization. The total cost of the components and the SMS service is less than \$500,- which is at least \$100,- less than comparable systems which furthermore make use of GPRS which is more expensive than using SMS for transferring messages in this case.



**Figure 1: Biometrics Terminal [34]**

Findings of a study indicate that innovativeness and optimism are the key drivers for gauging readiness of users to embrace the fingerprint technology [35]. A successful biometrics technology however must all of the three; secure, acceptable and voluntary [31].

Fujitsu's PalmSecure biometric technology has proved itself to be an ideal way for Primary School pupils in Scotland to pay for their school meals [31]. The system uses an infrared sensor to recognize the blood veins pattern in a pupil's hand. This solution has been rolled out successfully in Scotland and is under consideration for schools across the UK. This is not only one of the applications of biometrics in educational institutions but also forms a familiarization process for pupils in primary schools to work with biometrics. This is an advantage in case biometric identification for attendance monitoring will be used in secondary schools. Now they are already familiar with the use of biometrics, they probably agree faster with the use for attendance monitoring.

Hand recognition is one of the popular biometry technologies, especially in physical access control and time and attendance monitoring [36].

A highly reliable Fingerprint Access Control System is presented in [37]. The system provides multiple functions among which are fingerprint access, PIN Code access, fingerprint management and historical attendance review.

Yongqiang and Ji designed and implemented a wireless fingerprint attendance system and described it in [38]. The system consists of two main parts, a fingerprint verifying machine and a pc workstation which are connected through wireless communication nodes used at both sides. Fingerprint matching is performed at the fingerprint verifying machine which sends an attendance registration to the backend pc workstation. This workstation further processes the attendance data. The system is mainly used to monitor employee attendance, however the performance of this system meets the needs of daily attendance management and is for this reason able to support attendance monitoring in various enterprises and institutions, so for example in school buildings.

At the university of L'Aquila, Italy, a wireless biometric badge has been developed. This badge contains a chip that has a ZigBee based location engine [39]. The location is derived based on the received power level. Hence this requires a wireless sensor network and is mainly usable indoor. This badge furthermore contains a fingerprint sensor to verify that the dedicated user is wearing the badge and not somebody else pretending that someone is in a place just because his badge is located there [40].

**Advantages:**

- No cheating due to necessity of personal biometric sample at verification time.
- Highly reliable identification of correct user through biometric sample.

**Disadvantages:**

- Need to show biometric example at every single verification event.
- Acceptance of use of biometric systems in school.

Table 6 shows the effectiveness of the biometric applications presented in this section. As can be seen, most applications

**Table 6: Effectiveness of Biometrics in school**

<b>Biometrics</b>		
<b>Problem</b>	<b>Approach</b>	<b>Effectiveness</b>
Deploy abuse-free pay sys. at school [31]	Fujitsu Palmprint reader for pupil authentication; no cash involved	Children now have a simple method of receiving their meals; system can be extended



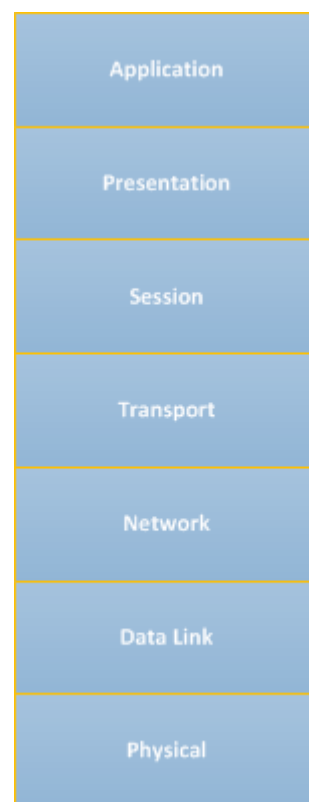
Proof attendance of international students [32]	UniNanny® Fingerprint attendance registration system	Produced reports are suitable to present to UK Border Agency; approach is effective.
Application of biometrics in developing regions [34]	Fingerprint reader attached to a laptop in turn attached to a mobile phone to exchange data via SMS	Incentives are needed to sustain the usage over time, technically satisfying system
Acceptance of biometrics in schools [35]	Perform a user study	An indication of the success factors of a biometric authentication system at schools
Reliable fingerprint access system [37]	design highly reliable fingerprint access control system based on an C8051F020 single chip	n/a
Spurious attendance (cheating with the registration system) [38]	Design of a wireless fingerprint attendance registration system	Practices proved effectiveness and user friendliness. Performance meets the needs of daily attendance management
Capabilities of WSN for positioning [39]	Analyze achievable performance of a WSN testbed platform	n/a
High security access control to restricted areas [40]	Wireless biometric badge equipped with fingerprint reader for authentication	n/a

### 2.3.3 Data traffic analysis

Where application files show a clear connection to computer users, system and network information may be equally as telling of user activities [41]. Network traffic analysis can be done at different levels. The OSI stack model consists of several layers (Figure 2) and each of the separate layers can be considered as a different level.

An enterprise should keep records of network events such as logging in or out of a computer and accessing network services such as remote Telnet, or FTP sessions. These records are very useful during an investigation of the network use, because they contain information about the activities of a specific user, as well as dates and times of those activities [41].

At application level, both HTTP and SMTP traffic contain valuable information to anyone investigating network activity. These protocols can be tracked in network devices such as routers and servers [41].



**Figure 2: OSI stack model**

To monitor traffic on a certain network, among others a network interface card (NIC) can be used. A NIC works in two modes, non-promiscuous mode (normal mode) and promiscuous mode. In order to capture the packets, the NIC has to be set in the promiscuous mode. Packet sniffers set the NIC card of its own system to promiscuous mode and receive all packets even they are not intended for it [42].

To give an indication about the amount of data involved in network traffic monitoring by packet sniffing, the workload of 195 users during a three days conference was analyzed. The total amount of transmitted data was 4.6 GB [43].

Kotz and Essien [44] traced network activity at Dartmouth College in 2001. They used three techniques to collect data about wireless network usage: syslog events, SNMP polling, and tcpdump sniffers. The network covers over 400 wireless access points and serves more than 2,000 users. To capture the syslog events, all Cisco Aironet 350 access points used on the Dartmouth campus have been configured to transmit a syslog message for different interesting events. The APs published a syslog message every time a client (or in fact an 802.11b NIC) authenticated, associated, re-associated, disassociated, or de-authenticated with the access point. During the eleven weeks of this experiment, over 3.5 million syslog messages arrived via UDP at a server which recorded all of them for later analysis. The server adds a timestamp to each message as it arrives. Each message further contains the AP name, the MAC address of the card, and the type of message, as described before. The network however does not use MAC-layer authentication in the APs, or IP-layer authentication in the DHCP server. Any card may associate with any access point, and obtain a dynamic IP address. This means that although the APs emit an "authentication" message for each card, there is no user authentication. We thus do not know the identity of users, and the IP address given to a user varies from time to time. It is assumed that DHCP data allows to associated MAC addresses with users [44].

The Simple Network Management Protocol (SNMP) was used to periodically poll the APs with intervals of 5 minutes. Each poll returned the MAC addresses of recently associated client stations, and the current value of two counters, one for inbound bytes and one for outbound bytes. The AP does not reset the counters when polled, so the difference between the values retrieved by one poll and the values retrieved by the next poll needs to be computed in order to obtain the actual values. Although each SNMP record contains a list of cards associated with the AP, the syslog data has been chosen to use for tracking cards because the syslog data provides the exact series of events for each card, whereas the SNMP polling data is less precise. The SNMP records support the analysis of the sniffer data [44].

Additionally, packet headers were recorded using the packet sniffer application tcpdump. Four computers were used, with the sniffer in promiscuous mode, to capture wireless traffic from 22 different access points. The total amount of traffic monitored during the eleven-weeks experiment consumes 183 GB memory. 90 GB is generated by HTTP (including HTTPS) [44].

According to [45], detecting the presence of nodes is a largely unexplored field. Although there are a number of research directions that deal more or less directly with the presence of nodes in wireless networks, this covers ad-hoc networks where the devices periodically send presence information into the network, while the current goal is presence detection based on already available information, so without sending and receiving presence information messages.

In this section, several approaches of analyzing data traffic are proposed. As can be seen, there are various methods to extract useful information from data traffic, which will be generated by students anyway. This means that, assuming that students generate enough data traffic, there is no additional action required and we could use data which is already available in everyday life. The challenge is to detect presence and identify mobile nodes and in line with this, to identify the users (students) and confirm that they attend their lectures.

#### Advantages:

- (Almost) every student has a mobile phone [46].
- No check-in and check-out events necessary.

#### Disadvantages:

- Cheating by taking fellow student's mobile phone and generate data traffic, pretending presence of an absent student.
- Requires reasonable and periodical network use/generation of data traffic, which cannot be guaranteed (yet).
- A lot of data traffic requires a lot of processing.

Table 7 shows the effectiveness of data traffic analysis, mainly focused on the problems approached in the corresponding papers. However, most papers only explain the technologies and do not report about the effectiveness of the presented solutions.

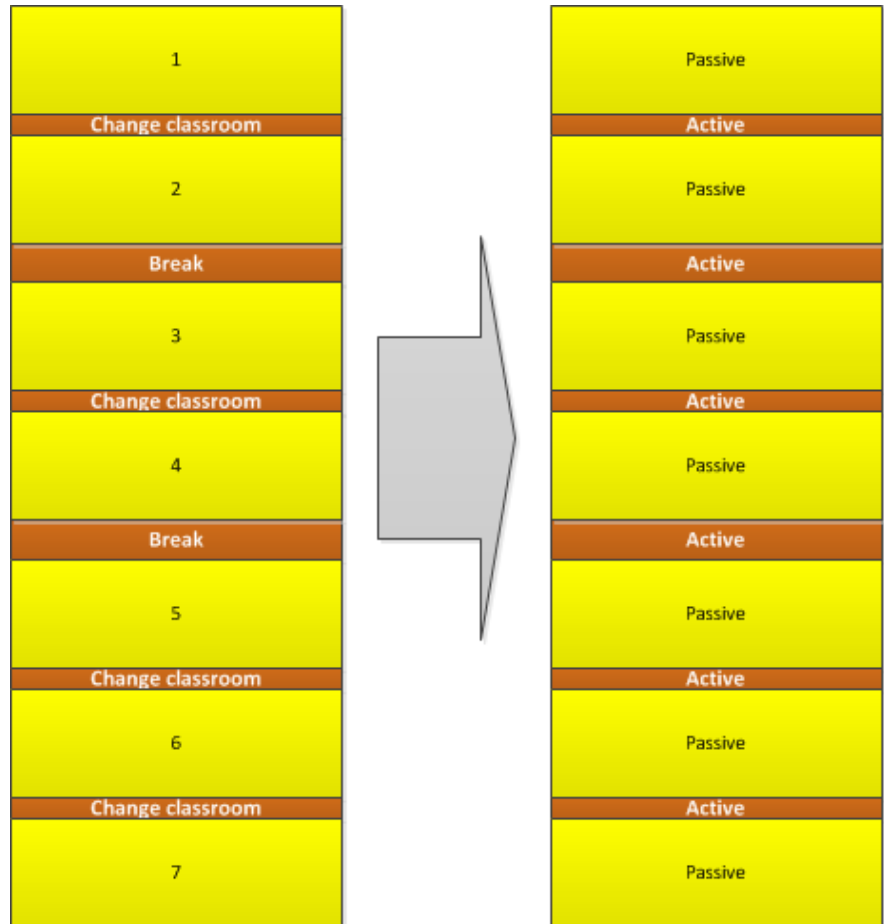
**Table 7: Effectiveness of data traffic analysis**

<b>Data traffic analysis</b>		
<b>Problem</b>	<b>Approach</b>	<b>Effectiveness</b>
Characterization of user behavior in public WLAN [43]	Analyzing a three days trace of the network traffic during a conference	n/a
Crimes with and against computers / networks [41]	Six proposed policies, enterprise network and computer related.	Deter computer crime and enhance recovery from attacks
(Unauthorized) network traffic analysis [42]	Packet sniffer use for intrusion detection and how to detect the presence of sniffer software	Sniffer can detect some threats, definitely not all. Sniffer can only be detected by doing tests, not by an average network user
Characterization of user behavior in public WLAN [44]	Analyze an eleven weeks trace of the activity of two thousand users from a general campus population	n/a
Detect presence/ absence of nodes in MANETs [45]	Send queries into the network to retrieve actual state	n/a

#### 2.3.4 Accelerometer

From a project during the course *Ubiquitous Computing* at the University of Twente, I know that it is easy to derive the activity level from accelerometer data [47]. There is a clear difference between the signals from the accelerometer if the person wearing the accelerometer is in rest or if the person is active. So, using these measurements it is easy to say something about the physical activity of a person [48]. Assuming that a student sits still in the classroom during a lecture, and knowing the schedule of a student, one should be able to know what times the accelerometer of a certain student should show a rest state, assuming that the student is attending a lecture.

As can be derived from the previous paragraph, the conclusions about attending a lecture based on accelerometer data is mainly based on assumptions. Obviously this is not desirable. An accelerometer for example does not give any information about its location, so the activity or inactivity could take place somewhere else than in the classroom. However, the schedule of a student could be seen as a more or less fixed pattern of alternation between inactivity in the



**Figure 3: Translation of events to activity levels**

classrooms and activity while changing from current classroom to the next, to attend the next lecture on the schedule. At the secondary school where I followed education for six years from the age of 12 until 18, the lessons all had a duration of 50 minutes with the first lesson starting at 8.30 in the morning. Assuming that it takes 5 minutes on average to change between classrooms, there are 45 minutes spent in the classroom. Furthermore, there is a break of 15 minutes after two lessons and a 20 minutes break after another two lessons. This schedule is drawn in Figure 3. The different events during a day at school can be translated to active and passive time slots.

Advantages:

- Easy processing based on low level data (less detailed).
- Not adhering expected pattern easily recognized.
- Less privacy sensitive; no personal data transmitted, no positioning.

Disadvantages:

- Much assumptions involved.
- No concrete localization.
- Accelerometer necessary for every user.

**Table 8: Effectiveness of applications in accelerometer papers**

<b>Accelerometer</b>		
<b>Problem</b>	<b>Approach</b>	<b>Effectiveness</b>
Activity monitoring with off-the-shelf devices [47]	Use off the shelf components instead of dedicated/custom made hardware to monitor activity and movement	Non-intrusive and potentially easily accepted methodology to monitor and analyze daily activity characteristics using an already familiar handset
Sitting at the desk in the office for longer periods [48]	SunSPOT® measures activity of employee, ambient tool reminds him to take short breaks if necessary	n/a

### 2.3.5 GPS

Global positioning electronic monitoring systems (GPS) are among the most recent form of electronic technologies used to supervise offenders [49]. In the criminal justice system, GPS can be used for detention, restriction and surveillance purposes. The technology eliminates the need for a device to be installed in the wearer's home and is currently being used or introduced in a number of jurisdictions in the United States [19].

The person is monitored to ensure curfew hours are kept. Place-restriction is enforced through an alert that is triggered if the person goes into prohibited areas. Surveillance is achieved by continuously monitoring the person's location. Miniature tracking devices are also currently being developed and tested. These can be implanted beneath the skin and can track an individual's location as well as monitor physiological signs. Although these may be removed using a simple surgical procedure, the potential for civil action for any adverse consequences of the surgery or the implant itself demands serious consideration before any such developments take place. Professional ethical issues also arise for doctors involved in the implantation and removal procedures. In the United Kingdom, there have been indications that the government may consider the use of surgically implanted devices for convicted pedophiles [19].

Although today GPS is integrated in all luxury cars and performs well for navigation purposes, the indoor performance of GPS remains weak. Cars have an external GPS antenna which is, of course, used outdoors. Receiving GPS signals indoors suffers from problems with weak signals up to complete blocking of the signals. [50]

Standard GPS receivers take one to two minutes to search for and acquire satellites. At the same time, they cannot acquire satellites if the signals are much attenuated from the outdoor minimum of -130dBm. A standard GPS receiver, when doing a cold-start, must search the entire frequency/code space. To find the signal they must sequentially search all 1023 code chips, in each adjacent frequency bin. To do this in a reasonable amount of time demands no longer than one millisecond of dwell time in each frequency/code bin. This short dwell period limits the detectable signal strength, and initial acquisition is typically only possible when outdoors with a clear view of the sky [51].

Assisted-GPS works by giving the receiver a hint of which frequency bins to search. The frequency offset is a function of the Doppler shift produced by satellite motion. Assistance is provided by sending the satellite orbit information, or derived information (such as estimates of the Doppler for an assumed location). With this assistance, the frequency/code search space reduces. This results in longer dwell time which increases the sensitivity with approximately 10 dB. However, for proper indoor operation, a processing gain of 20dB to 30dB is necessary. To achieve this, massive parallel correlation is applied, where all possible codes are searched in parallel. This requires more resources but allows a longer dwell time and the processing gain increases as the dwell time does. So, indoor-capability, or more precisely high-sensitivity GPS, is a combination of Assisted-GPS (A-GPS) and massive parallel correlation. Assisted-GPS is an old idea that improves regular GPS performance. Massive parallel correlation is a requirement, only recently feasible in a phone, to enhance A-GPS performance enough to get the 20 to 30 dB of processing gain required to acquire GPS signals indoors [51].

#### Advantages:

- GPS infrastructure (outdoor) already exists and globally available.

#### Disadvantages:

- Requires GPS device for each user.
- GPS does not have very well performance with indoor use.
- With GPS, the device knows where it is, but an additional step is required to inform other interested (supervising) devices about its location.
- Location error of A-GPS is between 30 and 100 meters [52].

**Table 9: Effectiveness of GPS-supported solutions**

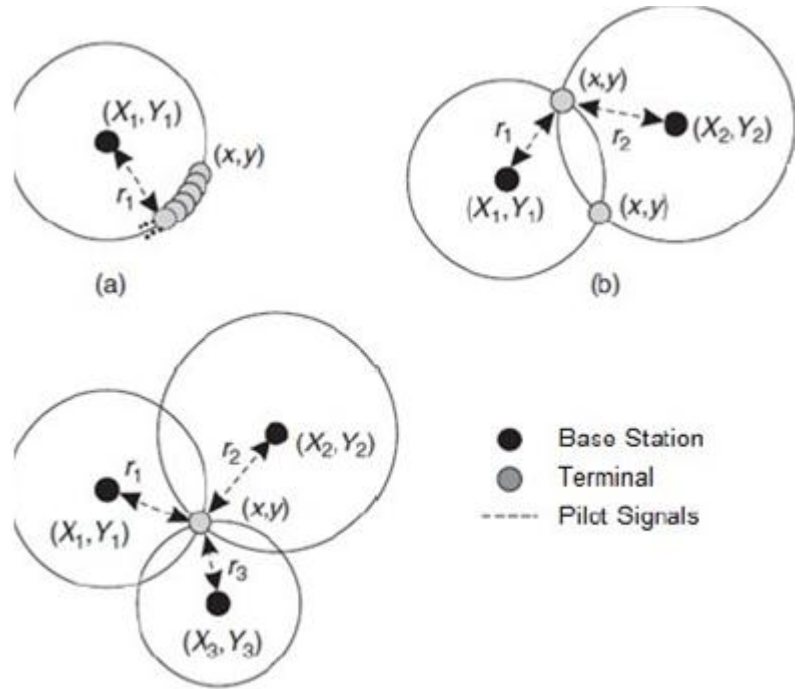
<b>GPS</b>		
<b>Problem</b>	<b>Approach</b>	<b>Effectiveness</b>
Workload during supervision of sex offenders [49]	Use GPS for real-time tracking of sex offenders	GPS reduced the workload.
Lack of prison space [19]	Passive-/Active-/Global Positioning- Systems	Correctional programs have potential, lack of physical restrictions is however a risk factor.
Availability of GPS at Schiphol Airport [50]	Use High-Sensitivity GPS (HSGPS) receivers to analyze availability and performance	GPS could be used outdoor but the indoor performance is too weak
Usual bad indoor GPS performance [51]	Combine assisted GPS (A-GPS) with massive parallel correlation	It is shown that GPS can be made to work practically anywhere in the cell-phone's native environments; inside cars, urban canyons, offices



### 2.3.6 WiFi triangulation

The number of wireless access points (WiFi APs) increases almost daily. Besides the main task of routing internet traffic, there are other so-called secondary tasks a wireless access point can contribute to. One of these secondary tasks is supporting positioning of mobile devices. This positioning is based on range, the fixed position can be calculated by means of a circular lateration, determining the intersection of the circles formed by the radii of the target in relation to nearby base stations.

To determine the ranges, one can take into consideration the signals intensity received by the terminal from the nearby base stations. If the terminal is receiving more than a signal from surrounding base stations, and some of these come with greater intensity than others, it means that the terminal is closer to the base-stations whose signal received is stronger than those in which the signals perceived by the terminal are weaker.



**Figure 4: Circular triangulation**

Depending on the received signal strength, the approximate distance  $r_1$  between the mobile device and the base station is calculated [53]. This results in a circle around the base station and the mobile device could be anywhere around that base station at the calculated distance, as depicted with the circle in Figure 4a. Now the same calculation is performed for another access point from which the signal strength is received and this will result in another circle with radius  $r_2$  around that other base station. The circles will overlap which results in two intersections as depicted in Figure 4b. If now the distance to a third access point is calculated, a third circle can be drawn in the picture and the intersection point of the three circles is then determined as the location of the mobile device, as in Figure 4c. Usually, to ensure proper positioning, all system implementations are preceded by a training phase where a sensor map of the environment is built [54].

#### Advantages

- Relative high availability of wireless access points in buildings [55].

#### Disadvantages

- High energy consumption [53].
- High density of access points necessary for precise positioning.
- Deviation in accuracy due to estimation of distance to AP based on signal strength.

**Table 10: Effectiveness of WiFi triangulation approaches**

WiFi triangulation		
Problem	Approach	Effectiveness
Localization of ambulances in large cities [53]	Android platform based solution that discovers and sends its current location via SMS to central server	Accuracy of 50 meters, which can be improved by applying advanced mathematical techniques
Determination of the position of a mobile device [54]	Build a sensor map of the environment and measure RF signal strength of mobile device	Localization of stationary device is accurate within 1.5 meters over 80% of the time, track a moving device to within 1 meter over 50% of the time
Location privacy in public WLAN [55]	analyze mobility of neighboring nodes to decide which mobile nodes are going to update their MAC addresses	It can protect user's location privacy quite well and can reduce the number of MAC address change rounds quite a lot

### 2.3.7 GSM cells

Since a mobile phone is always connected to at least one base station, which is always the nearest base station, it is possible to derive the location of a mobile phone and accordingly its user on a certain level of accuracy. Based on the location of the base station to which the device is currently connected. The question is however, if it will be possible to gain access to this data since this will decrease the privacy of the users of the network. Additionally, the location error of positioning based on GSM cell-ID lies in the interval of 100 – 1000 meters [52].

An approach that might be more successful in retrieving GSM antenna data is having a private GSM antenna. If this antenna is privately monitored and controlled then you will be able to look into the administration logs where the connected devices are registered. An example of such a private GSM device is the Private GSM Server fabricated by the Dutch company RadioAccess B.V. [56].

Advantages:

- Mobile phones are always "online," i.e. always have a connection with a base station in the vicinity.

Disadvantages:

- Statistics and user information might not become available from telecom operators.
- Accuracy probably not high enough, localization not exact due to cell size.
- And with a private GSM system, if someone is connected with the private GSM system belonging to a school, he is not necessarily in the classroom where he is supposed to be. He could also be in the cafeteria or probably even outside the school but only passing by along the street.

**Table 11: Effectiveness of GSM based approaches**

GSM cells		
Problem	Approach	Effectiveness
Accurate GSM based positioning [52]	Use TA (timing advance) parameters from all visible BTSs (Base Transceiver Station)	The positioning performance based on GSM increases but other technologies like A-GPS are still more accurate
Use of GSM network in office instead of PSTN [56]	Private GSM server at the office	Same devices (mobile phones) can be used to make calls via PSTN in the office. Saves a lot of money

### 2.3.8 Software based (Topicus)

Registration of absent students often still happens manually, also for example at the Etty Hillesum Lyceum in Deventer where I talked to the ICT manager (Bas Penris) at October 15<sup>th</sup>, 2010.

The procedure is that during the day at school, the teachers write down the names of the pupils who are not present. This information will be entered in a student information system, a software program built by Topicus that keeps track of the progress of the students and maintains other information, for example the absence. By default, the status of a pupil is set to "present." It is possible for the teachers to change this status to "absent."

For each of the different phases of education in The Netherlands, Topicus developed a student information system. This software is further described below.

**ParnasSys** is a pupil information system for primary education. It stores information about general behavior, performance and all other information about a pupil that is relevant for an educational institution. Regarding the presence indication, with ParnasSys it is possible that different parts of the day (morning, afternoon) can have different presence statuses. Since June 2009, there are already 3000 primary schools connected to ParnasSys. The number of pupils registered with ParnasSys is approximately 750,000. With 35% is Topicus leader in the market of pupil information systems for primary education.

**Vocus** is a student information system that is used for secondary school students. Since students at the secondary education institutions have different lessons every single hour, Vocus allows registration of presence information for every hour separately. The Vocus application is used by a small group of institutions. The database contains approximately 50,000 students. Since February 2010, Vocus is called "@VO 3.0."

**EduArte** is a student information system that is used for Dutch MBO institutions. It is the successor of NOISE, the system that is currently most in use at this type of educational institutions. As of April 2010, EduArte is up and running.

**Table 12: Effectiveness of Topicus software solutions**

Topicus Software		
Problem	Approach	Effectiveness
Proper student administration	Student information systems ParnasSys, Vocus, Eduarte	Able to store and edit a lot of data concerning a student

## 2.4 Discussion

The different solutions are able to localize people at different levels of accuracy. Positioning based on GSM cells, either public or private cell, will never be very precise. This is due to the fact that a private cell will cover an entire building, for example a school building but a public cell will even cover a larger area than only one building.

GPS systems are more precise in positioning but they have bad performance inside buildings.

Wireless access points, which are densely present in buildings these days, can be used for accurate positioning for example by applying the triangulation principle.

Depending on the chosen infrastructure, RFID systems can perform reasonable good (periodical) positioning, or at least control the attendance registration by registering the check-in events of student cards. The accuracy depends on the chosen infrastructure. Considering a ring or tier based structure, an inside and an outside ring (or tier) are distinguished. The outside ring registers the attendance at building level, where the inside ring registers the attendance at classroom level.

The same holds for the cases where biometrics is used as personal token for authentication instead of RFID cards.

The solution in which accelerometers are involved, as described in Section 4.4, is not performing any positioning at all. In this case, the attendance monitoring is based on an expected behavior (activity/movements) pattern and the deviations with respect to this pattern.

For several of the different solutions holds that a token is involved which is used to authenticate every single user. This might be an RFID chip card, a mobile phone, an accelerometer or a biometric sample such as a fingerprint. Although for some of these tokens it is less likely to accidentally forget, like for example your mobile phone, it might happen that you forget for example your RFID card. This might lead to a dilemma. On the one hand, you want the students to attend the lectures, but if they are not able to authenticate/check-in by means of their token, would the access to the school or classroom be denied? In that case your own system will have a negative influence on the attendance while the system is designed to have a positive influence on the attendance. This problem obviously does not occur in case biometrics is used.

Every system built for authentication is subject of cheating. Students might for example want to pretend that they are present while they are actually absent. This seems easy in case of RFID cards. However, if teachers supervise the check-in event, cheating will be less easy. With the use of accelerometers there is the problem that one could also wear another student's accelerometer, generating an activity pattern. As with an accelerometer, students could also

bring other student's mobile phone and generate data traffic, pretending the presence of the physically absent student. The same holds in case the attendance is based on GSM data. It is however less likely to bring another's mobile phone than bring other student's RFID card or accelerometer. Of course this problem does not occur with the use of biometrics.

Despite the fact that biometrics cannot be forgotten and are hard to cheat with, they need to be shown at every single verification moment. The same holds for RFID cards. This might become annoying and will possibly also be forgotten after a while. Solutions that are based on data traffic analysis do not suffer from this problem. It does not require a number of user initiated events every day to show a user's presence. This data traffic analysis works in the background and is not recognized by the visitors of the building. This makes the system less dependent of the cooperation of the users. It however requires a reasonable amount of data traffic. The accelerometer proposal is also working in the background and does not require check-in and check-out events from users. The use of accelerometers relies however more on assumptions and interpretations of the pattern of its output signals and can also be carried by a fellow student and so is generating the expected pattern while the dedicated user is in fact absent.

According to [45], detecting the presence of (mobile) nodes is a largely unexplored field. Although there are a number of research directions that deal more or less directly with the presence of nodes in wireless networks, this covers ad-hoc networks where the devices periodically send presence information into the network, while the current goal is the achievement of presence detection based on already available information, so without sending and receiving presence information messages.

In this section, several approaches of analyzing data traffic are proposed. As can be seen, there are various methods to extract useful information from data traffic, which will be generated by students anyway. This means that, assuming that students generate enough data traffic, there is no additional action required and we could use data which is already available in everyday life. The challenge is to detect presence and identify mobile nodes and in line with this, to identify the users (students) and confirm that they attend their lectures. As mentioned before, this is a rather unexplored field of research.

## Chapter 3 Classroom observations

As part of the research, several lectures have been attended at different schools for secondary education in The Netherlands. The visits to the schools have been made to investigate the behavior of students in general and the behavior of students with respect to modern media especially. This investigation has been performed in two different ways. First, students were observed to get acquainted with their behavior. This is necessary because the development of the truancy detection system should be adapted to the actual behavior of the students. Second, the school visits also gave the opportunity to do a survey among the students. Chapter 4 further elaborates on this subject and contains the analysis of the results following from this survey.

This chapter reports about the observations that have been done during these visits. For the observations, a standard form was used to make notations. This "Classroom Observation Form" is depicted in Appendix A.

### 3.1 Research methodology

The so-called "Classroom Observation Form" contains a number of fields that are filled in at the start of each lecture. This gives information about the school, class type and number of students that attend the specific lecture. There are also three technical features investigated. First is checked if there are any computers or laptops in the classroom. Next, with a mobile phone is checked whether there is a wireless network available. Last, the same mobile phone is used to do a Bluetooth scan, which searches for other Bluetooth devices in the vicinity.

Divided over 8 days, 40 different lectures have been attended in total. All educational levels (VMBO, HAVO and VWO) were visited and all study years, ranging from 1 up to 6 have been attended. The courses which were taught were also very diverse

### 3.2 Phones and usage

During the visits at the different schools it has been seen that at each of the schools mobile phones were used during the lecture, except for one school. This single school, at which no mobile phone use was recognized during the lectures, is a school for so-called "special education." Students at this kind of schools require extra and/or special attention because they suffer from behavioral problems or learning disabilities. For those two reasons, the groups are small (15 students at maximum) which results in a more strict and dense supervision.

Furthermore, their rules and the adherence to them are expected to be more strict than at average secondary schools.

During more than half of all attended lectures, phone use has been recognized while sitting the entire lecture at the same place in the back of the classroom. Due to the relative high number of students to supervise per lecture and the fact that views might be blocked by bodies of students or other materials like workbooks, phone use during lectures might be more intensive in reality. This also follows from the statistical research as can be read in Chapter 3.

More or less intensive phone use also partially depends on the intensity of supervision, normally performed by a teacher. It was for example clearly visible that during starts and ends of lectures, when students pack and unpack their bags, more phones were used than moments less close to start and end, i.e. when the lecture is approximately halfway. Chapter 5 reports about a practical experiment covering internet use with mobile phones at school and will elaborate on this.

Another fact is that during every single lecture it was visible that students have their phone with them all the time, either in the pocket of their pants or at another place in their direct vicinity.

Regarding the devices, the mobile phones used by the students, almost all phones are smartphones, very modern state of the art technology. It seems also that, although never explicitly stated, that having a (brand) new mobile phone is a kind of never ending competition among students in secondary education. This is in line with the expectations about smartphones for the next four years; it has already been forecasted that the number of smartphone users around the world will increase exponentially [57]. This is an advantage in the context of the truancy detection system, because the newer the device, the higher the chance that it is equipped with modern connectivity interfaces like a WiFi port with which the wireless network of the school could be accessed.

### 3.3 Bluetooth

While attending lectures, there was enough time to perform some additional “on site” research on technical issues. One of the research subjects was in the context of Bluetooth. The initial reason was to discover whether there are any mobile phones in the near area without the need to physically see one. Obviously this requires that mobile phones have activated their Bluetooth port and is set into visible mode. During 39 of the 40 lectures, one or more devices were found doing a Bluetooth scan. The average number of Bluetooth devices per lecture is 4.7 so almost 5 devices were recognized per lecture.

The names of the devices, which become visible after a scan attempt, can roughly be divided into three groups. In the first group the devices have the name of their owner. In the second



group, the devices have the name of the brand and type of the mobile phone. The third group consists of nonsense names like for example the occurrence of a device with the name "*Geiten*" (*English: goats*).

This small research offers interesting insights into usage by and attitude of students regarding mobile phones at school. Besides, the number of devices which had their Bluetooth port enabled was higher than initially was expected. Bluetooth is however not a suitable technique for application in a truancy detection system because of the ad-hoc infrastructure. This does not allow central monitoring of data traffic so presence detection within a school building would be very hard. Additionally, Bluetooth is not used as often as WiFi because it is intended for small amounts of data transfer and mostly during short time intervals.

### **3.4 School policies regarding mobile phones and adherence**

Most of the schools for secondary education in The Netherlands have a policy regarding the use of mobile phones inside the school building. The majority of these policies state that phone use in general is prohibited in the classrooms during the lectures. This policy also held for the schools that have been visited in the context of this research. It has been observed however that individual staff members as well as the managing board of the school did not fully adhere to this policy at all times.

It turned out that despite the policy the students carry their phones with them during the lecture and do not hesitate to use the devices. So teaching staff attempted to benefit from this, which seems to be successful at least at specific points. Students were allowed to use the calculator function of their mobile phone. Also, students were allowed to listen to music with their mobile phone (of course by using earphones) while making exercises independently. Another staff member allowed his students to make notations with their phones. Remarkable fact was that these students verified their spelling with their teacher more than when using traditional pen and paper. This is a positive effect in the context of education caused by the usage of mobile phones in the lecture.

### **3.5 Music**

Listening to personally selected music is very important to a major part of the students. They take any available chance to listen to their own music. This behavior could support the recognition of students online. They for example download music using iTunes, or stream music from YouTube or other websites. These services require login events to access personal accounts which amongst others contain playlists and/or credits.

These login events can be used as input for the truancy detection system. The attitude of students towards listening to music develops the idea for making a music service or music server at school, to which students can connect. This is a possible source of data traffic which could be used as input for the truancy detection system.

### 3.6 Other interesting facts

One of the students said that McDonald's does have free WiFi while school does not. This was said in a way that the students would have multi-benefit from going to McDonald's; first they do not have to attend lectures, second they have the opportunity to eat snacks and third they have the opportunity to connect to the free WiFi using their mobile phones. So, the presence of free WiFi within a fast food restaurant near a school without WiFi, might be a reason to play truant. (Food and free internet in one place is attractive to students).

At several schools covered in this research, a so-called "smartboard" is used to support the lectures. These boards are very multifunctional. They are connected to the pc of the teacher to the teacher can export his own screen to the smartboard and show everything he does on his pc, for example show pictures or notations, or visit (educational) websites with additional information, demo's and animations which could support the clarification of the lecture contents. The smartboard is also able to act as a traditional blackboard. It is a full touch screen on which can be written using a special pen, similar to a stylus.

Further developments could include integration of this system with mobile phones of students, so that students for example could post questions to the smartboard, using their mobile phones.

### 3.7 Conclusions

For almost all students in secondary education holds that a mobile phone is a very important thing to possess, and to have with you all day. There is no visible difference recognized between the different schools regarding the attitude of students towards mobile phones.

Bluetooth is not a suitable technique to use in a truancy detection system due to its decentralized infrastructure.

Allowing students to listen to music and let the school facilitate a service for this, has the potential to become an additional source of data traffic which would increase the amount of input for the truancy detection system.

## Chapter 4 Statistical analysis

This chapter has a slightly different outline than the rest of this thesis. It might also contain some redundant information. This is due to the fact that this chapter is intended to be published in the near future.

# Towards automated truancy detection

R.H.J. Ottenhof, P.H. Hartel, M. Junger

*University of Twente, Enschede - The Netherlands*

## 4.1 Introduction

The reason for this research project is partially based on the aim of the Dutch Ministry of Education to drastically decrease the number of school dropouts. The Ministry wants to reduce the number of dropouts, currently 40.000, to at most 25.000 in the year 2016 [58].

Fact is that truancy is an indicator for future dropouts [59]. So to reduce the number of dropouts, one of the possibilities is to reduce truancy among students.

Meanwhile, a trend is visible namely the increasing digitalization of education. With that, teachers are now expected to fill in absence information directly in the school administration system. However, this takes too much time according to these teachers, and since they are teaching staff and not administrative staff, they refuse to perform this task. Then there is also the group of teachers who say it is too difficult to use the administration system. Due to these reasons, absence registration on several schools decreased with 50%.

Electronic/automated truancy detection is based on the assumption that in general students have a mobile telephone, use their telephone during school time, and are online using their mobile phone. Also, it assumes that the students who are most likely to be truants are at least just as likely as other students to have and to use their mobile telephone. If truants are less likely to be online, electronic/automated truancy detection is not an adequate mean to detect truancy.

To investigate these questions we developed the present study. The study aims to answer the following questions:

1. How many students have a mobile telephone?
2. How often do they use it during school time?

3. Are they using the internet with their mobile phone during school time?
4. What are they doing when they are online?
5. Does 'mobility predict truancy'?

Furthermore, differences between boys and girls will be investigated and it is attempted to indicate trends in order to find out if and how a variable changes over time.

Additionally, a typical person, or group of persons, will be described, namely the "online truant." More specific, these are boys and girls who play truant today and are online with their mobile phones on a regular base.

## 4.2 Method

### 4.2.1 Sample

Data were collected by means of a convenience sample of schools for secondary education in The Netherlands. However, care was taken to collect data in schools at different locations, to cover the entire age-range of high school students, and to include all Dutch school-types which contain VMBO, HAVO and VWO.

First, Students from all educational levels, namely, VMBO, HAVO and VWO filled in the questionnaire.

Furthermore, different geographical locations within the Netherlands were selected for the research.

In total 49 schools were approached, in 14 cities. 7 schools participated in the study, divided over 4 cities. In each school, data were collected in different classes with respect to education level and study year.

Data were available from two schools in Gorinchem, two schools in Oldenzaal, one in Oss and two in Raalte.

### 4.2.2 Data collection

Knowledge has been gathered in two different ways.

1. One of the researchers made observations while attending a total of 40 lectures, divided over the different schools.
2. After each lecture the students were asked to fill in a questionnaire. So, data were collected by means of a written questionnaire during class. Anonymity of the students was guaranteed because students did not have to fill in their name.

#### 4.2.3 Concepts

- *Mobile telephone use:* 20 questions were asked about whether students have a mobile phone and how and how often they use it, especially at school.

**Table 13: Variables in the context of mobile phone use**

1	How many mobile phones do you have?
2	What make and type is your mobile phone of?
3	How long do you own this phone yet?
4	What is your main reason to buy a new / other phone? (subscription end / phone lost stolen / follow trends / phone broke / n a)
5	What kind of subscription do you have, or do you have pre-paid? (1-year / 2-year / monthly / pre-paid)
6	Have you got internet facility on your mobile phone?
7	Have you got a WiFi port in your mobile phone?
8	How much time (in minutes, on average) do you use your phone during lectures?
9	How much of the time in (8) is spent using the internet?
10	How often do you use your mobile phone between lectures and during breaks?
11	How much of the time in (10) is spent using the internet?
12	What is your main purpose to use your phone? (sms / call / internet)
13	Do you use any other devices at school to communicate with others?
14	Is there a wireless network available for use at your school?
15	Do you use this wireless network?
16	If there would be a free wireless network in your school, would you use it?
17	What is your main purpose to use the internet? (e-mail / chat / read news / social networks / watch movies / internet calls)
18	Which website is your favorite?
19	Have you got an account with Hyves?
20	Have you got an account with any other social network site?

- *Truancy:* 6 questions were asked to investigate intensity and motivation for truancy.

**Table 14: Variables in the context of playing truant**

1	Did you play truant in the past?
2	Do you still play truant today?
3	Do you experience the control on truancy as strict?
4	What is your most important motivation if you play truant? (slept bad / job / didn't prepare test / help family / other)
5	What should be changed at school that causes you to play truant less?
6	Is there anything you want to do with your phone which is prohibited at school?

- *Control variables (2)*: Information was asked about sex and which class students are in. In the results only age differences are distinguished based on the school class a student is in.

#### 4.2.4 Analysis

We investigated the frequencies of mobile phone use and truancy. We systematically checked whether mobile phone use and truancy differed by sex and by age, using cross-tabular analysis. Only statistically significant differences are reported.

We also investigated whether mobile phone use and the socio-demographic factors sex and age are related to frequency of truancy, using multiple regression analysis.

With stepwise linear regression it is possible to investigate how different variables have influence on today's truancy, while all these variables are taken into account at the same time. The outcome of every single figure is adjusted for the effects of the remaining variables that are used in the model.

### 4.3 Results

In this chapter all interesting results following from the questionnaire are displayed systematically. Further analysis and discussion about these results will be reported in Section IV.

#### 4.3.1 Sample

A total of 701 students completed the questionnaire.

The gender of the respondents is almost equally distributed, 50% females (352) and 50% males (349).

Students of all classes ranging from 1 up to 6 were involved. This means that the age of the respondents is in the range from 12 to 18 years.

Also, students from all different education levels in The Netherlands, VMBO, HAVO and VWO have filled in the questionnaire.

During the all the lectures that have been visited, 16 students were marked as absent. This holds that in total 717 students could have filled in the questionnaire. Since 701 students indeed filled in the questionnaire, the response rate of this survey is 97.8%.

#### 4.3.2 Descriptives

Descriptives of the discrete research results from the survey are reported in Table 15 and Table 16. The next sections elaborate on the figures in these tables.

**Table 15: Variables split for men/women**

	Sex	
	Male N = 349	Female N= 352
Student has at least one mobile phone	98,0%	100%
Student owns phone for less than a year	63,6%	63,5%
Student has mobile phone with internet	55,5%	52,0%
Student has mobile phone with WiFi port	35,3%	30,8%
Student uses mobile phone during lectures**	58,4%	77,0%
Student uses internet on mobile phone during lectures	22,0%	25,8%
Student knows whether WLAN is available in school*	66,9%	57,8%
Student has <i>Hyves</i> account**	89,6%	96,9%
Student has account with other social network**	60,8%	75,1%
Student played truant in the past	50,1%	43,0%
Student plays truant nowadays**	19,4%	11,7%

**Table 16: Variables per study year**

	Year				
	1 N=60	2 N=109	3 N=278	4 N=237	5+6 N=17
Student has at least one mobile phone	98,3%	99,1%	98,9%	99,6%	94.1%
Student has internet on his mobile phone	55,4%	61,1%	52,4%	53,0%	31.2%
Student has mobile phone with WiFi port**	21,1%	29,2%	32,1%	39,7%	20,0%
Student uses mobile phone during lectures*	48,2%	54,6%	73,1%	71,7%	81.2%
Student uses internet on mobile phone during lectures	14,9%	18,6%	26,2%	27,2%	7.14%
Student has account with <i>Hyves</i> **	81,7%	92,7%	94,9%	94,1%	100%
Student has account with other social network	65,0%	68,8%	69,3%	67,8%	52.9%
Student played truant in the past**	18,3%	24,8%	44,9%	65,3%	52.9%
Student plays truant nowadays**	5,0%	6,4%	15,4%	21,2%	35.3%

\*  $p < 0.05$ , \*\*  $p < 0.01$

#### 4.3.3 Mobile telephone use

*Mobile phone ownership and use:* Only 7 out of 701 respondents do not have a mobile phone (1%). The respondents who do not have a mobile phone are equally distributed over the six different years. However, all seven students without a phone are male respondents.

Most of the respondents do have only one mobile phone (68.9%). 20.1% has two mobile phones, 5.3% three and 4.6% even more than three mobile phones.

Almost half of the students have a *Samsung* phone (48%). Second largest group form the *LG* users with 19%. Third is *Nokia*; 11%. Only 2.5% of the respondents have an *iPhone*. 17.5% had another brand like *HTC*, *BlackBerry* or *SonyEricsson*.

*Length of ownership and reason of change:* Only 36% of the students have the same phone for over a year. This means that 64% own their current phone less than a year. 37% own their phone even less than half a year.

*Reasons for change:* The main reason to change to another phone is that the phone is broken, this is the case for 44.4% of the students who changed telephone during the past year. "Trend following" is for 21.6% of the respondents the main reason to change phones. "Subscription continuation" however, is only in 11.6% of the cases the reason to change to another phone.

*Type of subscription:* 28.4% of the respondents do have a subscription with a one or two year contract. In that group (N=191) only 28% has a one year subscription, which turns out to be only 8% of all respondents. Most of the students (68%) use pre-paid SIM cards.

54% of the students say they have internet on their mobile phone. 33% of all students have a WiFi port on their cell phone with which they could access the wireless LAN in the school.

Another 25% say that they are not sure whether or not their phone is equipped with a WiFi port. This means that 42% does not have a phone equipped with WiFi.

*Mobile phone use during the lectures:* 68% of the students say that they use their phone during the lessons. From this group (N=433), 33% indicates that they access the internet while using their phones during a lecture. Regarding this usage of phones during the lectures, there is a difference between boys and girls since 77% of the girls but only 58% of the boys use their phone in the classroom ( $p < 0.001$ ).



*Mobile phone use between the lectures and during breaks:* 86% of the respondents use their phone during a break or while going from one classroom to another in order to attend the next lecture. At these moments, 38% of the phone usage involves accessing the internet.

Regarding this phone use between the lectures, again this significant difference between boys and girls shows up, where 92% of the girls but "only" 75% of the boys make use of their cell phones between lectures and/or during breaks. The corresponding p-value is 0.001.

*Division into main purposes "call," "sms" and "internet":* 61% answers that the majority of their mobile phone usage is in the category "sms." Another 28% indicates that besides another of the two remaining categories, "sms" is one of the main purposes of their phone usage. Only 5% say that "internet" is the main activity performed with their phones and an additional 11% indicates that besides another category, "internet" is one of the main tasks while using a mobile phone.

*Other communication devices:* 6% of the respondents use an *iPod* to communicate with others. The computer (or laptop) is used by 41% of the respondents as communication device.

#### 4.3.4 Wireless networks at school

*Knowledge of network availability:* 37% of the respondents state that they do not know whether or not a wireless LAN is installed in their school. There is even one school in this research which has a wireless network available for their students to use but still 37% of its students said they do not know whether a wireless LAN is available. 24% of the students at that particular school even say that there is no wireless network while there actually is one.

Regarding this knowledge about the availability of wireless networks, a significant difference exists between boys and girls where the boys seem to be better informed than the girls. 67% of the boys know about the presence of a wireless network in the school. Among the girls 58% know about a wireless network at their school. The corresponding p-value is 0.003.

*Use the wireless network if available:* 79% says they would definitely use the school's wireless network if it had one and if it would be accessible for free. An additional 16% says they would probably use it. Only 5% say they would not use free WiFi at school.

#### 4.3.5 Internet use and online social networks

*Different activities on the internet:* Six categories have been defined as possible main purposes. These categories are "E-mail," "Chat," "Social networks," "Watch movies," "Read news" and "Internet calls." Chatting and updating social networks seem to be the two most important purposes since both are marked by 66% of the respondents as one of their main activities on the internet. The next most important purpose is watching movies, mostly via YouTube. 49% of the respondents stated this as one of their main purposes during internet usage. E-mail is for (only) 33% of the students one of their main purposes. Reading online news portals is marked popular by 9% of the respondents and making internet calls is popular with 7% of the respondents.

The popularity of online social network is also shown by the answers to the question "what is your favorite website," where 60% of the respondents indicated that *Hyves* is one of their favorite websites while an additional 16% indicate that another social network site like *Facebook* or *Twitter* is among their favorite websites. 25% says YouTube is one of their favorite websites.

*Accounts with social network sites:* 93% of the respondents have an account at the Dutch social network site *Hyves*. Furthermore, also 68% of all respondents have an account at (at least one) social network site other than *Hyves*, mainly *Facebook* or *Twitter*.

There are also a few respondents without a *Hyves* account but with an account at another social network site. All figures about online social network accounts are displayed in the Venn diagram below (Figure 5). Most important fact is that only 25 out of 689 students do not have any account with a social network site. This brings the total percentage of respondents who hold an account at a social network site to 96%.

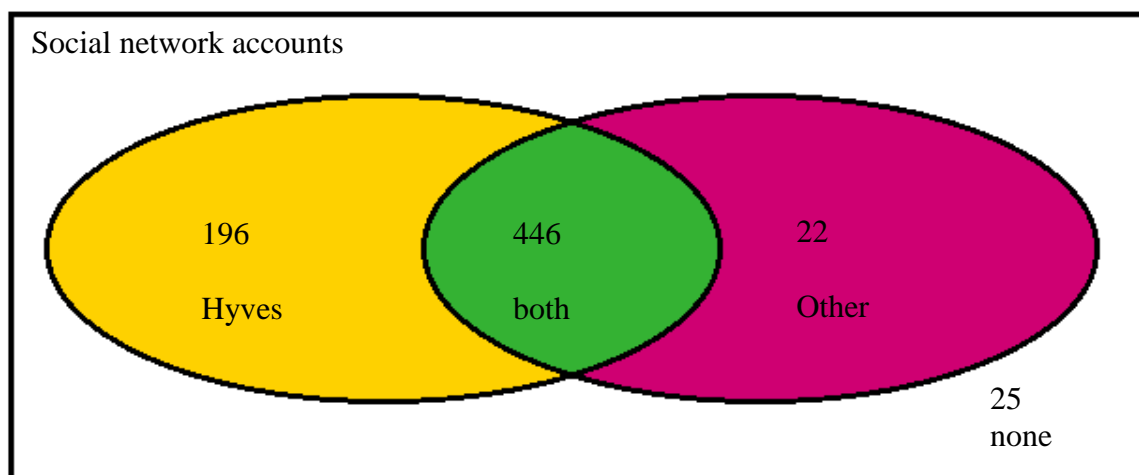


Figure 5: Venn diagram accounts social network sites

At *Hyves*, girls seem to be the majority since 97% of the girls have a *Hyves* account while “only” 90% of the boys have a *Hyves* account. This is a significant difference with chi-square  $\chi^2 (1, N = 698) = 14.6, p < 0.001$ . The same difference occurs in the answers to the question “do you have another social network account?” where 75% of the girls but only 61% of the boys answer “yes” This difference is significant  $\chi^2 (1, N = 689) = 16.2, p < 0.001$ .

Focusing only on the students who indicate they play truant, figures show that 97% of them have an account with any social network site.

#### 4.3.6 Attending lectures

*Truancy in the past and present:* It is expected that the figures about these questions might turn out to be higher in reality because probably not all students will have truthfully filled in questions about truancy. Nevertheless, 47% of the respondents stated they have played truant in the past. From this group (N=323), 32% still plays truant today, which happens once a month per student on average.

The overall rate of truancy today among all respondents turns out to be 16%. Regarding truancy that occurred in the past, a difference between boys and girls is shown by the figures which indicate that 43% of the girls but 50% of the boys played truant in the past. This difference is not significant:  $\chi^2 (1, N = 698) = 3.559, p = 0.059$ .

Considering today's truancy, the same difference occurs since 12% of the girls say that they still play truant every now and then, against 19% of the boys. This difference is significant  $\chi^2 (1, N = 694) = 7.772, p = 0.005$ .

*Reasons for playing truant:* Main reason (18%) is the fact that students did not learn for a test they had to make the specific day they play truant. Second reason is having slept badly the night before (15%). Other important reasons are that they had to help their family or friends (10%) or they simply planned to do something else instead of attending the lectures (8%). Playing truant in order to be able to work at a part-time job is only in 2.6% the reason of absence. Boys and girls share their opinion regarding why they would play truant.

*Strictness of the control on truancy:* Students experience the strictness of the control on truancy differently. 13% of all respondents say the control is not (too) strict. 51% thinks that the strictness is average but still 36% says that the control is very strict.

Calculating the same figures for the group of students who still play truant these days shows that 19% think that the control is not (too) strict, 46% find it an average strictness and 35% say that the control is very strict.

#### 4.3.7 Multiple regression analysis

The figures from the regression analysis are shown in

Table 17. Column B shows the degree of influence and whether it is directly proportional (positive values) or inversely proportional (negative values) with the degree of truancy playing. The more the values in column B deviate from zero (positively or negatively) the higher the influence of this variable on truancy. The dependent variable (first row of Table 17) is "plays truant nowadays." This regression analysis investigates the influence of all variables in the "model" -column on this dependent variable.

Column "Beta" shows the standardized coefficients. These Beta-weights allow for comparing the strength of each predictor. It shows for example that the number of mobile phones and the length of ownership almost have the same degree of influence on truancy, while being a prepaid user or not is of less influence.

The last column in the table ("Sig.") contains the p-value and indicates whether the relation between the specific variable and playing truant is significant. The significant variables in this regression analysis are printed bold and are reported below.

This regression analysis is performed in two steps, which results in two models (1 and 2). In the first model only sex and school class are taken into account in order to investigate the effect of only one of those two variables, adjusted by the effect of the other one. The variable School class distinguishes age differences with a range between 12 and 18 years.

In the second model, all variables in the questionnaire (see Table 17) which turned out to have a significant relation with the variable "playing truant nowadays" are added to the variables from the first model in order to investigate each variable individually, adjusted for the effect of the other variables taken into account.

**Table 17: Two step multiple regression analysis**

<b>Dependent variable: Plays truant nowadays</b>					
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	-0,082	0,063		-1,30	0,193
Sex	<b>0,074</b>	0,028	0,103	2,62	<b>0,009</b>
School class	<b>0,059</b>	0,014	0,160	4,06	<b>0,000</b>

2	(Constant)	-0,155	0,092		-1,69	0,091
	Sex	<b>0,115</b>	0,029	0,160	3,91	<b>0,000</b>
	School class	<b>0,051</b>	0,014	0,139	3,56	<b>0,000</b>
	Number of mobile phones	<b>0,042</b>	0,019	0,096	2,28	<b>0,023</b>
	Time the phone is owned	<b>0,036</b>	0,016	0,094	2,30	<b>0,022</b>
	Usage during lectures	0,016	0,014	0,053	1,14	0,253
	Usage between lectures	<b>0,030</b>	0,010	0,138	3,07	<b>0,002</b>
	Prepaid user	0,048	0,030	0,063	1,58	0,115
	WiFi port	0,025	0,033	0,032	0,753	0,452
	Uses free WiFi at school	0,008	0,036	0,009	0,226	0,821
	WiFi available at school	0,049	0,031	0,062	1,60	0,111

Since girls are coded 1 and boys are coded 2, the positive value for sex in column B shows that boys on average play truant more than girls. In model 2 sex is the factor that has the highest influence on playing truant among all other variables that are taken into account since the (standardized) Beta value is the highest for this variable. If however only sex and school class are taken into account, it follows from model 1 that school class is of higher influence than sex. So, the influence of sex becomes higher if more factors are taken into account.

The positive values for school class in column B show a directly proportional trend which states that "the older the students, the more they play truant."

The number of phones a student owns is directly proportional with the amount of time he plays truant. This holds that students who have more phones, play truant more intensively, on average.

The length of the ownership is inversely proportional with the amount of time a student plays truant. This shows that the students with the newer mobile phones are more likely to play truant.

The positive value for "usage between lectures" in column B, shows that truancy playing students use their phone more intensively between the lectures than students who do not or

play less truant. The same trend is visible for students who use their phones during the lectures but the relation with truancy is not significant for the variable "usage during lectures."

Summarizing the statements above it can be said that sex, school class, number of mobile phones, length of phone ownership and the usage of phones between the lectures are predictors for truancy among students in secondary education.

Furthermore, intercorrelations between the similar variables "phone usage during lectures" and "phone usage between lectures" and "internet use during lectures" and "internet use between lectures" are all strong and significant (all p-values < 0.01) in a specific degree as shown below:

**Table 18: Intercorrelations between measures of media usage**

Measure	1	2	3	4
1. Phone usage during lectures	--	0.444	0.623	0.445
2. Phone usage between lectures		--	0.273	0.348
3. Internet use during lectures			--	0.705
4. Internet use between lectures				--

## 4.4 Discussion

The results in the previous chapter show clear figures about mobile telephone use among secondary school students. The chapter also reports about the amount of truancy played by the same youth. This chapter will elaborate on these two subjects and at the same time there will be reported about the relation between the two subjects.

### 4.4.1 Mobile telephone use

Almost all students in secondary education in The Netherlands have a mobile phone (99%). The majority has only one phone (69%) but still a minority of 30% owns more than one phone. Students in secondary schools seem not very attached to their mobile phones. Although it seems from the classroom observation sessions that the device is very personal, this probably only holds for the content of the phone, like messages and photos, the physical device seems to be easily replaceable. Students change from one phone to another relatively often: only one third has his phone for over a year.

There can be different causes for this fast development. First, mobile phones are highly subject to trends, especially in the world of secondary schools students where having the newest mobile phone sometimes almost turns out to be a competition. Second, the technical development of phones and their performance and capabilities all increase rapidly. This causes a wide range of choices among phones. Third, phones are easily given away “for free,” for example with almost all new subscriptions. These subscriptions have a contract with duration of one or two years. If the subscription is continued, the customer can get a new phone “for free” along with the prolongation of the subscription.

In the context of changing reasons, this continuation of subscription is however only the third most important reason to change to another phone, after “phone broke” (most important) and “follow the trend” (second most important).

The relative low percentage of occurrence of the first reason might be caused by at least two facts. First, the overall percentage of subscriptions is low, in contrast to the number of pre-paid users. Second reason is the possibility of “SIM-only” -subscriptions with which the monthly costs for a subscription can be reduced significantly. In this case, as the name of the subscription already states, only a SIM card is provided with the subscription but no phone. Also, at prolongation of this kind of subscriptions, the opportunity to get a brand new phone for free does not apply.

It is expected that the high number of pre-paid users (68%) mentioned in the previous paragraph, is caused by parents who buy a phone for their child to be able to reach them whenever necessary and to allow their children to call in case of an emergency, but at the same time want to keep the costs under control instead of providing unlimited credit which would probably be used for extensively sending text messages, since this research indicated that sending text messages is by far the most popular activity when using a mobile phone.

Several factors can be identified that cause this popularity of the "short message service." First, sending text messages is very easy with every mobile phone because the device already has a user-friendly built-in interface for sending and receiving text messages. It might require more steps and possibly also a login event before being able to send messages with your phone via the internet. Furthermore, the receiver of the message will be alerted directly after having received the message, while he possibly had to connect and login somewhere to read the same message if it was sent via the internet. Second, mobile internet would require an additional subscription, often called "bundle." If the user does not have such a bundle, he has to pay for the use of internet a fixed amount of money for every megabyte, which is relative expensive, so bundle or not, accessing the internet via the GSM network costs money.

As this research already showed, a large number of students (68%) use pre-paid. In those cases it is likely that there is no additional internet bundle which makes it rather expensive to use the internet unless a WiFi hotspot is accessible (for free) like for example in modern schools. This however requires a mobile phone that has a WiFi port.

Among the students who have internet on their phones ( $N=365$ ), 47% has a WiFi port in their phone. The correlation of having a phone with internet and a phone with WiFi port is highly significant,  $r(672) = 0.183$ ,  $p < 0.01$ .

Students who have internet on their phone more often have a phone with a WiFi port than students without internet on their phone.

This supports the assumption that if someone intends to access the internet with his phone on a more regular base, he is supposed that he also thinks about the necessary capabilities of his cell phone, which enables him to use the internet more frequently and besides that take advantage of the fact that public places are more and more equipped with WiFi hotspots which - in most cases - can be accessed for free.

At best, this will cause the interaction that people will buy a phone equipped with a WiFi port because the number of places with access points increases, and the managers of public places will facilitate hotspots because the number of people with a WiFi enabled phones increases over time.



Despite the fact that most school policies state that phone use in general is prohibited in the classrooms during the lectures, mobile phones are used intensively. These policies might influence the answers to the question "how much time on average do you use your mobile phone during a single lecture," especially because a teacher is nearby while answering this question. Nevertheless, 68% of the students use their phone during the lessons and a third of these users (N=433) say that they access the internet while using their phones during a lecture. The amount of internet users will only increase in the near future because of two reasons. First, the number of smartphones increases exponentially the next four years [57] and second, more and more schools will deploy a wireless network which allows students to access the internet for free, using their mobile phones.

The clear difference of use intensity between boys and girls - girls use their phones more intensively than boys do - is also confirmed by teaching- and management staff of the schools that took part in the research. "Girls evaluate every minute that just passed by" is one of the remarkable quotations which has been heard.

Mobile phone use during the lectures already happens intensively, the usage between the lectures and during breaks is even more intensive since 86% of the respondents use their phone during a break or while going from one classroom to another in order to attend the next lecture. At these moments, 38% of the respondents are on internet with their mobile phone.

Further analysis of these figures show that phones are significantly more used during breaks and while changing lecture rooms, than during the lectures. This offers additional opportunities regarding student recognition and truancy detection. If a student uses his mobile phone between every lecture, so when he goes from one lecture to another, a specific pattern in use intensity could be recognized. Based on the schedule of the student, a phone use expectation can be defined. This could be compared with the detected phone use in order to check whether the student is present when he is expected to be present.

#### 4.4.2 Other communication devices

There are two devices namely the computer (one of the workplaces at school) and the iPod, which are used by the students besides their mobile phone to access the internet. Instead of the computer, a laptop could be used too. For the purpose of the truancy detector a laptop and computer are not necessarily distinguished and considered the same category of devices, because the data traffic that both devices generate, and deliver to the truancy detector, is of the same format.

The computer in most cases has a wired connection with the internet; iPods (only the type "*iPod Touch*") have a WiFi port. Interesting fact is that iPods send the device name with the

data traffic during communication via a wireless network. Often this device name contains the name of the owner of the device, for example "*iPod\_van\_Suzanne*" (*English: "this is Suzanne's iPod"*). This supports the recognition of students.

The general use of iPods is low, only 6% of the students use it to communicate with others. Main reason is probably the fact that iPods are relative expensive and since they are mainly used for listening music, there are a lot of MP3 players which form a less expensive alternative. The computer is used as communication device by 41% of the respondents. This might not seem to be a lot but one has to keep in mind that many schools have very strict rules regarding the use of their (public) computers. In most of the cases it is prohibited to visit social network sites and often these sites are blocked so that they cannot be visited at all. Also, in these computer rooms there is a supervisor present who can remotely view the activities on the computers in the room which will also de-motivate students to visit prohibited websites using these computers.

There is however an advantage caused by the previously named regulations, namely that students now are more likely to use their mobile phones for visiting these websites, at best via the wireless network of the school. This has positive influence on the amount of input for the truancy detection system. It might furthermore cause a long-term effect with which students will always use their mobile phones via the wireless network to visit social network sites. Students will be more and more familiar with the mobile application and in the end prefer it over the computer application.

#### **4.4.3 Wireless networks at school**

It is by far not always clear to students that there is a wireless network in the school available for use. This follows from the fact that on average 37% of the students answer "I do not know" to the question "Is there a wireless network available at school?" This tends to disinterest but this is not true. In fact it is the other way round. The figures about popularity of a wireless network in the school are positive in the context of automated truancy detection; 79% of the students would definitely use the wireless network if it was available and accessible for free. Another 16% says they would probably use the wireless network. This is important because the truancy detection system relies (at least partly) on the data traffic going through this wireless network.

If such a large amount of students (79-95%) would use this network then the coverage of the truancy detection system would be very high, expressed in the number of students and with respect to the total number of students.

#### 4.4.4 Social networks

Chatting and updating social networks are the two main purposes for secondary school students when using the internet. These activities are closely related since both activities basically comprise the communication with others and sharing thoughts.

The research also confirms the popularity of social network sites with the high number of positive answers to the questions "do you have a *Hyves* account" (93%) and "do you have an account at other social networks" (68%).

The high percentages of students who have an account with a social network site are useful for a truancy detection system based on data traffic because research on data traffic showed that it is relatively easy to recognize people based on data traffic coming from social network sites. This means that a large number of students can be detected in a way that requires relatively low effort.

#### 4.4.5 Attending lectures

As can be seen in Section 4.3.2, girls use their cell phone more intensively during the lectures than boys do, 77% against 58% respectively. On the other hand, boys play truant more than girls (Section 4.3.2); 19% against 12% respectively indicate they play truant nowadays. These figures tend to show that the largest target group, the truancy playing boys, generates the least input, namely data traffic, for the truancy detection system. Nonetheless, this research shows that 84% of the group of students who play truant nowadays, use their mobile phone during the lecture and even 90% of these truants use their phone between the lectures and during breaks.

Phone use during the lectures is correlated with playing truant with a Pearson correlation coefficient of  $r(682)=0.137$ ,  $p < 0.01$ . This means that although not very strong, there exists a relation between phone usage during lectures and playing truant. The correlation shows that students who play truant tend to use their phone in the classroom.

The overall opinion of students about strictness of truancy control tends more to "very strict" than to "not strict." Focusing only on the group of students who play truant nowadays, a slight change in the distribution between "not strict" and "quite strict" is observed. This tends to indicate that the group of students that plays truant nowadays experiences the control less strict than an average student. This is not very surprising because it is expected that if they would have experienced the control as very strict, they would not play truant, or at least less. Nevertheless, students who play truant do not significantly differ from average students on this point. The corresponding p-value is 0.376.

## 4.5 Conclusions

The mobile phone has the potential to become the token which will be used within an automated truancy detection system in secondary education. This has several reasons, which follow from the analysis of the survey:

- 1.) Almost all students in secondary education (99%) have a mobile phone.
- 2.) Students carry their mobile phone with them all day.
- 3.) Students use their phones intensively during the day at school, amongst others for internet.
- 4.) Online social networks are very popular among students.

Research in the field of data traffic furthermore shows two additional reasons:

- 5.) Data traffic from social network sites supports easy recognition of students.
- 6.) The amount of data traffic coming from mobile phones will increase in the next years, together with the exponential increase of the number of smartphones [57].

A relation exists between playing truant and using your mobile phone during the lectures. This research shows that, on average, students who play truant every now and then, use their phones more frequently than students who do not play truant. Analysis shows a significant correlation between truants and how students use their mobile phone during the lectures. It has also been shown that among the students who play truant nowadays, 84% uses his mobile phone during the lectures. If it becomes possible to detect this large percentages of students, this will be very satisfying compared to the registration percentages of traditional absence registration systems [30].

In general, the perspectives for the future are positive for a truancy detection system which is based on data traffic. The amount of data traffic will only increase in the near future, based on the increasing number of online services. Besides, the ability for students to be online whenever and wherever they want, will extend based on the exponential increase of smartphones and their rapid integration into everyday life.

## Literature

1. Marja van Bijsterveldt-Vliegenthart Dutch Secretary of State of Education, *Approach of the problem of leaving school early*. 2010. p. 1-2.
2. Steketee, M.J., et al., *Kinderen in Tel - Databoek 2010*, in *Kinderrechten als basis voor lokaal jeugdbeleid*. 2010, Verwey-Jonker Instituut: Utrecht.
3. Sheehy, A., *Worldwide Smartphone Markets: 2011 to 2015*. Generator Research, May 24, 2011: p. 280.
4. Nedap Education Groenlo. *Examples - ROC van Amsterdam*. [cited 2010 October 5]; Available from: <http://www.nedap-education.com/en/examples.php>.

## Chapter 5 Prototype

The number of students at an average secondary school in The Netherlands easily reaches 1,000. More than a third of these students use their mobile phone for internet purposes. Assume that all these students would use the wireless network of the school. Add all educational related computer use during the lectures and the traffic from school's employees. This results in Gigabytes of traffic every day and this amount of data will only increase in the next few years since the number of digital educational textbooks is increasing exponentially. Additionally, there are already schools which are doing experiments with full digital education with which each student has his own personal laptop to work with, instead of paper textbooks. With the amount of data available today, it is already impossible to manually inspect these Gigabytes of traffic, but what to search for if this traffic inspection should be automated?

In the chapter containing the statistical analysis, valuable research directions were found to support the handling of the huge amounts of data traffic that traverse through a school building every day. From the questionnaires it has been discovered that online social networks like for example *Hyves*, *Facebook* and *Twitter* are very popular among secondary school youth. This not only follows from the large number of accounts at social networks but also from answers to questions like "what is your favorite website" and "what is your main purpose for using the internet." With this in mind, a prototype has been developed using the latest version of JAVA™. This chapter explains the working of this prototype in detail.

### 5.1 Architecture

The working of the application follows a general process which is shown in Figure 6.

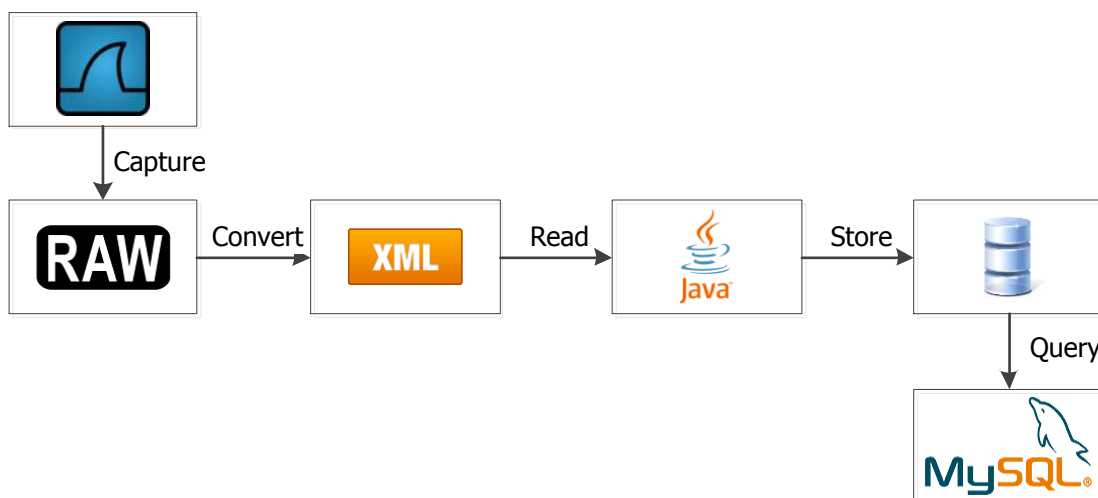


Figure 6: General process flow

Wireshark is used to capture network traffic directly from a network interface card (NIC). Wireshark stores all captured data traffic in "*libpcap*" format. This is represented by the raw file icon. These raw data files are converted to so-called PDML files which are, according to Wireshark, "XML Packet Detail" files. These PDML files are in fact just ordinary XML files with a pdml tag in it. They can be handled in the same way as normal XML files.

These PDML/XML files are read by the JAVA™ application, which uses handlers to filter the useful personal information from the students out of the entire bunch of data traffic. These handlers will be discussed in more detail later in this chapter.

Upon filtering personal data it is directly stored into a MySQL database using JDBC. From this moment on, data which can be used for presence registration is distilled from the large amount of available data traffic and is stored in a database which can be easily queried to get the desired information about a person or a specific day or time. See also section 5.4 for further elaboration on data storage.

The system is entirely based on the local data traffic flowing through the infrastructure which is monitored by the system. This is due to the fact that already at the start of the project, the decision has been taken not to use (dedicated) client software in the truancy detection system. This decision causes that there is no additional cooperation from the targets (the students) necessary and on the same time it limits the influence of the targets on the system. In fact, students should not notice the difference between schools with and without this software monitoring their network. The Prototype chapter further elaborates on this.

## 5.2 User interface

For the ease of use, a graphical user interface has been built; the main screen is shown in Figure 7. From this figure the core functions of the application immediately become clear. In the center of the main screen there are four buttons. It is possible to inspect one single file by clicking the "Inspect file" -button.

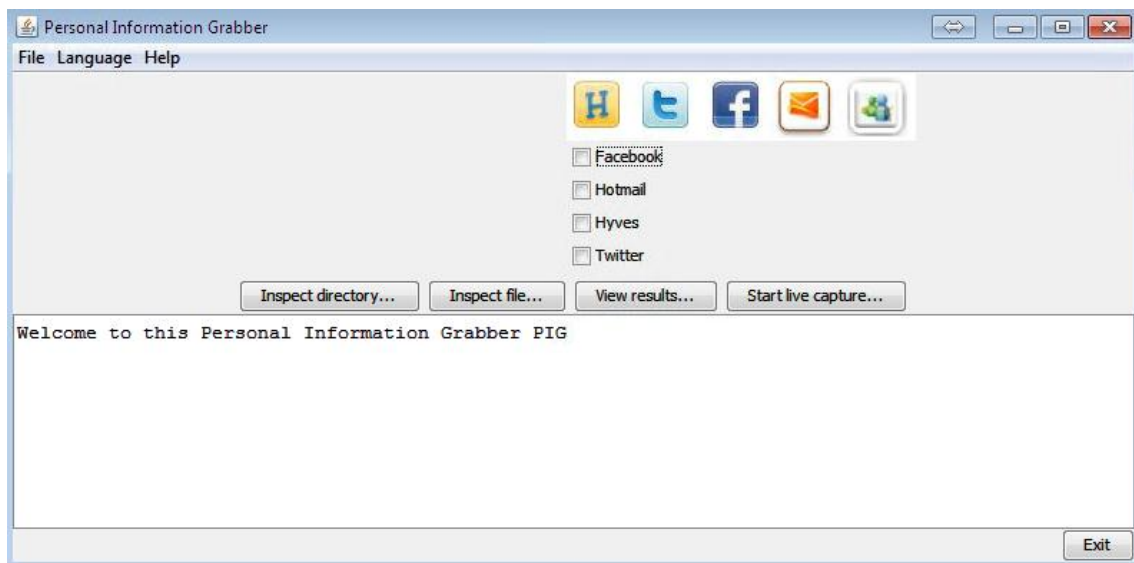
However, if a file becomes larger, the application and the Java Virtual Machine (JVM) require a lot of resources. To tackle this and let the application run more smoothly, file processing should happen using (reading) smaller files; smaller batches of input. These batches are put in a designated folder in the file system. The files in this folder can be read and processed consecutively by using the "inspect directory" functionality.

The current results that have been stored so far in the database can be queried with the button "View results." A small dialog appears in which a start date and an end date can be specified. After clicking the button "Apply dates" a table will be shown with all days between and including the selected start and end date. The table then shows all persons in the database with which the program is currently connected, and shows for every person whether it was

present on one or more of the selected dates, or that there is no information about him in the database. In the latter case "N/A" will be displayed in the table, if the person was however present, "Present" will be displayed in the table.

Finally, the button "Start live capture" can be clicked in order to start a live capturing process immediately after choosing the capturing network interface in a dialog that appears.

For each of the functionalities except the "view results" -functionality holds that one or more handlers need to be selected. This is done by checking the desired checkboxes in order to filter events of the corresponding services.



**Figure 7: Graphical User Interface of the prototype (main screen)**



## 5.3 Handlers

To facilitate easy extension of the prototype in the future, inheritance has been applied in the design of the prototype. An interface has been defined which is called "Handler." This interface specifies all methods that should be implemented by a handler class which handles a specific type of data like for example *Facebook*. Whether these methods will get a body implementation in the handler classes and what these bodies would look like depends on how data is received from a web site for which a handler is built.

In the future, additional handlers can be build and work parallel to the currently running handlers. The new handler class should implement the Handler interface and inherit all methods in this interface. The methods which are relevant for the service for which a handler is built, should be implemented. Basically the methods search in String values for keywords in the data from the service. If there is for example an e-mail address found, the specific method should return this e-mail address as a String. Finally the building process of the handler array in the class "HandlerUI" has to be adjusted in such a way that the new handler will be recognized by the engine. The handler needs to be registered with the engine. From this moment on, the engine is able to recognize events from a new service.

Currently there are four handlers available. One for *Hyves*, one for *Facebook*, one for *Twitter* and one for *Hotmail*. All handlers search in the data they receive for specific keywords or tags which are unique for the different services. An example from *Hyves* is "*var global\_member\_username*" which contains the account name of a *Hyves* user. Detailed descriptions of the exact working of all different available handlers can be found in Appendix C.

## 5.4 Data storage

The data filtered by the prototype is stored in a MySQL database. The schema which gives an overview of how the database is organized is depicted in Figure 8. All tables have a first column called "id" which automatically increments in case a row is added to the table. This is for quick reference and supports anonymous reporting about table entries, which is necessary because the tables contain personal data of which publication is prohibited without explicit permission of the involved person.

For each of the different handlers a table exists. Since the handlers have to implement the already mentioned interface "Handler," the tables that store the accounts of the different services look similar.

An account can have a maximum of four properties: an account name, a user name which is the real name of the owner of the account, a user id which is a unique number at the level of the specific service to identify and refer to a user, and the e-mail address of a user.

An account table is different if a service does not use or support one of these properties. In the table with *Facebook* accounts there is for example no column for account names because Facebook does not have account names like *Hyves* and *Twitter*. The latter uses screen names as account names for users. *Hotmail* neither uses account names nor uses unique user identification numbers. Just only the name of the user and its e-mail address.

Besides the account tables, there exists a table which keeps track of all persons recognized in the past and another table which acts as a list containing all events that occurred. This table is therefore called "eventList."

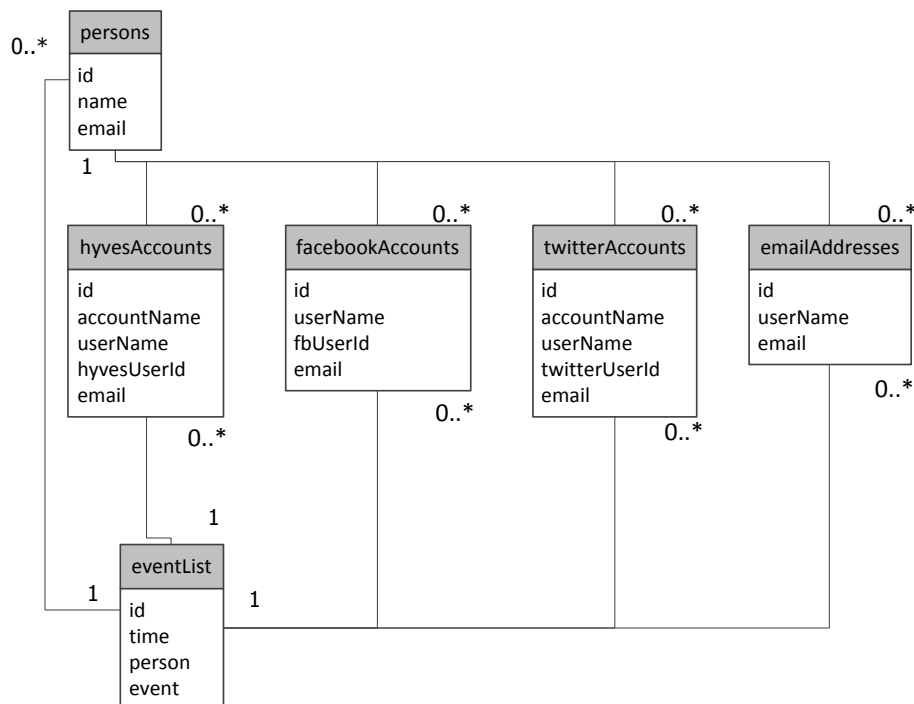


Figure 8: Database schema

Considering the database schema in Figure 8, the question might rise why the hardware (MAC) address of the source device from which the data traffic originates is not stored.

A communication device can be identified according to its MAC address which is supposed to be unique for every device. However, this is not entirely true because a MAC address can be cloned or spoofed, which allows for masquerading, i.e. pretending to be someone else who is generating messages.

Spoofing and cloning are official security vulnerabilities but are by far not the worst problem. The act of masquerading will probably happen a lot nowadays, although it will happen more or less accidentally. An example is a laptop owned by the school and which is used by another student every consecutive lecture. In the data traffic the same MAC address of this laptop is visible but the user differs every next hour. This causes the MAC address of a device to be unsuitable to apply for explicit authentication of users. For this reason, the current prototype does not register MAC addresses. There are however other features with which the MAC address could be an interesting property to keep track of. Section 6.5 elaborates on this.

## Chapter 6 Network traffic analysis

For practical research purposes and to check the working of the prototype, we gained permission at a secondary school in The Netherlands to monitor the entire wireless network during one week in April 2011. Findings during this experiment are reported in this chapter.

### 6.1 Background

The main goal of this Masters project is to investigate the possibilities of automated truancy detection in a school. Absence of students cannot be registered directly based on data traffic events, but presence can. Based on this presence registration in combination with a presence expectation, usually a roster, the absence of students can be investigated. To prove the concept of presence registration based on local data traffic, the in Chapter 4 proposed prototype is used in a real world scenario.

### 6.2 Test environment

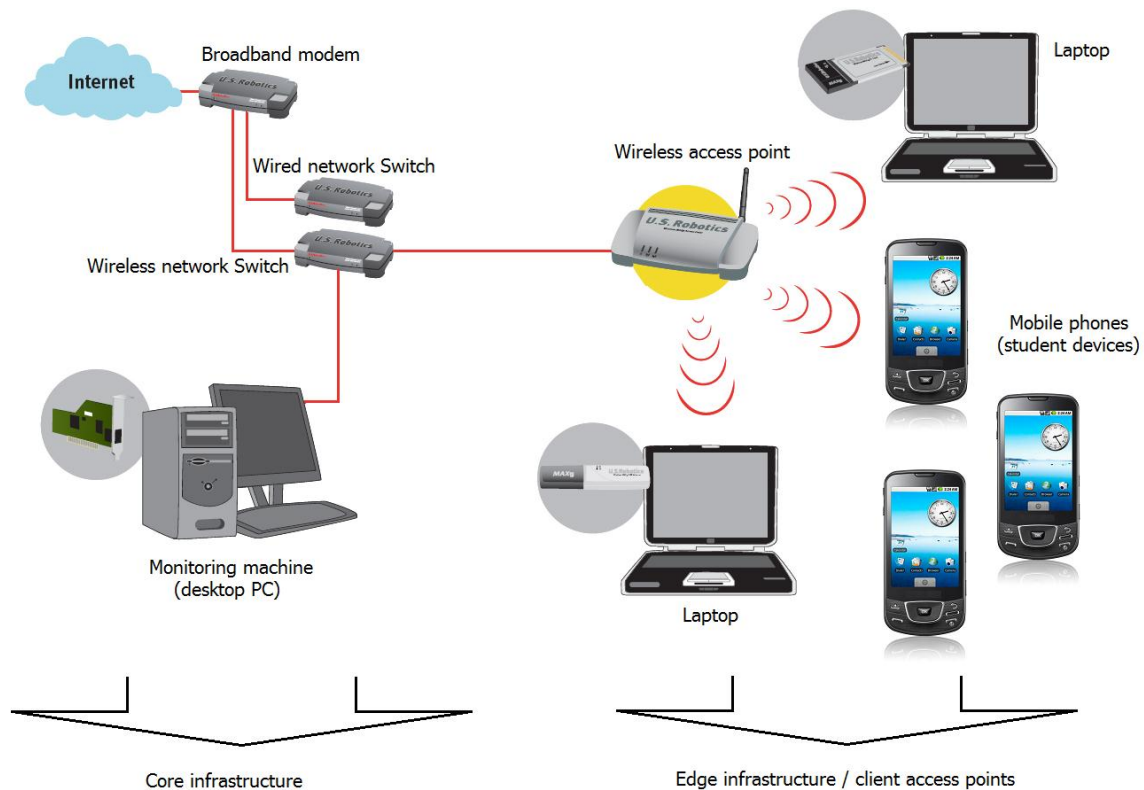
The prototype monitored the entire wireless LAN of a secondary school from April 5 until April 14, 2011. The school has approximately 1255 students and educates in all three levels of Dutch secondary education. The lectures are given following the time schedule below on the left, except for Tuesdays. On Tuesdays, the right schedule is followed and school is out at last at 14.05 hr. Afternoons on Tuesdays are used by the staff to perform non-educational but education related tasks.

Table 19: Lecture times at the test school

1.: 8.30 – 9.30	Les 1: 8.30 – 9.30
2.: 9.30 – 10.30	Les 2: 9.30 – 10.30
<b>break</b>	<b>break</b>
3.: 10.50 – 11.50	Les 3: 10.45 - 11.45
4.: 11.50 – 12.50	Les 4: 11.45 - 12.45
<b>break</b>	<b>break</b>
5.: 13.20 – 14.20	Les 5: 13.05 - 14.05
6.: 14.20 – 15.20	(Les 6: 14.05 - 15.05)
7.: 15.20 – 16.20	(Les 7: 15.05 - 16.05)

The prototype was running on a separate desktop system and was connected to the infrastructure of the school using an Ethernet connection. An overview of the system is depicted in Figure 9. The left part of the picture shows the equipment in the ICT room of the school where the core of the infrastructure is located. There is amongst others a main switch for the wired network and a separate main switch for the wireless network. The monitoring device was

connected to the main switch of the wireless network, via a so-called mirror port. All data traffic that goes through the main switch is also directed to this mirror port, so this mirror port received all data of all communication sessions on the wireless network. The right part of Figure 9 represents the environment in which the students use their mobile phones. It consists of a wireless access point which is connected to the main switch of the wireless network. Students connect to the wireless access point through the IEEE 802.11a/b/g/n protocol.



**Figure 9: Infrastructure test environment [data traffic analysis secondary school]**

The prototype makes use of WireShark to capture data traffic. Since a mirror port was used, all traffic that we are interested in is already routed to the monitoring machine. This makes configuration a lot easier because we do not need to modify settings in the network interface of the monitoring machine, and WireShark does not need to be necessarily in promiscuous mode. During the process of capturing data traffic, no filter was used. To ensure that no important parts of data would be ignored, the choice was made to capture and save everything that passed by. The equipment has enough performance to capture and store everything, so the only issue would be storage capacity. It turned out however that the amount of captured data is in terms of Gigabytes and since there are already hard disks which are not expensive and with the capacity of several Terabytes, there was no issue left regarding data capturing without any filters.

## 6.3 Results

The next three tables show the figures from the data capturing experiment at a secondary school. Table 20 contains the dates and times at which data has been captured, together with the amount of data in Gigabytes per day, the number of files and the number of recognized events at each day.

**Table 20: Data capture dates**

Day	Date	Capture Start	Capture End	Gigabytes	Files	Events
Tuesday	April 5	12:40 hr	16:25 hr	1.17	2399	283
Wednesday	April 6	07:45 hr	15:00 hr	4.27	8744	1132
Thursday	April 7	-	-	0	0	0
Friday	April 8	07:47 hr	16:25 hr	3.76	7708	721
Monday	April 11	09:15 hr	16:25 hr	0.068	140	9
Tuesday	April 12	08:25 hr	16:24 hr	0.078	159	3
Wednesday	April 13	07:45 hr	16:28 hr	0.091	187	1
Thursday	April 14	07:45 hr	13:12 hr	1.98	4070	224
8 days		48 Hours, 57 Minutes		11.42 Gigabytes	23407 files	2373 events

Table 21 shows the number of events per service that has been monitored, including the number of occurrences for each column in the account tables in the database.

**Table 21: Events per service**

	Hyves	Facebook	Twitter	Windows Live
<b>Total events</b>	681	1278	329	80
<b>Total accounts</b>	107	52	44	50
<b>Unique user ID's</b>	77	46	0	-
<b>Unique account names</b>	83	-	44	-
<b>Unique person names</b>	82	2	0	50
<b>Unique e-mail addresses</b>	29	31	0	50

Table 22 shows relations between two individual accounts. A relation means that an e-mail address in the first table also occurs in the second table.

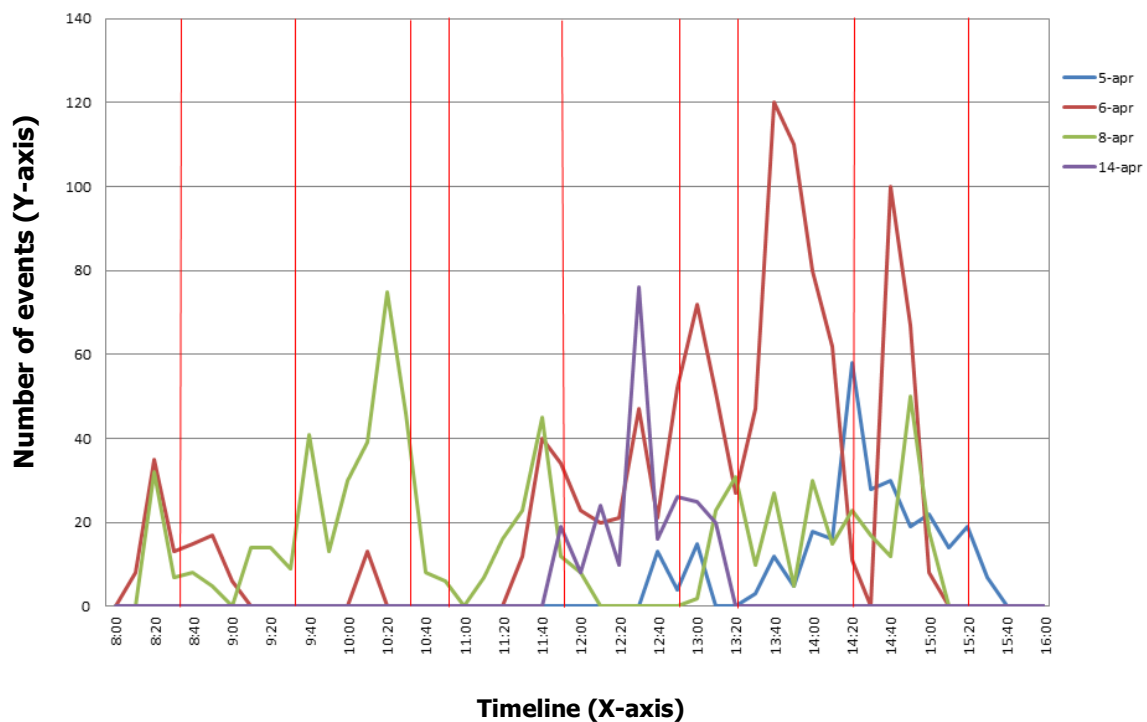
**Table 22: Relations between services**

	Hyves	Facebook	Twitter	Windows Live
<b>Relations with Hyves</b>		2	0	30
<b>Relations with Facebook</b>	2		0	6
<b>Relations with Twitter</b>	0	0		0
<b>Relations with Windows Live</b>	30	6	0	

From the *Hyves* accounts table it has been discovered that there are two *Hyves* accounts which belong to one single person. This conclusion could be made because these accounts are both linked to the same e-mail address.

The information in Table 20 and Table 21 has been combined and analyzed in more detail. This results in the graph depicted in Timeline (X-axis)

Figure 10, which shows the number of events over time. It contains the number of events recognized during four different days as can be seen in the legend on the right of the graph. The vertical red lines are starting times of lectures or breaks, see also Table 19.



**Figure 10: Data traffic intensity during 4 days**

The first lecture of the day starts at 8.30 am. At April 6 and April 8, events have been recognized between 8.00 am and 8.30 am. The peak at approximately 8.20 is interesting. This indicates that students use their mobile phones intensively prior to the first lecture because there is some spare time and/or they do not have anything else to do. Then at 8.30 the peak drops drastically because the lessons start and the number of events decreases further during the first lesson.

Around 9.30 the number of events increases. At this time the first lecture ends and students go from one classroom to another. At 9.40 when the second lecture finally started, the number of events decreases again. This is however does not last long because while the lecture is going on the number of events increases again. Then at 10.30 when the first break starts, the number of events is relative low and decreasing. This might be caused by the fact that students of different classes are able to meet face-to-face so they do not need to communicate by using their mobile phones. Furthermore, they are now able to chat and eat and drink, and they might save the time for surfing on the internet until a less interesting (part) of a lecture starts.

During the third and fourth lecture again a lot of events occur with again peaks in the second half of both of the lectures. It is expected that this trend is inversely proportional with the motivation of the students, at the end of the lecture students are expected to pay less attention and are more likely to grab their mobile phone.

In the first half of the lunch break which starts at 12.50 (Tuesdays (blue line) at 12.45) there occurs a peak of events at each of the individual days which, except for April 8, decreases when the break comes to an end.

Although with different amplitudes (intensity), during the fifth lecture similar patterns are recognized at April 5 and April 8. This is interesting but since these are only two days, decent considerations cannot be stated.

Broader perspectives that cover an entire day show that in general there occur more events during the second half of the day. This might be caused by the fact that during the day more things happen and students experience more, so there is more to discuss among the students. Furthermore it is expected that students have more motivation to attend lectures in the morning than in the afternoon, so to prevent from getting bored, they use their mobile phones more in the afternoon.

Previously stated facts indicate that most peaks in the graph occur around starts and ends of lectures. During these moments students have relative more opportunities to use their mobile phone. They are packing and unpacking their bags, there is less (strict) supervision on what students exactly do if the lecture has not officially started. This and the previous



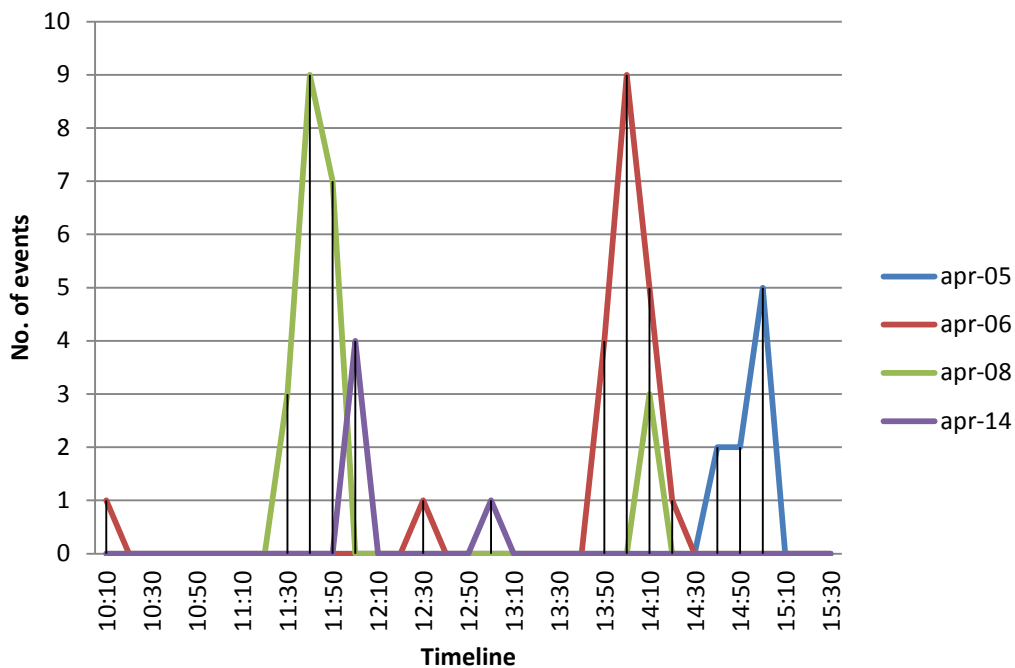
paragraph however cause future work to conduct research in order to verify these findings. (See also Section 7.2.1).

During the analysis of the captured data traffic, there were several assumptions that data traffic belonging to different services could be related to one single person. The assumptions in these cases are based on the fact that the events occurred within one minute and person names and/or account names had a close match.

An example is the Twitter event of the user with id=23 in the Twitter accounts table. Within half a minute, at April 6 at 12:55hr, a Twitter event was generated, followed by a Hyves event. The person name at Hyves has a very close match with the account name at Twitter. The same happens again at April 8 at 14:00hr. So based on the occurrence within such a small time frame and the close match of the names, the assumption was made that the Twitter account and the Hyves account both belong to the same person.

The prototype however does not keep track of the source device. If for example this data was sent from the same device within a small time frame, it is more likely that the same person generated the different data. If however the source devices differ, it is less likely that the same person generated these different events.

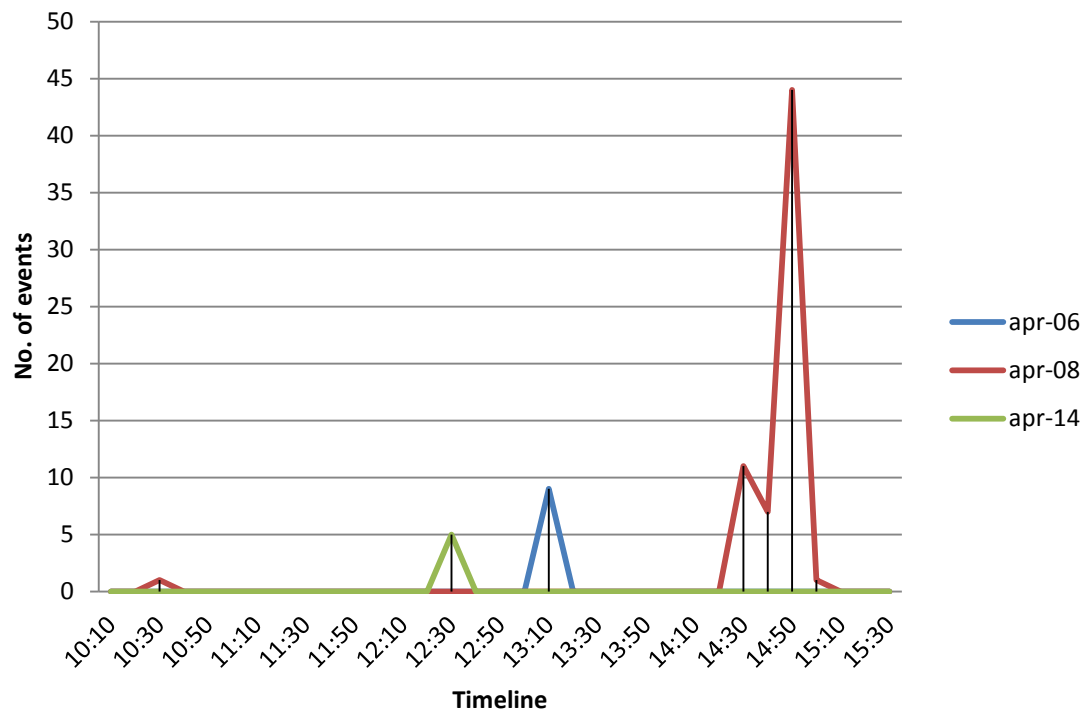
The student with the *Twitter* account at id=5 in the *Twitter* accounts table in the database, used *Twitter* the most intensive of all *Twitter* users. There were 57 *Twitter* events (not to be confused with tweets) from this account recognized within 4 days. Out of 44 *Twitter* accounts this is the only one which was recognized at each of the four selected days. These days were selected because by far the most events (in general) were recognized during these days. The number of events of this particular *Twitter* user is plotted against the time and shown in Figure 11.



**Figure 11: Timeline of events of most intensive Twitter user**

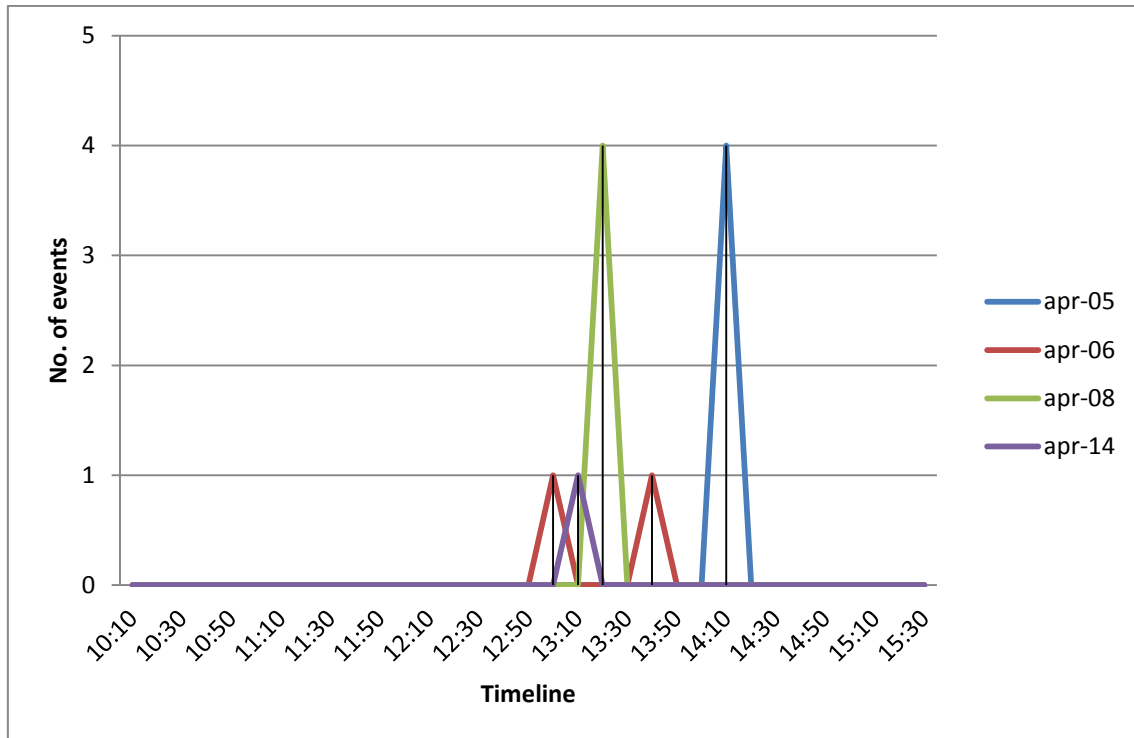
As can be seen in Figure 11, which contains the *Twitter* events of 4 days from one student, there is not directly a pattern visible that is similar and is repeated at each of the individual days.

The graph in Figure 12 shows the events of the most intensive *Hyves* user, who generated 78 *Hyves* events during 3 days. It turns out however, that there occurred a peak at one day namely at April 8. The other, significantly smaller, peaks both occur at different times. So again no daily pattern could be recognized. Further research showed that the second most intensive and third most intensive *Hyves* users generated relatively many events during only one single day of the research period.



**Figure 12: Timeline of events of most intensive Hyves user**

For the services of Windows Live Mail and Messenger, the most intensive user has been analyzed. The three most intensive users only generated 8 events each. One of them, with id=5 in the email addresses table, seems to check his e-mail every day during or after the lunch break. Despite the small number of events, they occur within the same time span of one and a half hour every individual day. This is visualized in Figure 13.



**Figure 13: Timeline of events of most intensive Windows Live user**

Finally, the intensity and periodicity of Facebook users has been analyzed. The student who used Facebook most heavily, generated 457 events during two days. The second and third most intensive Facebook users generated approximately the same number of events during two days as well. Because the events are spread out over only two individual days, further analysis is not relevant since any pattern recognition of daily or weekly presence is impossible over such a short time span of only two days. This holds that data traffic needs to be monitored during a longer period in order to perform analysis and conclude about patterns.

Compared to computers and laptops, mobile phones that access the internet receive less user specific data. For example during a *Hyves* login event on a traditional computer or laptop, the full name of the user, the account name of the user and the unique Hyves-user-ID can be rendered from the data traffic. This in contrast to a Hyves login event via a mobile phone, with which most of the time only a service-specific (unique) user ID is sent. This supports unique user recognition but makes it harder to map a network user to a student since there is no real/full name available.

A solution to the problem above is to combine the results collected from the login events of the different services. This is for example possible if at two different login events of different services, exactly the same e-mail address was recognized. This e-mail address is the relation between the two strictly different events. This relation can be guaranteed since e-mail addresses are unique and personal and belong to exactly one person.

## 6.4 Conclusions

The occurrences of interesting events in data traffic follow similar patterns at different days. This can be concluded from the graph depicted in Figure 10. This graph also shows that mobile phones are used more intensively by students at the end of individual lectures. Furthermore, it has been discovered that mobile phones are used more intensively in the afternoons. All previously stated conclusions are important and support the making of expectations regarding moments when students are expected to be online. This in turn supports the recognition of absence.

The best truancy detection system supports all different (mobile) platforms available at the moment of detection. With Hyves for example, this holds that the non-mobile web login, the mobile web login, the Hyves application for Android, the iPhone app, the BlackBerry app, and the Nokia application (by *Ovi for Nokia*) should be supported. And this is only for one single service yet. So if this has to be done for all different services, the two-dimensional matrix setting out the available platforms against the services to be supported quickly becomes large. To be short: for a decent truancy detection system, a lot of small parts of dedicated software needs to be written.

Despite possible assumptions, data traffic from different services (different social networks) can only be related to one another in the case there is a match on the e-mail address.

Since the database structure of the current prototype keeps track of four different properties of each single account, namely the account name, person name, user id and e-mail address, only the e-mail address can be guaranteed to be globally unique. The user id is unique too, but only within the specific service to refer to a user. Outside the scope of the service in which this user id is used, this id is meaningless.

There is however one drawback concerning this relation method, namely the fact that persons could use different e-mail addresses at different services. In that case there are no unique personal identifiers to relate, so the match between two profiles of the same person cannot be made.

## 6.5 Recommendations

Based on the design choices made during the development of the current prototype, there is only one possibility to compare events from different tables. Comparison can only take place based on e-mail addresses recognized during login events, either because this e-mail address is the login name, or because the e-mail address of the user was sent back by the server of the specific service at which login took place.

Future extensions of the current presence detection software could contain a feature that keeps track of sessions. This feature registers two or more consecutive login events at different services but with the same source MAC address. Since a MAC address is unique for a device, the assumption can be made that these consecutive login events belong to the same person and so the profiles at the online services also belong to the same person. Based on this recognized relationship, missing information from one account entry in the database can be completed with information received from another account entry.

## Chapter 7 Conclusions and future work

In this chapter, the research questions which were stated at the start of the research project are answered based on the research results collected during two types of research: the figures from the questionnaires (1) and the data from the practical experiment of network traffic monitoring (2).

Furthermore there are recommendations for future work proposed in order to continue the development.

### 7.1 Answers to research questions

#### 7.1.1 Sub research question 1

*"How do students use modern media at school?"*

The mobile phone has the potential to become *the* client device, in the context of tools, to detect presence of students at school. Success factors of the previously stated conclusion are derived from the figures gathered from the questionnaires filled in by 701 secondary school students, and from classroom observations during visits at Dutch secondary schools. Amongst these success factors are the following important figures:

- 1.) Almost all students in secondary education (99%) have a mobile phone.
- 2.) Students carry their mobile phone with them all day.
- 3.) Students use their phones intensively during the day at school (68% during the lectures), amongst others for internet purposes.
- 4.) Online social networks are very popular among students. 96% has an account with an online social network site.

A relation exists between playing truant and using your mobile phone during the lectures. This research shows that, on average, students who play truant every now and then, use their phones more frequently than students who do not play truant. Analysis shows a significant correlation between truants and how students use their mobile phone during the lectures. It has also been shown that among the students who play truant today, 84% uses his mobile phone during the lectures. If it becomes possible to detect this large percentages of students, this will be very satisfying compared to the registration percentages of traditional absence registration systems [30].

### 7.1.2 Sub research question 2

*"How can this behavior of the students be registered automatically?"*

Since online social networks are very popular among secondary school youth, data traffic from these services has been analyzed in detail. Research by means of a practical experiment with which the developed software tool has been used showed that it is relatively easy to recognize the users of the network, based on the data traffic that passes by. Especially data traffic from and to the servers of social network sites like for example *Hyves* or *Facebook* turns out to be very suitable for recognizing students. Besides that, since these responses are the result of login events, a form of authentication is built in. In line with the previously mentioned statements, the practical experiment with the developed software tool showed that it is possible to detect the presence of students based on their online behavior.

### 7.1.3 Main research question

*"How to automate the process of truancy registration?"*

In general, the perspectives for the future are positive for a truancy detection system which is based on data traffic. The answer to the first sub research question shows potential for an automated truancy detector.

Putting all previously stated conclusions together, it can be said that the in this thesis proposed truancy detection system has chances to become successful in the future. The amount of available input for the system will increase due to:

1. The increasing number of online services
2. The integration and acceptance of mobile phones in school environments
3. The number and performance of available wireless networks that will increase
4. The number of smartphones that will increase exponentially.

So, more availability of the networks, easier accessibility of the networks and more and better services over these networks. All three points plus the fact that wireless networks in the schools can be used without extra costs for the students, will trigger an increasing number of students to use this network and use it more and more. Finally, the truancy detection system will not suffer from shortage of valuable input.



The coverage, the number of students who already are a target for the system, which is currently about 33% is reasonable yet and will increase in the next few years as already mentioned at the beginning of this section.

However, the results so far only proofed presence registration. To go from presence registration to absence registration, additional steps need to be taken and additional research has to be conducted. The next chapter will elaborate on this with some future work proposals.

## 7.2 Future work

The proposals for future work are divided into two categories. First, two research proposals are done. Next, the during this Masters project developed prototype is considered and three out of many improvement possibilities are further described.

### 7.2.1 Research

In order to commercially publish the automated presence registration system, research needs to be conducted in at least two areas to verify assumptions which have been made during the initial design and development phase which resulted in the current version of the prototype.

#### *Authentication method*

The current authentication method is based on the fact that users of social network sites only know their own login credentials. Or the other way round; login credentials of an account are only known by the owner of this account. Then a login event with an account has been recognized in the data traffic, it is assumed that the person who corresponds to this profile, is the person who is recognized and is physically present.

Since this is an assumption it is interesting to perform research to investigate the usage and distribution of credentials and whether or not they are shared with for example family or friends.

#### *Peaks in data traffic*

Section 6.3 describes the discovery of peaks in data traffic around the start and end times of the lectures. The assumption is made that the students are less motivated to attend the lectures around these moments or there is less supervision so better opportunity for students to use their mobile phone, or a combination of both. Since these are assumptions, they need to be verified by means of research in order to conclude about the usefulness of these peaks in the data traffic, with respect to pattern recognition.

### *Acceptance*

The prototype uses personal data with more or less private information as its input. Despite the fact that the impact on privacy of the students is low, not everybody might be able to correctly evaluate the impact on personal privacy. For that reason, research has to be conducted to investigate how an automated truancy detection system as proposed in this thesis, will be received by the people, especially the target group namely the students.

### *Absence recognition*

The main research question of this project is how to automatically register truancy, so how to register absence. It turned out that this is an ambitious goal. This research proofed automated presence registration, absence registration is however at least one step beyond. Research needs to be conducted in order to investigate how the current results gathered with presence registration can be used for and transformed into absence registration.

## **7.2.2 Prototype**

For the prototype there are a lot of extensions possible. Three important ones are described below. The others are stated in Chapter 7 of Appendix C.

### *Handlers*

In order to be able to receive even more input than currently possible, additional handlers can be developed for any available service. The next service to propose a handler for is Google. Google offers a lot of different services which can be used with a single Google account. Being able to recognize login events of Google accounts is interesting in at least two ways. First, as already mentioned, this Google account can be used with a reasonable number of different services, so the usage intensity is expected to be relatively high. Second, the fact that this account can be used at different services supports the recognition of relations between login sessions of the same user at different services. This relation principle is one of the key aspects that make this prototype robust and self-learning.

### *Connection with school database*

The prototype recognizes people, but it also needs to know who it has to recognize. During the practical experiment performed in the context of this project for example a teacher of the school has been recognized in the data traffic. This is nice but this is not the goal of the system. Besides the knowledge about who needs to be there, the truancy detector also has to know when the target persons need to be present. This information is available from the

administration system of the school. One of the next steps to take with the prototype is facilitating a connection with this system or at least the database, so presence expectations can be made which need to be compared with the results of the data traffic monitoring process. The difference in this comparison shows the possible cases of truancy.

#### *Show presence per student per hour*

The current prototype shows the presence of all recognized students on a daily base. If a student has been recognized he is considered present the entire day. This has been done to show the proof of the concept in a clearly visible and understandable manner.

Schools however register the presence of their students per hour. The events recognized by the prototype are stored in the database together with a timestamp which is exact up to a second. So the storage of the data already happens precisely enough, the presentation of the data requires some improvements.

## Literature

1. Bland, N. and T. Read, *Policing Anti-Social Behaviour*, in *Police Research Series*. 2000: London, UK.
2. Alderson, P., *School students' views on school councils and daily life at school*. *Children & Society*, 2000. **14**(2): p. 121-134.
3. Vazsonyi, A.T., et al., *Routine activities and deviant behaviors: American, Dutch, Hungarian, and Swiss youth*. *Journal of Quantitative Criminology*, 2002. **18**(4): p. 397-422.
4. Weinstein, S.M. and R. Mermelstein, *Relations Between Daily Activities and Adolescent Mood: The Role of Autonomy*. *Journal of Clinical Child & Adolescent Psychology*, 2007. **36**(2): p. 182 - 194.
5. Schmiedl, G., T. Grechenig, and B. Schmiedl. *Mobile enabling of virtual teams in school: An observational study on smart phone application in secondary education*. in *Education Technology and Computer (ICETC), 2010 2nd International Conference on*. 2010.
6. Marcum, C.D., G.E. Higgins, and M.L. Ricketts, *Potential Factors of Online Victimization of Youth: An Examination of Adolescent Online Behaviors Utilizing Routine Activity Theory*. *Deviant Behavior*, 2010. **31**(5): p. 381 - 410.
7. Scornavacca, E., S. Huff, and S. Marshall, *Mobile phones in the classroom: if you can't beat them, join them*. *Commun. ACM*, 2009. **52**(4): p. 142-146.
8. popcenter.org. *Crime Scripts*. 2010 [cited 2010 November 10]; Available from: <http://www.popcenter.org/learning/60steps/index.cfm?stepNum=35>.
9. Baker, M.L., J.N. Sigmon, and M.E. Nugent, *Truancy Reduction: Keeping Students in School*, in *Juvenile Justice Bulletin*. 2001, Office of Juvenile Justice and Delinquency Prevention: Washington DC, US.
10. Mccray, E.D., *It's 10 a.m.: Do You Know Where Your Children Are?* *Intervention in School and Clinic*, 2006. **42**(1): p. 30-33.
11. Sutphen, R.D., J.P. Ford, and C. Flaherty, *Truancy Interventions: A Review of the Research Literature*. *Research on Social Work Practice*, 2010. **20**(2): p. 161-171.
12. Corville-Smith, J., et al., *Distinguishing Absentee Students from Regular Attenders: The Combined Influence of Personal, Family, and School Factors*. *Journal of Youth and Adolescence*, 1998. **27**(5): p. 629-640.
13. Yahaya, A., et al., *The Effects of Various Modes of Absenteeism Problem in School on the Academic Performance of Students in Secondary Schools*. *European Journal of Social Sciences*, 2010. **12**(4): p. 624-639.
14. Claes, E., M. Hooghe, and T. Reeskens, *Truancy as a contextual and school-related problem: a comparative multilevel analysis of country and school characteristics on civic knowledge among 14 year olds*. *Educational Studies*, 2009. **35**(2): p. 123 - 142.
15. Attwood, G. and P. Croll, *Truancy in secondary school pupils: prevalence, trajectories and pupil perspectives*. *Research Papers in Education*, 2006. **21**(4): p. 467 - 484.
16. Pauchli, W., *Spijbelen in het MBO - Praktijkgericht onderzoek*. 2010, University of Utrecht.
17. Sun, P., et al., *Internet accessibility and usage among urban adolescents in Southern California: implications for web-based health research*, in *Cyberpsychology & Behavior*. 2005. p. 441-453.
18. Bessems, K.M.H.H. and M.T.W. Leurs, *Research Safe School 2003 / Onderzoek Veilige School 2003*. 2004, GGD Zuidelijk Zuid-Limburg: Maastricht.
19. Black, M. and R.G. Smith, *Electronic Monitoring in the Criminal Justice System*, in *Trends and issues in crime and criminal justice*. 2003, Australian Institute of Criminology: Canberra, Australia.
20. Hoshen, J., J. Sennott, and M. Winkler, *Keeping tabs on criminals [electronic monitoring]*. *Spectrum, IEEE*, 1995. **32**(2): p. 26-32.

21. Newman-Ford, L., et al., *A large-scale investigation into the relationship between attendance and attainment: a study using an innovative, electronic attendance monitoring system*. *Studies in Higher Education*, 2008. **33**(6): p. 699 - 717.
22. Doyle, L., et al., *An evaluation of an attendance monitoring system for undergraduate nursing students*. *Nurse Education in Practice*, 2008. **8**(2): p. 129-139.
23. Rukzio, E., et al., *An Experimental Comparison of Physical Mobile Interaction Techniques: Touching, Pointing and Scanning*, in *UbiComp 2006: Ubiquitous Computing*, P. Dourish and A. Friday, Editors. 2006, Springer Berlin / Heidelberg. p. 87-104.
24. Ervasti, M., M. Isomursu, and M. Kinnula. *Experiences from NFC Supported School Attendance Supervision for Children*. in *Mobile Ubiquitous Computing, Systems, Services and Technologies, 2009. UBICOMM '09. Third International Conference on*. 2009.
25. Silva, F., V. Filipe, and A. Pereira. *Automatic Control of Students' Attendance in Classrooms Using RFID*. in *Systems and Networks Communications, 2008. ICSNC '08. 3rd International Conference on*. 2008.
26. Qaiser, A. and S.A. Khan. *Automation of Time and Attendance using RFID Systems*. in *Emerging Technologies, 2006. ICET '06. International Conference on*. 2006.
27. Lim, T.S., S.C. Sim, and M.M. Mansor. *RFID based attendance system*. in *Industrial Electronics & Applications, 2009. ISIEA 2009. IEEE Symposium on*. 2009.
28. Hsu, I.C. *Mobile ubiquitous attendance monitoring system using wireless sensor networks*. in *Education Technology and Computer (ICETC), 2010 2nd International Conference on*. 2010.
29. w3schools.com. *XML tutorial*. 2010 [cited 2010 October 20]; Available from: <http://www.w3schools.com/xml/default.asp>.
30. Nedap Education Groenlo. *Examples - ROC van Amsterdam*. [cited 2010 October 5]; Available from: <http://www.nedap-education.com/en/examples.php>.
31. Nelson, M., T. Wright, and K. Ashida, *Fujitsu's PalmSecure-based e-POS system for school cafeteria*. *Fujitsu Scientific and Technical Journal*, 2007. **43**(2): p. 236-244.
32. Network75. *International Student Attendance Monitoring*. [cited 2010 September 28]; Available from: <http://www.uninanny.co.uk/content/public/main/internationalstudents.aspx>.
33. Stringer, R., *Get 'em young*. *Infosecurity*, 2009. **6**(2): p. 34-37.
34. Paik, M., et al., *A biometric attendance terminal and its application to health programs in India*, in *Proceedings of the 4th ACM Workshop on Networked Systems for Developing Regions*. 2010, ACM: San Francisco, California. p. 1-6.
35. Mohd Said, R., et al., *User Technology Readiness Measurement in Fingerprint Adoption at Higher Education Institution*, in *Computational Science and Its Applications – ICCSA 2008*, O. Gervasi, et al., Editors. 2008, Springer Berlin / Heidelberg. p. 91-104.
36. Leong Lai, F. and S. Woo Chaw. *A Comparison Study on Hand Recognition Approaches*. in *Soft Computing and Pattern Recognition, 2009. SOCPAR '09. International Conference of*. 2009.
37. Chunjiang, L. and L. Fang. *Design of highly reliable Fingerprint Access Control System based on C8051F020 single chip*. in *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on*. 2010.
38. Zhang, Y. and J. Liu. *The Design of Wireless Fingerprint Attendance System*. in *Communication Technology, 2006. ICCT '06. International Conference on*. 2006.
39. Tennina, S., et al., *Locating ZigBee nodes using the TI's cc2431 location engine: a testbed platform and new solutions for positioning estimation of wsns in dynamic indoor environments*, in *Proceedings of the first ACM international workshop on Mobile entity localization and tracking in GPS-less environments*. 2008, ACM: San Francisco, California, USA. p. 37-42.
40. Tennina, S., et al., *Distributed localization, tracking, and automatic personal identification: a solution based on a wireless biometric badge*, in *Proceedings of the 4th ACM international workshop on Experimental evaluation and characterization*. 2009, ACM: Beijing, China. p. 97-98.

41. Yasinsac, A. and Y. Manzano. *Policies to Enhance Computer and Network Forensics*. in *2001 IEEE Workshop on Information Assurance and Security*. 2001. United States Military Academy, West Point, NY.
42. Qadeer, M.A., et al. *Network Traffic Analysis and Intrusion Detection Using Packet Sniffer*. in *Communication Software and Networks, 2010. ICCSN '10. Second International Conference on*. 2010.
43. Balachandran, A., et al., *Characterizing user behavior and network performance in a public wireless LAN*. SIGMETRICS Perform. Eval. Rev., 2002. **30**(1): p. 195-205.
44. Kotz, D. and K. Essien, *Characterizing Usage of a Campus-wide Wireless Network*. 2002, Dartmouth College: Hanover, New Hampshire, USA.
45. Tran, T.M.C., B. Scheuermann, and M. Mauve. *Node presence detection with reduced overhead*. in *Wireless On-Demand Network Systems and Services, 2009. WONS 2009. Sixth International Conference on*. 2009.
46. Duimel, M., *8-18 jarigen en mobiele telefoons*, in *Altijd binnen bereik*. 2010, Stichting mijn kind online: Den Haag.
47. Hynes, M., et al. *Monitoring of activity levels of the elderly in home and community environments using off the shelf cellular handsets*. in *Consumer Electronics (ICCE), 2010 Digest of Technical Papers International Conference on*. 2010.
48. Krolovitsch, A., et al., *STRIDE: A ubiquitous System to improve healthiness during VDT work*. 2010, University of Twente: Enschede.
49. Payne, B.K. and M.T. DeMichele, *Electronic supervision for sex offenders: Implications for work load, supervision goals, versatility, and policymaking*. Journal of Criminal Justice, 2010. **38**(3): p. 276-281.
50. Odijk, D. and F. Kleijer, *Performance of high-sensitivity GPS for personal navigation at Schiphol Airport, the Netherlands*, in *IGNSS Symposium 2007*. 2007: University of New South Wales, Sydney, Australia.
51. van Diggelen, F. *Indoor GPS theory & implementation*. in *Position Location and Navigation Symposium, 2002 IEEE*. 2002.
52. Borenovic, M.N., et al. *Enhanced Cell-ID + TA GSM Positioning Technique*. in *Computer as a Tool, 2005. EUROCON 2005. The International Conference on*. 2005.
53. Bellini, A., et al., *A low cost positioning and visualization system using smartphones for emergency ambulance service*, in *Proceedings of the 2010 ICSE Workshop on Software Engineering in Health Care*. 2010, ACM: Cape Town, South Africa. p. 12-18.
54. Ladd, A.M., et al., *Robotics-based location sensing using wireless Ethernet*. Wirel. Netw., 2005. **11**(1-2): p. 189-204.
55. Horng, S.-J., et al., *Enhancing WLAN location privacy using mobile behavior*. Expert Systems with Applications, 2011. **38**(1): p. 175-183.
56. RadioAccess BV. *Private GSM Server*. 2010 [cited 2010 September 30]; Available from: <http://www.radioaccess.nl/index.php/produkten>.
57. Sheehy, A., *Worldwide Smartphone Markets: 2011 to 2015*. Generator Research, May 24, 2011: p. 280.
58. Marja van Bijsterveldt-Vliegenthart Dutch Secretary of State of Education, *Approach of the problem of leaving school early*. 2010. p. 1-2.
59. Steketee, M.J., et al., *Kinderen in Tel - Databoek 2010*, in *Kinderrechten als basis voor lokaal jeugdbeleid*. 2010, Verwey-Jonker Instituut: Utrecht.

## Appendices

### Appendix A Classroom observation form

Date: \_\_\_\_\_ - \_\_\_\_\_ - 2011

Time: \_\_\_\_\_ - \_\_\_\_\_

School name: \_\_\_\_\_

Place: \_\_\_\_\_

Course name: \_\_\_\_\_

Class type: \_\_\_\_\_

Number of students: \_\_\_\_\_

Number of absents: \_\_\_\_\_

Computer in classroom:

YES NO amount: \_\_\_\_\_

WLAN detected with own mobile phone?

YES NO comment: \_\_\_\_\_

Number of different device recognized with Bluetooth: \_\_\_\_\_

Mobile communication devices in the classroom	
Device type	Use intensity

General remarks (behavior, interests):

## Appendix B Questionnaire

### Enquête moderne media op school

Deze mini-enquête gaat over het gebruik van multimedia tijdens het dagelijks leven op school. De enquête begint met een aantal vragen over multimedia op school en de laatste vragen van de enquête gaan over het naar school gaan zelf. Door het invullen van deze enquête help je mee aan een onderzoek van de Universiteit Twente ter verbetering van de informatievoorziening en communicatiefaciliteiten voor leerlingen in het voortgezet en middelbaar beroepsonderwijs. Het invullen kost slechts 5 minuten en is uiteraard geheel anoniem. Er staan vragen op de voor- en achterzijde. Alvast bedankt voor je deelname aan deze mini-enquête!

#### Over jezelf

1. Wat is je geslacht? ☐ MAN ☐ VROUW

2. In welke klas zit je? \_\_\_\_\_

#### Mobiele telefoons

3. Hoeveel mobiele telefoons heb je? ☐ Ik heb geen mobiel ☐ 1 ☐ 2 ☐ 3 ☐ meer dan 3

4. Van welk merk en type is de mobiele telefoon die je het **meest** gebruikt? \_\_\_\_\_ ☐ N.V.T.

5. Hoe lang heb je die telefoon al?  
☐ korter dan een maand ☐ 1 - 6 maanden ☐ 6 - 12 maanden ☐ langer dan een jaar  
☐ N.V.T.

6. Wat is de hoofdreden om van telefoon te wisselen?  
☐ Abonnement loopt af/wordt verlengd ☐ Telefoon was gestolen/kwijt  
☐ Ik wil de trend bijhouden ☐ Telefoon is stuk gegaan ☐ Geen mening / n.v.t.  
☐ Anders, \_\_\_\_\_

7. Heb je een abonnement of maak je gebruik van pre-paid, met de meest gebruikte mobiele telefoon?  
☐ 1-jarig abonnement ☐ 2-jarig abonnement ☐ maandelijks opzegbaar abonnement  
☐ pre-paid ☐ N.V.T.

8. Heb je internet op je mobiel? ☐ JA ☐ NEE ☐ N.V.T.

9. Heb je een WLAN / WiFi poort op je mobiel? ☐ JA ☐ NEE ☐ WEET IK NIET ☐ N.V.T.

10. Hoeveel tijd in minuten maak je **gemiddeld** gebruik van je mobiele telefoon **tijdens één les**? ☐ N.V.T.  
☐ NIET ☐ 0-5 minuten ☐ 5-10 minuten ☐ 10-15 minuten ☐ meer dan 15 minuten



11. Hoeveel van die tijd in vraag 10 gaat op aan internetten op je mobiele telefoon? ☐ N.V.T.  
☐ GEEN ☐ 0-5 minuten ☐ 5-10 minuten ☐ 10-15 minuten ☐ meer dan 15 minuten

12. Hoe vaak maak je gebruik van je mobiele telefoon **tussen de lessen door**? ☐ N.V.T.  
☐ niet ☐ tussen sommige lessen ☐ tussen iedere les ☐ alleen in de pauzes  
☐ tussen iedere les én in de pauzes ☐ tussen sommige lessen maar wel altijd in de pauzes

13. Hoeveel van die tijd in vraag 12 gaat in totaal op aan internetten op je mobiele telefoon? ☐ N.V.T.  
☐ GEEN ☐ 0-5 minuten ☐ 5-10 minuten ☐ 10-15 minuten ☐ meer dan 15 minuten

14. Waar gebruik jij je mobiele telefoon het meest voor? ☐ N.V.T.  
☐ BELLEN ☐ SMSSEN ☐ INTERNETTEN ☐ ANDERS, nl.: \_\_\_\_\_

15. Gebruik je op school nog andere apparaten om met mensen in contact te komen?  
JA, namelijk met de volgende apparaten: \_\_\_\_\_ ☐ NEE

16. Is er draadloos internet (WLAN/WiFi) beschikbaar op school? ☐ JA ☐ NEE ☐ WEET IK NIET

17. Maak je hier gebruik van? ☐ JA, ALTIJD ☐ JA, VAAK ☐ JA, SOMS ☐ NEE ☐ N.V.T.

18. Als er gratis internet voor je mobiel beschikbaar zou zijn op school, zou je hier dan gebruik van maken?  
☐ JA ☐ NEE ☐ MISSCHIEN

19. Waarvoor gebruik je internet het meest?  
☐ E-MAILEN ☐ BIJHOUDEN VAN SOCIALE NETWERKEN (HYVES, FACEBOOK, etc.)  
☐ NIEUWSBERICHTEN LEZEN ☐ FILMPJES BEKIJKEN (YOUTUBE, DUMPART, etc.)  
☐ CHATTEN (messenger) ☐ BELLEN VIA INTERNET (Skype, Voice-over-IP)  
☐ IK GEBRUIK GEEN INTERNET ☐ ANDERS,  
namelijk: \_\_\_\_\_

20. Wat is je favoriete website? \_\_\_\_\_ ☐  
GEEN MENING

21. Heb je een Hyves account? ☐ JA ☐ NEE

22. Heb je een account bij andere sociale netwerken? ☐ JA, namelijk: \_\_\_\_\_ ☐ NEE

## Lessen bijwonen

23. Heb je in het verleden ooit gespijbeld? ☐ JA ☐ NEE

24. Spijbel je tegenwoordig nog weleens? ☐ JA, namelijk \_\_\_\_ keer per week / maand / jaar (omcirkel juiste termijn) ☐ NEE

25. Ervaar je de controle op spijbelen als streng?  
☐ NIET STRENG ☐ MATIG STRENG ☐ ERG STRENG  
☐ GEEN MENING

26. Wat is je belangrijkste reden om te spijbelen?  
☐ SLECHT GESLAPEN ☐ NIET GELEERD VOOR TOETS  
☐ IETS ANDERS GEPLAND ☐ BIJBAANTJE ☐ FAMILIE/VRIENDEN HELPEN  
☐ ANDERS, namelijk: \_\_\_\_\_  
☐ GEEN MENING ☐ VRAAG NIET VAN TOEPASSING

27. Wat moet er op school of tijdens de lessen veranderen zodat je minder of niet zou spijbelen?

( ☐ Vink dit hokje aan als je nooit spijbelt )

28. Zijn er zaken die je op school niet mag doen met je mobiele telefoon of computer, maar wel zou willen? Welke zijn dat?  
☐ NEE  
☐ JA, namelijk:

29. Heb je nog aanvullende opmerkingen? Noteer ze hier:

## Appendix C Documentation “Personal Information Grabber”

### Documentation “Personal Information Grabber”

A Java™ application to monitor user presence on your local network, based on social networks  
May 27<sup>th</sup>, 2011

Robbert Ottenhof

#### Table of Contents

1. Introduction
2. Installation and configuration
  - 2.1 System requirements
  - 2.2 Installation and configuration
3. Functionality: using the application
  - 3.1 Inspect directory
  - 3.2 Inspect file
  - 3.3 View results
  - 3.4 Start live capture
  - 3.5 Menu bar
4. Class diagram
5. Software description per class
  - 5.1 Package handlers
    - 5.1.1 Class HandlerUI
    - 5.1.2 Class DbConnect
    - 5.1.3 Class DirectoryReader
    - 5.1.4 Class FlowController
    - 5.1.5 Class StringToHex
    - 5.1.6 Interface Handler
    - 5.1.7 Class HyvesHandler
    - 5.1.8 Class FacebookHandler
    - 5.1.9 Class TwitterHandler
    - 5.1.10 Class HotmailHandler
  - 5.2 Package ui
    - 5.2.1 Class GUI
    - 5.2.2 Class NicDiscovery
  - 5.3 Package unused
6. Answers to research questions
7. Recommendations and extensions
  - 7.1 Small additions and modifications
  - 7.2 Handlers
  - 7.3 Database
  - 7.4 Other
  - 7.5 Different application areas of the software

## 1. Introduction

*<Chapter 1. Introduction already exists in this thesis and has been left out at this place>*

## 2. Installation and configuration

### 2.1 System requirements

Operating System: Windows OS preferred

Processor: 1.8 GHz

Memory: 512 MB minimal (1GB recommended)

Java: JRE 1.6 (Java 6)

Additional software: Wireshark's Tshark (available from [www.wireshark.org](http://www.wireshark.org)), MySQL

The application furthermore makes use of two external libraries, which are included in the software release of the truancy detection application.

### 2.2 Installation and configuration

1. Extract all files in the RAR archive or perform a SVN Checkout.

Using Eclipse:

2. File Menu -> New -> Java Project

3. In the "New Java Project" dialog choose "Create project from existing source" and click "Browse..." and choose the source directory

4. Click "Next >" and click on the tab "Libraries" and click on the button "Add External JARs..."

5. In the "JAR Selection" dialog, double click jcalendar-1.3.3 (in the root folder)

6. Click on the button "Add External JARs..." for a second time

7. In the "JAR Selection" dialog, double click mysql-connector-java-5.1.14-bin (in the mysqlconnector-java-5.1.14 folder)

8. Back in the "New Java Project" dialog, click "Finish"

Now the environment has been properly set but the JVM requires one additional setting:

9. In Eclipse, go to the Run Menu -> Run Configurations

10. In the "Run Configurations" dialog, click the "New launch configuration" button

11. Give this new launch a name, be sure the proper project is set in the field "Project" and enter the following into the field "Main class": ui.GUI

12. Now in the "Run Configurations" dialog, click the tab "Arguments" and enter the following in the "VM arguments" field: -Xms32m -Xmx512m

13. Click "Apply" and "Close" buttons (in that specific order)

Also be sure that the environment variable "JAVA\_HOME" on your machine is properly set.

Furthermore, this application is intended to communicate with a MySQL database. This is not necessarily a local database but for the application to run it needs a database to connect with. If there is however no remote database available, MySQL server is necessary to be installed which is available for free on the web. The SVN server and the latest version of the RAR archive contain a text file called "testdb.txt" which should be imported in the following way after installing MySQL:

14. Start a new command prompt and change to the folder ...\\MySQL\\MySQL Server X.X\\bin>

15. Type: mysql -u user\_name -p database\_name < testdb.txt

16. The database with the in 15. specified name should exist (tables will be imported from txt file)

17. Enter password and press ENTER

Finally, since the application prefers Windows OS and makes use of the tool "Wireshark," it is assumed that Wireshark (and with it Tshark) is installed in the folder C:\\Program Files\\Wireshark. If installed elsewhere, it is recommended to re-install it in the designated folder. The class GUI.java in the package "ui" contains a main method that does not require any arguments. This main method should be invoked to get it all running. What happens and the further possibilities after this invocation, is explained in

## Chapter 3: Functionality: using the application.

### 3. Functionality: using the application

The application has four different functions which can be selected from the main window. These functions will be explained in more detail in the four subsections of this chapter. To start with, a general introduction shows the necessary configuration steps before the main window shows up. After invocation of the main method of the GUI class, a configuration window will appear. The window

contains five text fields to set up a connection with a database. The first field contains the host name where the database is located. If the database is not hosted locally, the IP address of the host where the database is located should be filled in. Second text field contains the port number on the host where the database is listening. For example with MySQL this port number is 3306 by default. Third text field contains the name of the database, this is very user specific. Fourth and fifth text fields contain the credentials to login at the database that has been specified in the previous three text fields. Like with the database name, this is user specific. If all fields are filled in correctly and the “Apply” button is pressed, the main window will appear.

### 3.1 Inspect directory

Click the button “Inspect directory...” and a dialog window will appear in which a file can be selected from the file system. A random file can be chosen inside the directory that is intended to be read by the application. Clicking the “Open” button will start the reading of the directory in which the file has been selected. Instead of the “Open” button it is also possible to click the “Cancel” button to return to the main window of the application without taking any action. This might be useful in case the user forgot to select any of the four available handlers. By checking the boxes on the main panel, the corresponding handlers will be activated to use with the next action.

If however the “Open” button was selected, the application will start reading files in the folder in which the selected file is located. The application will try to read all available files in this folder but only capture (packet dump) files that are supported by Tshark and Wireshark can and will be processed. The supported file types are all raw data or packet dump files. Standard output file format of Wireshark / Tshark is libpcap format. This is also the format used by (amongst others) tcpdump. It is therefore supposed that the output of these tools will also work with the truancy detection application but this is not tested yet and outside the scope of the project.

The files do not need a specific file extension, the file format and an optional gzip compression will be automatically detected by Tshark. Compressed files however require the zlib library. Any subfolders in the directory that is being processed will be ignored. In case a file is not of one of the capture file formats, a parse error will occur (which will only be visible on the console, not on the GUI) and the application will proceed to the next file in the directory.

It is possible to stop a directory reading process at any moment by clicking the “Stop running” item in the File menu. This menu item appears after a directory reading process is activated, see also Section 3.5.

### 3.2 Inspect file

Click the button “Inspect file...” and a dialog window will appear in which a file can be selected from the file system. A so called PDML file should be chosen at this point, since this function only takes PDML files as input. Clicking the “Open” button will start the reading of the selected file. At this point it is also possible to click the “Cancel” button to return to the main window of the application without taking any action. Again like with directory inspection this might be useful in case the user forgot to select any of the four available handlers.

### 3.3 View results

Click the button “View results...” and a small dialog appears () in which a start date and an end date can be specified by either typing them in the text fields (format example: 20-mei-2011) or clicking on the icon on the right side of the text field and select the date on the calendar that will be shown.

After clicking the button “Apply dates” a table will be shown with all days between and including the selected start and end date. The table then shows all persons in the database with which it is currently connected, and shows for every person whether it was present on one or more of the selected dates, or that there is no information about him in the database. In that case “N/A” will be displayed in the table, if the person was however present, “Present” will be displayed in the table.

### 3.4 Start live capture

Click the button “Start live capture...” to start live data capturing data packets on one of the network interface cards of the current host. Clicking this button will result in the appearance of a window with quite a large dropdown which keeps a list of all network interfaces that are currently available at this host. By clicking the arrow on the right, the entire list will be shown and after

clicking on an entry in this list, live capture will start immediately using the selected network interface.

The application will create a folder C:\liveCap and will put all capturing files in this folder. Inside C:\liveCap it will also create a folder called “pdml” to store all converted files.

After converting a raw data (capture) file, the application will delete it to keep track of the current status and to avoid processing a capture file twice.

If it is not the first time that a live capture is about to start, all files in C:\liveCap will be deleted so the live capture will not process any outdated data. Subfolders of C:\liveCap will not be deleted.

In fact, live capturing is achieved with the help of the command “C:\Progra~1\Wireshark\tshark -i <capdevice> -b filesize:<capsize> -w <capfile>” where <capdevice> becomes the number of the selected capturing device (received using the drop down menu), <capsize> is the amount of memory after which Tshark switches writing to a new capturing file and <capfile> becomes the prefix of the output file names.

### 3.5 Menu bar

Besides the four main functions, the main window consists of a menu bar with three menus “File,” “Language” and “Help.” The File menu contains a Configuration item in which the database settings can be reviewed and edited, and an exit item to leave the application. If a directory inspection process is running, the file menu also contains a “Stop running” item, which can be clicked to abort a running directory inspection process.

The language menu contains five languages at the moment of writing (Dutch, English, German, French and Spanish) but only Dutch, English and German are supported at the moment. After clicking one of the supported languages, the text of the controls (buttons, menu items) will change to the chosen language. The help menu already has a “contents” item which is not implemented yet but this is a future extension in which the application should have built-in help functionality. The help menu also contains an “about” item which provides some information about the author of the application and the project it belongs to.

## 4. Class diagram

This page contains the class diagram, a structural overview of the application. Cooperation is explained below while all classes are explained in detail in the next chapter.

The GUI class contains the main method that should be invoked to start the application. A new GUI instance is created and in the constructor of the GUI class a new instance of HandlerUI is created. The new HandlerUI gets the just created GUI instance as argument.

After the HandlerUI instance has been created the configuration window can be shown to the user. It is important to first create a HandlerUI because it contains the important method createDbConnection which is used to establish a connection with the database.

So, after filling in the fields on the configuration window that appears and clicking the “Apply” button, this method createDbConnection in HandlerUI is called.

In this method, a new DbConnect instance is created, the object which handles all communication with the database. DbConnect only communicates with HandlerUI and GUI does not directly communicate with DbConnect so HandlerUI is the intermediate between DbConnect and GUI.

Now the three core objects of the application are initialized, the full functionality of the application can be used.

- If the user wants to inspect a PDML file, he clicks the “Inspect file...” button and the readFile method in HandlerUI will read the selected file.

- If the user wants to inspect a directory containing raw data files, he clicks the button “Inspect directory...” and selects a file in the directory he wants to process. Now the readDirectory method in HandlerUI is called and the class DirectoryReader will help HandlerUI in a separate thread to process the chosen directory.

- If the user wants to view the results stored in the database, he clicks the button “View results...” and a new small window will appear at which the user can specify a time span by

choosing a start and an end date using the JDateChoosers. For this function the library "jcalendar-1.3.3" is necessary

- If the user wants to start a live capture, he clicks the button "Start live capture..." and a window with a dropdown menu will appear. This menu contains all network interfaces available for use by Tshark. The values are received using the class NicDiscovery, see also Section 5.2.2. If an entry from the dropdown menu has been selected, live capturing will start using the selected network interface. This live capturing is further handled by the class FlowController which will run in a new created thread.

All functions except for the third listed above, depend on the four implemented handlers. These handlers can be thought of as filters which filter specific information from the data traffic, see also Sections 5.1.7 through 5.1.10 for detailed information per handler. Handlers can be enabled and disabled by respectively checking or un-checking the checkboxes on the main panel. All handlers have to implement the interface Handler and inherit its methods. Now each handler can be implemented differently but can be handled the same by for example instances of HandlerUI. This allows for easy extension of the application with additional handlers.

Any additional handler has to implement the Handler interface. Furthermore, the way the handler array is built in the HandlerUI class should slightly be adjusted to allow this new handler and finally the DbConnect class should get a method called "add<handler\_name>Account", analog to the other methods like for example addHyvesAccount.

First, Wireshark captures all data going through the selected network interface and stores it in a capture file, which is the raw data. Next, this data is converted to a PDML (XML packet detail) file, which is in fact XML, using Tshark. This PDML file will be read by the Java application which filters personal data, and stores all interesting events in a database. From this moment on, the database can be queried to view the results from the data capturing and filtering processes.

## 5. Software description per class

### 5.1 Package handlers

#### 5.1.1 Class HandlerUI

The HandlerUI class is more or less the core object of this truancy detection application. On the one hand because it acts as intermediate for all other objects, either direct or indirect. On the other hand, the HandlerUI class was one of the first classes that have been implemented and during the process of implementation the other classes have been built around it. Furthermore, this class does the heavy work of reading XML files. For three of the four functions of this application, this XML file reading is necessary. It might be clear that this is by far the most important task of the HandlerUI class. This reading task is performed by the method readFile which makes use of a SAX parser from the in Java available library.

Besides the file reading method, HandlerUI contains methods for configuration purposes.

- First, the current handler array can be set by invoking the method setHandlerArray with a String array as argument, where this String array is checked for codes which indicate the different handlers. If an array index contains the code "hy" it will put the Hyves handler in the new handler array. Other codes are "fb" for Facebook, "hm" for Hotmail and "tw" for Twitter.
- Second, a database connection can be established using the DbConnect class by invoking the method createDbConnection. This method takes five arguments; the hostname, the port number, the database name, the user name of the database user and the password for this user.
- Third, the method convertRawToPdml gets Tshark to convert raw packet dump files into pdml files. This is handled by an external process which is created after the invocation of the exec method of java.lang.Runtime. This exec method tries to execute the argument which is passed as a String upon invocation of the exec method. In this case this argument looks like: "conv.bat "+raw+" "+resFile where raw is the raw data file that should be converted and resFile is the file (including path) of the destination (PDML) file. The conv.bat is a file containing a batch script containing the following line: "C:\Progra~1\Wireshark\tshark -T pdml -r %1 > %2 -R http." This causes the system to run the command specified by this line, which causes the invocation of Tshark with the arguments "-T pdml" which states that the output format should be PDML. Furthermore, the "-R http" argument states that Tshark

should only process HTTP data. %1 and %2 represent the first and second argument from the String that is passed with the exec method, so “raw” and “resFile” respectively. The argument “-r” will cause the reading of the file specified by “%1” and the “>” causes that the output will be written to the file specified by “%2”.

#### 5.1.2 Class DbConnect

This class is responsible for all communications with the database. Upon startup of the application it establishes the connection with the database. Further on during the use of the application this class does all querying at the database. In fact, it translates from simple commands received from the class HandlerUI into correctly formed SQL queries, for example for adding events, accounts or persons to the database. For each of the four different handlers there is a separate method to add a corresponding account to the database while the method addEvent() takes care of adding an event to the eventList table.

The DbConnect class also has the responsibility to check whether it does not add duplicate information into the database, which is achieved by using the methods eventAppears() and accountAppears(). Since this class handles all database querying, it contains the method getPresence() which can be invoked to get an overview of the presence of the persons in the database during a specific time span; a start and end date can be specified on the GUI for this purpose, see also Section “3.3 View results”. So far this is the only possibility to show presence information. In the future this should be extended with additional functions in order to customize the request to the database, so to be able to be more specific in what the user wants to see.

#### 5.1.3 Class DirectoryReader

Instead of reading one single file, the DirectoryReader class helps the HandlerUI class to read all files in a specific folder which can be selected using the GUI. A new directory reading process is always performed in a new Thread because if the directory contains a lot of files, this should not prevent the user from doing something else in the meantime, like for example viewing results from the database. Upon starting a directory reading process, all files in the selected directory are put in a File[] array using the listFiles() method of the class File. Since this directory reading function requires raw data files for input, it will convert all files in the created array to pdml format and then the readFile() method in HandlerUI reads this XML-like pdml files. A for loop processes the File[] array.

#### 5.1.4 Class FlowController

The FlowController class knows almost the same behavior as the DirectoryReader class. There are two differences however, the first is that the input files become available from live real-time capturing (which are written into the folder C:\liveCap) and the second is that the raw data files will be deleted after they have been processed.

Since the input for this process is live capture data, the amount of data per time is considered not to be constant, it might fluctuate. For this reason, the FlowController has a built-in mechanism to avoid reading raw data files which are still in use by the system to write in. This is achieved by the second forloop

in the while-loop in the run-method.

The last two entries of the fileList are not processed by this for-loop. The last entry is the pdml folder in the liveCap folder, the predecessor of the pdml folder is the file which is not finished yet. This scenario requires that there are no other subfolders than the PDML folder in the liveCap folder, which is true by default (there could have been manually created another folder in the liveCap folder by the user though).

A probably better way to prevent from reading a file which is still written in by Tshark might be to check whether the file is in use. This could be done using a FileLock (java.nio.channels.FileLock).

#### 5.1.5 Class StringToHex

The values of the XML fields with tag names “show” and “showname” that appear in the PDML files often contain important and useful personal information about the users of the network. The format of these tag values is human readable ASCII text but the values are often limited to a specific number of characters, or their view is truncated by Wireshark. To overcome this, the corresponding “value” field needs to be read. The content of this field is in hexadecimal and needs to be converted to ASCII text before it is useful. The method convertHexToString() in the StringToHex class provides this conversion functionality. The class is however named “StringToHex” since it is also able to convert the other way round (ASCII text to hexadecimal).



#### 5.1.6 Interface Handler

This interface specifies all methods that should be implemented by a handler class which handles a specific type of data like for example Facebook. Whether these methods will get a body implementation in the handler classes and what these bodies would look like depends on how data is received from a web site for which a handler is built.

All methods that have to process data from an XML document are equipped with two String arguments. The first argument is the actual data that needs to be analyzed; the second argument identifies the type of data. This second argument allows to distinguish between for example ASCII text and hexadecimal data, so the handler knows whether it should convert it (in case of hexadecimal) or not (in case of ASCII).

- The method `getUserName()` tries to get the real name of a person for example "Robbert Ottenhof" out of the data.
- The method `getAccountName()` tries to get an account name out of the data, like for example the screen name belonging to a Twitter account.
- The method `getUserId()` tries to get an unique user identification number out of the data, if the particular service make use of user ids, like Facebook, Hyves and Twitter.
- The method `getMac()` is for future extension of the application. It is intended to match different profiles that belong to one person, like for example the Facebook account and the Hyves account of one person.
- The method `getTimeStamp()` gets the time stamp out of the packet in which information about some kind of login event has been found. This time stamp rendering is already performed in the HandlerUI class but it is redirected to the handlers to keep the correct time and event together. This works the same for all different handlers and will not be explained in the next sections that describe the different handlers.
- The method `getName()` returns the name of the specific handler so in case of communication with a Twitter handler this method will return a String with the value "twitter". This has proven to be useful not only for diagnostics during implementation but also when doing database operations.
- There are two different implementations of the method `getEmail()`. One without arguments which is useful in case the E-mail address is filtered by one of the other methods of the handler. This occurs for example in cases where the E-mail address is used as login name (so the account name or user ID is the same as the E-mail address because in fact it is the E-mail address itself, like with Facebook). If it is however not the case that E-mail address is the login name, the method `getEmail()` with two String arguments can be used for attempting to get the E-mail address out of the data.

In the next four sections, all currently implemented handlers will be described. Only the methods that have been implemented for this handler are explained.

#### 5.1.7 Class HyvesHandler

Because Hyves is a very popular and Dutch social network with eleven million users (or at least accounts)

and research showed that more than 70% of the students in secondary education have such an account, Hyves was chosen to be the first social network to build a handler for. At this moment it is also the only handler which does not need to use the hexadecimal converter but that is just a detail.

- The method `getAccountName()` filters the account names out of the data. With Hyves, an account can be viewed by typing the URL `<accountname>.hyves.nl` in the web browser. Hyves for computer browsers like Internet Explorer, makes use of "var global" to indicate a variable that is sent via the internet. The name of an account appears directly after the tag "var global\_member\_username." In case of a mobile Hyves user the account name is not always visible. The Android app sends a lot of information in XML format though. Each XML tag name and the tag's value are stored in an array with length two and then the tag name at the first index is checked and if it matches "username" the value which is in the second index will be stored as the account name.
- The method `getUserId()` check either for "var global\_member\_id" with computer users or for the tag "auth\_id" which is in the "HTTP GET"- message. User ID is always available, no matter what kind of device is used.
- The method `getUserName()` filters the real name of the Hyves user, if available. It searches

for "var global\_member\_fullname" in case of a computer login, and it also searches for the XML tags named "firstname" and "lastname."

- The method getEmail(String, String) is able to render the E-mail address in case of a login event with the Hyves Android app and searches for the XML tag "emailaddress."

#### 5.1.8 Class FacebookHandler

The second social network to build a handler for was Facebook.

- The method getUserId() searches for three different tags; navAccountInfo, m\_user and c\_user.

- The method getUsername() looks for the tag navAccountInfo. Besides that, there are refresh actions with which the name of the user is sent, so this method also looks for the tag "refresh name." The at-sign is not supported by every text format and is sometimes replaced for "%40." This is also the case with the Facebook data in the pdml files.

- For this reason, the getEmail(String, String) method searches for the tag "%40".

Provided that the tag "m\_user" is also available in the same String as where "%40" was found, the Facebook handler is now able to filter the E-mail address from the data. Note that this tag "m\_user" almost certainly indicates a mobile user, who will log in with his E-mail address being his user name.

#### 5.1.9 Class TwitterHandler

Third class in the row of handler development is the handler for the quite popular short message network Twitter.

Mobile Twitter events so far only provide account names, which are called "screen names" at Twitter.

Non-mobile events also contain user ID's, user names and E-mail addresses.

- The method getAccountName in this case searches for the tag "/mobile.twitter.com/" in combination with the tag "following" (both should appear in the same String). For computer users this method uses the tag "twtr.currentUserScreenName =".

- The method getUserId requires the tags verify\_credentials, id and friends\_count.

- The method getUsername requires the tags verify\_credentials, name and friends\_count.

- The method getEmail(String, String) requires the tags twtr.currentUser.sync and emailAddress.

The Twitter handler is not working properly. Although it filters all desired personal data using the methods stated above, it sometimes stores other data than the personal data that should be filtered from the data traffic. This requires improvement of the Twitter handler.

#### 5.1.10 Class HotmailHandler

The last handler that has been implemented so far is the one to handle data received from Windows Live, including Hotmail, MSN Messenger and eBuddy web messenger. Only two methods have been implemented since there are no account names and user ids involved.

- The getUsername method is able to find the name of a Hotmail computer user by looking for the tag "c\_meun" and finds a Hotmail mobile user name with the tag "<div class=\c\_ft\" ". An eBuddy user name is found in case the combination of the tags "Flcontent" and "screen" occurs in a single String. Finally the getUsername method is able to search in data provided by the MSN server in XML format. It then searches for the XML tags named "NameHint" and "LastNameHint".

- The method getEmail(String, String) is able to find the E-mail address of a user via the following ways: the combination of the tags "Bezig met aanmelden" and "@" or using the tag \"email\":" which in the data appears as "email:". The third possibility to find an E-mail address is at the occurrence of the tags "buddylist" and "unread" and "@".

## 5.2 Package ui

### 5.2.1 Class GUI

GUI.java has all responsibility regarding the communication between the application and its user. All windows that are shown to the user are built here and all actions the user takes are processed by this class. The latter is of course handled by the method actionPerformed() which belongs to the interface ActionListener which is implemented by the GUI class. The GUI class is the one of which the main method should be invoked to get the application running. In the main method a new GUI

instance is created and this constructor creates new instance of HandlerUI (see Section 5.1.1 Class HandlerUI) and then first builds and shows the configuration window (by invoking the method showConfigPanel() ) which should be used to enter the settings for a database connection. If these settings are applied and correct, the main window will be built and shown to the user. From this point on, a database connection is established and the full functionality of the application is available for use.

#### 5.2.2 Class NicDiscovery

This class helps the GUI class to identify all currently available network interfaces at the local host. First the method discoverNics() should be invoked. This method requests all for Tshark available capturing devices by executing "C:\Progra~1\Wireshark\tshark -D" using the exec() method of Runtime. The output of this command is collected using an InputStream and stored in a Vector<String>. Each line outputted by Tshark becomes a different entry in the Vector. Next action, which is the responsibility of the GUI class, is to invoke the method getNicList() which returns the Vector in which the identified network interfaces were stored during the last call of discoverNics(). The Vector returned by getNicList() is used as input for the JComboBox "nicList" in the class GUI.

#### 5.3 Package unused

This package contains all classes that were planned at the start of the implementation of this truancy detection application but turned out that will not be used in operational phase. This is mainly due to the fact that a database is used to store all events, persons and accounts that are found in the data that has been analyzed and probably will be analyzed in the future.

### 6. Answers to research questions

*<Chapter 6 Answers to research questions already exists in this thesis and has been left out at this place>*

### 7. Recommendations and extensions

#### 7.1 Small additions and modifications

- Warning dialog at start live capture that says: "all files in C:\liveCap will be deleted!" because now all files in this folder are deleted without any warning.
- Dialog when clicking Exit-button; "Do you really want to exit? YES/NO." This is already implemented but still commented in the GUI class.
- Dialog when selecting a network interface from the dropdown. Now the live capture will start immediately after selecting a NIC and releasing the mouse button in the dropdown menu. It is probably better to give the user the chance at this point to confirm or cancel live capture using the selected NIC, instead the chosen NIC is accidentally incorrect.
- Add language packs (French / Spanish / Italian / Swedish / Danish / Norwegian).
- Catch the NullPointerException if getAttributeNode("show").getNodeValue() is "null"
- ToolTipText of "Inspect directory" button is wrong; it says that this action requires \*.pdml files but actually, this action requires raw data (packet dump) files.
- Fix something for the "Inspect directory" function, if a file is not of the correct (capture) format. Now an error occurs and the application will proceed with the next file in dir.
- Fix something for the "Inspect file" function, if a file is not of the correct format (should be PDML). Check for example the file extension (it should end with "cap" because of the \*.cap and \*.pcap file formats).
- A radio button on the main window to select whether RAW data or PDML data will be read. This setting should become connected to both directory inspection and single file inspection, so both functions can handle both types of data formats.

#### 7.2 Handlers

- Twitter handler passes desired data, but it seems that it sometimes also passes more than that. This holds that the Twitter handler needs an additional revision.
- Facebook handler: add a "name search" functionality using the internet: Runtime.exec( C:\Progra~1\Intern~1\iexplore www.facebook.com/profile.php?id=<fbUserId\_that\_has\_been\_found> ) using this command it is possible to start Internet Explorer and search for the homepage of the user that matches the Facebook user ID that has been found. This requires a facebook account and a logged in user but this is no problem since it is easy and free to create

an account. Then the box “remember me” at login with Facebook should be checked once and now it is possible to do some additional research by actively searching for the name of the Facebook user that matches the account with the user ID found in the initially captured data.

- Add YouTube handler (uses google accounts). Since login at YouTube happens using a Google account, and Google accounts are also used to login elsewhere, for example at other Google services like Google docs, it is expected that this E-mail address could support the reciprocal linking of different accounts belonging to one person. This becomes useful in case it is possible to render a name at one account of a person, for example at Hyves, but not at another account for example YouTube. If it is however possible to see the E-mail address (i.e. the user ID) of the login event at YouTube, and this is the same E-mail address as used with a Hyves account, then it is known which users logs in with this Google account at YouTube and from that moment on the YouTube login events also become useful since it was able to connect a person’s name to this events based on the match of the E-mail addresses that were used.
- Add Gmail handler (uses google accounts) for the same purpose as the YouTube handler, a Gmail handler could be built, in order to collect Gmail E-mail addresses.
- At some points the handlers behave quite static regarding information retrieval. They look for a specific tag in a String and then take a specific part of this String by cutting the head and the tail at predefined String indexes, for example: `result = data.substring(5, 24)`. This might cause problems if web pages and apps of these social networks will release a newer version of the software or website. There is a reasonable chance that they will keep using the same tags but the positions of these tags in the String might be changed or shifted a few characters forth or back. For this reason it seems better to let the index depend on a tag or token in the String, instead of always cutting the String at the same indices.
- At some points, the HandlerUI class processes data twice but processing it once is sufficient. Solving this will increase the performance.

#### 7.3 Database

- Error table to create error log; every time an exception occurs, this should be put in the database. This log could give an overview of weaknesses of the application and could probably help identifying why something went wrong during the use of the application.
- Reverse engineering for “eventList” table in database; if additional information (like a person’s name) is available after a long time of only seeing a user-ID, for example with facebook or some mobile Hyves, update all previous entries in the person column of the “eventList” table.
- Cross-reference should take place between tables that hold accounts for the different social network sites. Information could be exchanged if E-mail addresses match.
- A list should be set up with exceptional cases that could cause an error in the SQL syntax. An example is names like “O’neill” since the ‘-character is used to indicate start and end of a String, so this could cause a wrong start or end (probably end) indication.

#### 7.4 Other

- Build help menu that becomes visible after clicking the item “Contents” in the “Help” menu.

#### 7.5 Different application areas of the software

- Emergency; in case of an emergency this application could provide an overview of people who recently entered the building or who entered the building after the last moment it has been opened (for example since the start of this day). This supports evacuation and gives an overview of people who should be searched for in case of an emergency. If the building is large, the system could send warning messages to all people who entered the building, for example in the last hour before the emergency alert.
- “TravelBuddy”; if this application would be installed in a train that offers free WiFi, it could give an overview of people who are currently using the network in this train. In this way, other people could see whether there are any friends of them travelling with them in the same train, without the need of walking through the entire train. This area of application requires client software that listens to a server which will keep track of the current or recent users in the train.