ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ -
ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΔΙΟΙΚΗΣΗ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Privacy Enhancing on Mobile Devices: Continuous Authentication with Biometrics and Behavioral Modalities.

Ενίσχυση της Ιδιωτικότητας σε Κινητές Συσκευές: Συνεχής Αυθεντικοποίηση με χρήση Μορφολογικών και Συμπεριφορικών Βιομετρικών.

Του Ιωάννη Στύλιου (Α.Μ. 323Μ/2014022)

Επιβλέπων : Αναπληρωτής καθηγητής Σπύρος Κοκολάκης

Σάμος, Μάρτιος 2016

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ -
ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΔΙΟΙΚΗΣΗ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Privacy Enhancing on Mobile Devices: Continuous Authentication with Biometrics and Behavioral Modalities.

Του Ιωάννη Στύλιου (Α.Μ. 323Μ/2014022)

Επιβλέπων : Αναπληρωτής καθηγητής Σπύρος Κοκολάκης

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή .

Αν. καθ. Κοκολάκης Σπύρος Επ. καθ. Ριζομυλιώτης Παναγιώτης Αν. Καθ. Καμπουράκης Γεώργιος

(Υπογραφή) *(Υπογραφή)* *(Υπογραφή)*
....

Σάμος, Μάρτιος 2016

(Υπογραφή)

…………………………….

Η παρούσα εργασία αφιερώνεται στους γονείς μου.

ΠΡΟΛΟΓΟΣ

Η διπλωματική αυτή εργασία πραγματοποιήθηκε στα πλαίσια του μεταπτυχιακού προγράμματος σπουδών «Ασφάλεια Πληροφοριακών & Επικοινωνιακών Συστημάτων» του τμήματος Μηχανικών Πληροφοριακών & Επικοινωνιακών Συστημάτων του Πανεπιστημίου Αιγαίου, κατά το χειμερινό εξάμηνο του Ακαδημαϊκού έτους 2015-16. Επιβλέπων καθηγητής και καθοδηγητής της εργασίας ήταν ο αναπληρωτής καθηγητής Σπύρος Κοκολάκης. Το θέμα της εργασίας είναι: «Privacy Enhancing on Mobile Devices: Continuous Authentication with Biometrics and Behavioral Modalities». Η έρευνα της εργασίας με τίτλο: “Mobile Phones & Behavioral Modalities: Surveying users’ practices” έγινε δεκτή με διαδικασία κριτών και παρουσιάστηκε στο International *IEEE* Conference TELFOR 2015. November 25, 2015, SAVA Center, Belgrade, Serbia. http://www.telfor.rs

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστώ τον επιβλέποντα καθηγητή μου, κ. Σπύρο Κοκολάκη για τις πάντα εύστοχες και πολύτιμες συμβουλές και παρατηρήσεις του, καθώς και για την βοήθειά του σε όλη την διάρκεια της παρούσας εργασίας.

Ευχαριστώ πολύ τον καθηγητή Σωτήρη Χατζή από το Τεχνολογικό Πανεπιστήμιο Κύπρου για τη βοήθειά του και τις πολύτιμες συμβουλές και παρατηρήσεις του.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και την Όλγα Θάνου για την πλήρη στήριξή τους στην ολοκλήρωση της διαδικασίας συγγραφής της παρούσας εργασίας καθώς και σε όλη την πορεία του μεταπτυχιακού προγράμματος.

Abstract

Mobile phones are one of the most popular means of access to the internet. Users, via the telephone, connect to different services such as: Google, social networks, work accounts, banks accounts, etc. Those services, are oftentimes, left running on their device. This practice entails risks, such as, loss or/and the violation of their personal data. Also, the stealing of the device, after login, grants full access to sensitive data and applications. For all the above reasons, Continuous Authentication (CA) systems have been suggested in literature. CA systems represent a new generation of security mechanisms that continuously monitor user behavior and use this as basis to re-authenticate periodically throughout a login session.

In the present thesis a literature review was carried out on topics including the following: Continuous Authentication, Privacy, Users Attitudes, Biometrics, Behavioral Modalities. In the literature review we present a collection of selected published sources relevant to the topic of the thesis, which are accompanied by annotation, critical analysis of contents and apposition in some cases of the main conclusions of each study. The purpose of the literature review is the critical analysis of the contents and the detection of possible gaps in the literature on the particular topic.

In order to answer to the research questions that have been posed from these research areas we conducted two corresponding surveys with two original questionnaires in which we had a total of 304 participants from Greece and Cyprus. The purpose of these surveys has been to identify users' attitudes with regard to the protection of their sensitive personal data, as well as users' practices pertaining to certain behavioral modalities.

In the first survey, we examine whether users adopt some basic practices to protect their sensitive personal data themselves, or there is a need to further strengthen their protection. For purposes of statistical analysis, our main variable is age because we wanted to evaluate the significance degree regarding users' attitudes and practices among different age groups. Finally, we seek the factors that influence the attitude of users with respect to their practices for the protection of personal data through statistical hypotheses.

In the second survey, we analyze the most salient patterns characterizing user practices regarding certain behavioral modalities including: the way of using various applications, power consumption, touch gestures and guest users' habits. This can offer qualitative information, for the different behaviors / "characters" of users. What we want to see via our questionnaire is whether users do perform similar tasks at a certain time of the day. In addition, through this approach we want to examine under what basis the user's profile can be created in order to be used in further research regarding user's Continuous Authentication.

In the third part of our research work we present an Experimental Procedure and the Behavioral Biometrics Data Collection Architecture for mobile devices. In the present experiment we recorded modalities of movement imprinting the user's walk patterns.

Our methodology imprints the modalities of movement, by the accelerometer and gyroscope sensors, of 10 volunteers in total. The procedure was designed in such a way so as to collect data from every participant for three sessions. The sessions recorded three sequences of 10 minutes each while the participant: walked and hold the device on his hand, walked and had the device on his pocket, was running and had the device on his pocket. These sessions were repeated for two days and gives us a total of 60 minutes' real use data of the smartphone for each user.

Keywords — Mobile Phones, Privacy Risk, Behavioral Modalities, Biometrics, Users Attitudes, Continuous Authentication, Survey.

Table of contents

1

Introduction

The wealth of services that were made available over the last few years including access to emails, social media, banking, etc. lead to the rise of the amount of sensitive data stored on or processed by handheld devices [ALZ2014]. The users choose easy to remember passwords, for all their tasks; thus, the level of protection decreases significantly [ACBS2009]. Even though the password is demanded frequently, an attacker could gain access to the device after the successful authentication of the legitimate user, and misuse all sensitive data [ALZ2014], [BZJ+2014]. In addition, despite the fact that mobile phone's security measures have been increased during the last years, users don't take the necessary measures to avoid a possible unauthorized access and/or sensitive data retrieval from their mobile phone [CF2005]. Finally, there is a plethora of recent work that indicates that password authentication is not appropriate for mobile devices [D1999], [FBM+2013].

Continuous Authentication (CA) systems represent a new generation of security mechanisms that continuously monitor user behavior and use this as basis to re-authenticate periodically throughout a login session. CA has been around for about a decade. As a result a limited amount of research work has been produced to date, and the first commercial products have only recently started reaching the market. We attempt, in this chapter, to provide some general perspectives in order to help achieve some common and better understanding of this emerging field. The chapter introduces basic CA concepts and terminologies, discusses the characteristics of CA data sources, and identifies major areas of application for CA systems [AT2001].

The term “biometrics” is derived from the Greek words “bio”, meaning “life”, and “metrics”, meaning “to measure”, and it is a method through which we can establish the identity of a person based on physical or behavioral attributes of that person.

Biometrics have been used for authentication purposes, that is to verify that a person is who he claims to be (also known as “positive recognition”). In that case, verification (or authentication) is established by comparing a biometric captured by the person to be verified (e.g. a fingerprint) against a previously captured biometric template of the same type and from the same person. The first verification biometric systems used hand geometry recognition and were mainly used for physical access control and for recording time and attendance. More recently many banks, across the world, use biometric authentication in order to verify their customers and grant them access to ATMs [G2007]. There are also stores that use fingerprint recognition for biometric payments. Car manufacturers incorporate biometric authentication systems into newest cars, in order to unlock their doors or start the ignition [I2015]. Mobile phones can now capture and store biometric templates as well, allowing their owners to authenticate to their devices by using their own biometrics.

On September 2013 Apple released the iPhone 5S, the first mobile device with an embedded fingerprint scanner. Since then, millions of users across the world have been using their thumbs in order to unlock their devices, purchase applications and authenticate to remote services. During the following years, many phone manufacturers followed: Samsung, HTC, LG, Huawei, Xiaomi, Oppo, ZTE are only some of the manufacturers that have included fingerprint scanners on their phones and tablets [S2015]. Moreover, many manufacturers have gone a step further and have incorporated (or plan to incorporate) sophisticated iris and/or retina scanners into their newest models (Microsoft 950 XL, Fujitsu NX F-04G, ZTE Grand S3, etc.). [S22015]. Biometric authentication seems to have gained mobile users’ acceptance, and, according to the Biometrics Research Group, biometric smartphone users will increase, from 200 million users in 2015, to two billion users by 2020 [O2015].

For many years, providing something the user knows (e.g. a PIN or a password) has been the most popular method to authenticate the identity of a person. Something the user has (e.g. a hardware token) is also often used, usually as a supplementary, 2nd factor authentication class for critical applications (e.g. e-banking). However, both the aforementioned methods have some serious drawbacks that have increased the need for the adoption of a third authentication class: what the user is. Using biometrics for user-to-device authentication has many advantages over the aforementioned methods. More specifically:

- Biometric authentication is based on traits that are unique to each individual and rarely change over time, thus providing a more reliable identification method than traditional authentication methods.

- Biometric traits are very hard to forge (although not impossible), and they can't be guessed, as is the case with PINs and passwords. Moreover, users cannot pass their biometric characteristics to other users as they can do with their passwords or cards, thus providing true and complete accountability.
- Users aren't required to remember anything (e.g. PINs or multiple complex passwords that need to be changed frequently) or carry things with them (e.g. cards, tokens). Our identity is always with us. We cannot lose it nor forget it.
- Using our biometrics to identify ourselves can be a very fast and easy process, thus providing a user-friendly and convenient authentication method.

Certainly, biometrics have some weaknesses but in the literature some countermeasures have been proposed. The weaknesses in the biometrics are summarized below [B2015]:

- A password is secure as long as it is hidden: Could biometrics be hidden? (Of great importance are physical presence and integrity of the biometric).
- Publication of geometric algorithms: In cryptography the algorithms are known and security derives from the secrecy of the key. In biometry knowledge of algorithms reduces safety. (e.g. Hill climbing attack).
- Recall of biometric patterns: If a code leaks, it can be recalled and a new one can be issued. If a biometric pattern leaks, what can be done?
- Deception (spoofing): How the creation of counterfeit biometrics can be addressed? (e.g. high resolution photos, facial casts, synthetic fingerprints etc.).

1.1 Subject of the thesis

In the present thesis a literature review was carried out on topics including the following: Continuous Authentication, Privacy, Users Attitudes, Biometrics, Behavioral Modalities. In the literature review we present a collection of selected published sources relevant to the topic of the thesis, which are accompanied by annotation, critical analysis of contents and apposition in some cases of the main conclusions of each study.

Afterwards, a grouping of the gathered literature sources took place, based on some of their common characteristics, such as the research problem, the goals / objectives, the research approach, the findings, etc. The grouping of the studies resulted in the following categories:

- Privacy on Mobile Devices.

- Continuous Authentication on Mobile Devices using Biometrics & Behavioral Modalities.

Furthermore, our methodology uses a categorization table that shows which studies fall into each category and the timeline of the posts.

In order to answer to the research questions that are being posed from these research areas we conducted two corresponding surveys with two original questionnaires in which we had a total of 304 participants from Greece and Cyprus. The purpose of these surveys is to find the users' Attitudes regarding to the protection of their sensitive personal data as well as the users' practices on certain behavioral modalities.

In the first survey we examine if the users adopt some basic practices to protect their sensitive personal data themselves or if there is a need to further strengthen their protection. For the statistical analysis, our main variable is *age* because we wanted to evaluate the significance degree regarding users' Attitudes and Practices between different age groups. The target group of the survey are 204: students, employees and members of University of Athens and University of the Aegean. Our survey was conducted using in-person delivery technique with a multiple-choice questionnaire. It consists of four subsections and is formed as follows:

- Demographics
- Storage Practices
- PIN Practices
- Device Protection

Our first survey answers three main research questions:

- What are the Users' Attitudes on Mobile Devices?
- Can the users' practices protect their sensitive data?
- Is there a need to strengthen the protection of users' personal data through a Continuous Authentication System with biometrics & Behavioral modalities?

Finally, we seek the factors that influence the attitude of users with respect to their practices for the protection of personal data. Furthermore, we search for the factors that influence the attitude of users with respect to their practices for the protection of personal data. To achieve this we have investigated, through statistical hypotheses, if the *age* and *gender* of the users relate to their practices for the protection of their personal data.

In the second part of the research we analyze the most salient patterns characterizing user practices regarding certain behavioral modalities including: the way of using the various applications, power consumption, touch gestures and guest users' habits. To this end, we used an original questionnaire, created for the needs of the specific survey, to examine whether we can find some trends among the users. This can give us a qualitative information, for the different behaviors / "characters" of users. The target group of the survey are 100: professors, students, employees and members of Technological University of Cyprus. What we want to see via our questionnaire is if users do perform similar tasks at a certain time of the day. In addition, through this approach we want to examine under what basis the user's profile can be created in order to be used in further research regarding User's Continuous Authentication.

Our survey answers two of our main research questions:

- What are the behavioral modalities among the users?
- Can analysis of Behavioral Modalities be utilized in the context of Continuous User Authentication?

We then examine some statistical hypotheses concerning age in relation to the use of applications in a specific part of the day, the time period of use by the users and the correlation between age and communication. To check if these hypotheses apply we use the non-parametric Kruskal – Walis test since we don't have a normal distribution and we have more than two groups to check. We examine if there is a statistically significant difference between the age groups in correlation to the variables. Survey responses were analyzed using descriptive analysis, Crosstabs, Frequencies and Kruskal–Wallis test ($p<.05$) on SPSS.

In the third part of the research we present an experimental biometric data collection process by a mobile device. In the present experiment we recorded modalities of movement imprinting users' walk patterns. Our methodology imprints the modalities of movement, by the accelerometer and gyroscope sensors, of 10 volunteers in total. The procedure was designed in such a way so as to collect data from every participant for three sessions of ten minutes each with a break of 5 minutes for instructions. The session recorded three sequences of 10 minutes each while the participant: walked and hold the device on his hand, walked and had the device on his pocket, was running and had the device on his pocket. These sessions were repeated for two days so as to effectively capture the biometric behavior of the user. This gives us a total of 60 minutes’ real use data of the smartphone for each user.

1.1.1 Contribution

Initially, the purpose of the literature review is the critical analysis of the contents and the detection of possible gaps in the literature on the particular subject / topic.

The purpose of the first survey is to find the users’ trends that relate to their attitudes regarding to the protection of their sensitive personal data. Moreover, to see if there is a need for further protection via a Continuous Authentication System.

The purpose of the second survey is to find the users’ practices on certain behavioral modalities like: the application’s way of use, power consumption, touch gestures and guest users’ habits. In addition, the users’ Authentication, by application’s way of use, is not proven in the literature. So, through our survey, we want to see if this would work well in order to be used in further research. The usefulness of this survey is important because it combines four behavior modalities, and gives information about the users’ practices concerning their mobile devices. The results can be used by other researchers, as a potential guide, in works about User’s Continuous Authentication using Behavioral Biometrics. Also, it can be used by mobile operators for their future technological investments.

In the third part of the research we will present an experimental biometric data collection process by a mobile device. In the present experiment we recorded modalities of movement imprinting the user's walk patterns. We will present the Data Collection Architecture by which we can collect the biometric data of the users, the way and type of storage and the Data Preparation for introduction to machine learning algorithms. This knowledge also can be used by other researchers, as a potential guide, in works about User’s Continuous Authentication using Behavioral Modalities.

The contribution of the thesis is summarized as follows:

- A literature review was carried out on subjects like: «Continuous Authentication, Privacy, Behavioral Modalities, Biometrics, Users’ Attitudes».
- A survey was conducted concerning Users’ Attitudes on Mobile Devices: Can the users’ practices protect their sensitive data?
- A survey was conducted concerning Mobile Phones & Behavioral Modalities: Surveying users’ practices”.

- Experimental Procedure: Behavioral Biometrics Data Collection Architecture.

1.2 Organization of text

In Chapter 2 we present the Literature review as well as the Methodology and Scope of the literature review.

In Chapter 3 we present the Background of the Study and we make a reference to: Continuous authentication, Assessment of Biometric Characteristics, Statistical Methods.

In Chapter 4 we present the first survey’s Results: “Users’ Attitudes on Mobile Devices: Can the users’ practices protect their sensitive data.

In Chapter 5 we make a presentation of the second survey’s Results: “Mobile Phones & Behavioral Modalities: Surveying users’ practices”.

In Chapter 6 we present the Experimental Procedure and the Behavioral Biometrics Data Collection Architecture.

In Chapter 7 we present the Conclusion, Summary and findings, and Future extensions.

In Chapter 8 we present the Bibliography.

In Chapter 9 the Appendix.

2

Literature review

In this chapter we present a collection of selected publications which are relevant to the subject of our thesis /research. Furthermore, they are accompanied by an analysis of context and apposition of the basic conclusions of every study/ research. In addition, the Categories and the Determination of entry and exclusion criteria are reported.

2.1 Methodology and Scope of literature review

Our methodology is based on the collection of selected publicated sources which are relevant to the subject of our thesis / research. Moreover, they are accompanied by annotation, critical analysis of content and apposition, in some cases, of the main conclusions of each study / research. There will be no limitation in books and journal articles only, but the subject of the literature review may also be other information material, such as websites. A necessary prerequisite of systematic search for suitable publications is the definition of indexing terms. In order to increase the efficiency of search we used combined indexing words like «and» / «or» / «not». Some of the indexing terms that we used are the following: Mobile Phones, Privacy Risk, Behavioral Modalities, Biometrics, Users Attitudes, Continuous Authentication, Survey. Given that decisions regarding the inclusion or exclusion of publications involve a degree of subjectivity, the appropriateness or not of the publications was considered by two researchers.

Afterwards, the grouping of the gathered literature sources took place, based on some of their common characteristics, such as the research problem, the goals / objectives, the research approach, the findings, etc. The grouping of the studies resulted in the following categories:

1. Privacy Risk on Mobile Devices: In this category the publicated studies that were selected investigated Privacy Attitudes and Preferences on Mobile Devices.
2. Continuous Authentication on Mobile Devices using Biometrics and Behavioral Modalities: In this category were selected studies which subject of research was Continuous Authentication with Biometrics and Behavioral Modalities.

In addition, our methodology uses an extremely useful tool, the literature distribution table that shows which studies fall into each category as well as the timeline of the publications presented below.

The purpose of literature review is the critical analysis of the contents and the detection of possible gaps in the literature of the particular subject / topic.

2.2 Privacy on Mobile Devices

The timeline of publications concerning various issues of Privacy on Mobile Devices are presented in the following table:

Study	Context	Methodology	Participants
N.L. Clarke, S.M. Furnell. 2005	Authentication of users on mobile telephones – A survey of attitudes and practices	Survey	Mobile subscribers
Shane Ahern, Dean Eckles, Nathaniel S. Good, Simon King,Mor Naaman, Rahul Nair. 2007	Over-exposed?: privacy patterns and considerations in online and mobile photo sharing	Study-Interviews	Users
Stan Kurkovsky, Ewa Syta. 2010	Digital natives and mobile phones: A survey of practices and attitudes about privacy and security	Survey	Young people
Erika Chin, Adrienne Porter Felt, Vyas Sekar, David Wagner. 2012	Measuring user confidence in smartphone security and privacy	User study	Smartphone users
Mark J. Keith, Samuel C. Thompson, Joanne Hale, Paul Benjamin Lowry, Chapman Greer. 2013	Information disclosure on mobile devices: Re-examining privacy calculus with actual user behavior	Controlled experiment	Consumers

In 2005, Clarke and Furnell [CF2005] conducted a survey of 297 mobile subscribers, with the attempt to assess their use of mobile devices, their use of current authentication methods, and their attitudes towards future security options. The findings revealed that the majority of the respondents make significant use of their devices, with clear demands for protection against unauthorized use. However, the use of current PIN-based authentication was marked as problematic, with a third of the respondents indicating that they do not use it at all, and other problems being reported amongst those that do. In view of this, the respondents' opinions in relation to future security options are interesting, with 83% being willing to accept some form of biometric authentication on their device.

Ahern et al [AEG+2007] used context-aware camera phone devices to examine privacy decisions in mobile and online photo sharing. Through data analysis on a corpus of privacy decisions and associated context data from a real-world system, they identified relationships between location of photo capture and photo privacy settings. Their data analysis led to further questions which they investigated through a set of interviews with 15 users. The interviews revealed common themes in privacy considerations: security, social disclosure, identity and convenience.

Kurkovsky and Syta [KS2010] presented the results of a survey of over 330 young people, namely known as digital natives, aged 18 to 25. They attempted to evaluate their use of mobile technology, their attitudes about security and privacy as it relates to mobile phones, as well as their perceptions of different ways how security and privacy could be improved in future mobile devices. Despite a commonly held belief that digital natives are technologically savvy, their self-assessment did not appear to support this statement. Furthermore, despite the respondents' awareness of various threats to security and privacy, very few of them actually took any concrete steps to protect their devices from unauthorized access.

Aviv et al [AGM+2010] conducted an experiment to test the feasibility of a smudge attack via photography. A smudge attack is a method to discern the password pattern of a touchscreen device such as a cell phone or tablet computer. The smudge attack relies on detecting the oily smudges left behind by the user's fingers when operating the device using simple cameras and image processing software. Under proper lighting and camera settings, the finger smudges can be easily detected, and the heaviest smudges can be used to infer the most frequent user input pattern (the password). The researchers were able to break the password up to 68% of the time under proper conditions. Smudge attacks are a threat for three reasons. First, smudges are surprisingly persistent in time. Second, it is surprisingly difficult to incidentally obscure or delete smudges through wiping or pocketing the device. Third and finally, collecting and analyzing oily residue smudges can be done with readily-available equipment such as a camera and a computer. A smudge attacker is within reason considering search and seizure procedures in many countries.

Chin et al [CFS+2012] conducted a user study involving 60 smartphone users. First, they interviewed users about their willingness to perform certain tasks on their smartphones to test the hypothesis that people currently avoid using their phones due to privacy and security concerns. Second, they analyzed why and how they select applications, which provided information about how users decide to trust applications.

Keith et al [KTH+2013] proposed and tested an experimental methodology designed to replicate real perceptions of privacy risk and capture the effects of actual information disclosure decisions. Subsequently, they reported the results of a controlled experiment involving consumers (n=1025) in a range of ages, levels of education, and employment experience. Based on their methodology, they found that only a weak, albeit significant, relationship exists between information disclosure intentions and actual disclosure. In addition, this relationship is heavily moderated by the consumer practice of disclosing false data.

2.3 Continuous Authentication on Mobile Device using Biometric & Behavioral Modalities

Similarly, a recent survey of Androulidakis et al. [ACB+] presented, at conclusions, some behavioral modalities regarding power consumption. Kim et al [KCH2010] proposed an enhanced multimodal personal authentication system for mobile device security, which fuses information obtained from face, teeth and voice modalities to improve performance. In addition, some other behavioral modalities are presented like the touch screen behavior, where screen touches are a behavioral biometric according to Frank et al. [FBM+2013]. Feng et al [FZS2013] exploited mobile motion data as a novel biometric modality and their experimental results showed that user movements (e.g., walking) have a high impact on the verification performance. Moreover, Bo et al. [BZJ+2014] showed that the touch screen behavior can identify transitions or change of hands between the device owner and a guest who may or may not be a known entity.

Study	**Context**	**Methodology**	**Participants**
Dong-Ju Kim, Kwang-Woo Chung, Kwang-Seok Hong. 2010	Person authentication using face, teeth and voice modalities for mobile device security	Experiment	Volunteers
Frank, M., Biedert, R., Ma, E., Martinovic, I., Song, D. 2013	On the applicability of touchscreen input as a behavioral biometric for continuous authentication	Experiment	Smartphone users
Tao Feng, Xi Zhao; Weidong Shi. 2013	Investigating Mobile Device Picking-up motion as a novel biometric modality	Experiment	Volunteers
Androulidakis, I., Levashenko, V., Zaitseva, E. 2014	Smart phone users: Are they green users?	Survey	Smartphone users
Bo, C., Zhang, L., Jung, T., Han, J., Li, X.-Y., Wang, Y. 2014	Continuous user identification via touch and movement behavioral biometrics	Experiment	Smartphone users

Study	Context	Methodology	Participants
Kwapisz, J.R., Weiss, G.M., Moore, S.A. 2010	Cell phone-based biometric identification	Experiment	Users
Shi, E., Niu, Y., Jakobsson, M., Chow, R. 2011	Implicit authentication through learning user behavior	Experiment	Users
Riva, O., Qin, C., Strauss, K., Lymberopoulos, D. 2012	Progressive authentication: deciding when to authenticate on mobile phones	Experiment	Users
Zhang, L., Tiwana, B., Qian, Z., Wang, Z., Dick, R.P., Mao, Z.M., Yang, L. 2010	Accurate online power estimation and automatic battery behavior based power model generation for smartphones	Experiment	Users
Murmuria, R., Medsger, J., Stavrou, A., Voas, J.M. 2012	Mobile Application and Device Power Usage Measurements	Experiment	Users
Shye, A., Scholbrock, B., Memik, G. 2009	Into the wild: studying real user activity patterns to guide power optimizations for mobile architectures	Experiment	Users
Rahul Murmuria , Angelos Stavrou, Daniel Barbará, Dan Fleck. 2015	Continuous Authentication on Mobile Devices Using Power Consumption, Touch Gestures and Physical Movement of Users	Experiment	Users

Kwapisz et al. [KWM2010] published a system to identify and authenticate users based on accelerometer data. They used a dataset of 36 users, labeled according to activities such as walking, jogging, and climbing stairs. These labels were used as context and the authors presented analysis with and without these labels. For feature extraction, the authors divided the 3 axes readings of the accelerometer into windows of 10-seconds, and for each window they extracted features such as mean, standard deviation, resultant, and binned distribution. For identification, the authors performed a 36-class classification, whereas for the task of authentication, the authors reduced the problem to a 2-class problem. They achieved a classification accuracy of 72.2% for 10-second windows. While they concluded based on their results that it is not critical to know what activity the user is performing, their dataset was generated by users repeating a limited set of predefined activities.

Shi et al. [SNJ+2011] presented an approach that was built on the concept that most users are habitual in nature and are prone to performing similar tasks at a certain time of the day. The researchers collected a wide range of behavioral information such as location, communication, and usage of applications, in order to create a user profile. Their method is based on identification of positive events and boosting the authentication score when a \good" or habitual event is observed. The passage of time is treated as a negative event in that scores gradually degrade.

Riva et al. [RSL2012] presented an architecture that grants users access to any content on the device only when the authentication system evaluates the device operator's level of authenticity to be higher than what is required to access that content. Their system utilized face and voice recognition, location familiarity, and determining possession by sensing nearby electronic objects as signals to establish the legitimate user's level of authenticity. They motivated their work with a user study that explored models where there are at-least 3 levels of security: public, private, and confidential. With this framework, they tested nine users, and were able to reduce the number of explicit authentications by 42%.

Zhang et al. [ZTQ+2010] presented an automated power model construction technique that uses built-in battery voltage sensors and knowledge of battery discharge behavior to monitor power consumption of each application on an electronic device. They achieved an absolute average error rate of less than 10%.

Murmuria et al. [MMS+2012] demonstrated that the power consumption by individual device drivers on a smartphone varies by state of operation of that particular device driver. Shye et al. [SSM2009] presented a power estimation model by leveraging real user behavior. They presented evidence that system power consumption patterns are highly correlated with user behavior patterns, but stopped short of trying to profile users on this basis. Finally, Murmuria et al. [MSB+2015], succeeded in proposing a continuous user monitoring using a machine learning based approach comprising of an ensemble of three distinct modalities: power consumption, touch gestures, and physical movement. They were able to verify that their system is functional in real-time while the end-user was utilizing popular mobile applications.

Seo et al. [SKK2012], proposed a specially designed biometric identification method for intelligent mobile devices by analyzing the user’s input patterns, such as a finger’s touch duration, pressure level and the touching width of the finger on the touch screen. They collected the input pattern data of individuals to empirically test their method. Their testing results show that this method effectively identifies users with near a 100% rate of accuracy.

Saevanee et al. [SCF2008], investigated three behavioral biometric techniques based on SMS texting activities and messages, looking to apply these techniques as a multi-modal biometric authentication method for mobile devices. The results showed that behavior profiling, keystroke dynamics and linguistic profiling can be used to discriminate users with overall error rates 20%, 20% and 22% respectively. To study the feasibility of multi-modal behavior biometric authentication system, matching-level fusion methods were applied. Two fusion methods were utilized: simple sum and weight average. The results showed clearly that matching-level fusion can improve the classification performance with an overall EER 8%.

2.4 Additional Results

Study	**Context**	**Methodology**	**Participants**
Clarke et al. [CFR+2002]	Acceptance of Subscriber Authentication Methods For Mobile Telephony Devices	Survey	Users
Clarke et al. [CF2005]	Authentication of users on mobile telephones – A survey of attitudes and practices	survey	Users
Karatzouni et al. [KFC2007]	Perceptions of User Authentication on Mobile Devices	survey	Users
Jones [JH2012]	Do Business Students Practice Smartphone Security?	survey	Business Students

In 2002, Clarke et al. [CFR+2002] presented the findings of a survey concerning the opinions of subscribers regarding the need for security in mobile devices, their use of current methods, and their attitudes towards alternative approaches that could be employed in the future. Surveyed users responded positively towards alternative methods of authentication, such as fingerprint scanning and voice verification.

On a survey of 297 mobile subscribers, conducted by Clarke et al. [CF2005], they attempted to assess the use of mobile devices, the use of current authentication methods, and the attitudes towards future security options. The findings revealed that the majority of the respondents make significant use of their devices, with clear demands for protection against unauthorized use. However, the use of current PIN-based authentication was marked as problematic, with a third of the respondents indicating that they do not use it at all, and other problems being reported amongst those that do. In view of this, the respondents' opinions in relation to future security options are interesting, with 83% being willing to accept some form of biometric authentication on their device.

Karatzouni et al. [KFC2007] examined four research questions: whether users recognize a need for security on their current devices; how they perceive the current authentication facilities, and whether they use them; whether they envisage a need for greater security provision in the future; and their perceptions of alternative authentication methods and the ways in which they could operate. The overall results showed that users envisage a need for enhanced security as their usage of the device changes to incorporate more sensitive functions. Furthermore, from the options discussion, a preference towards the use of biometric authentication was expressed by the majority of the participants.

While intentional misuse of data is a concern, Muslukhov et al. [MBK+2013] showed that users are also concerned about sharing mobile phones with guest users.

Jones [JH2012] presented a survey with the topic: Do Business Students Practice Smartphone Security? The purpose of this study was to investigate the degree to which business students practice smartphone security. A survey of security-related practices was administered to students in business classes at a regional public university. The results of the survey showed students to be lax in their smartphone security with men more willing to engage in risky behaviors than women. There were no differences in behaviors based upon maturity level or use of smartphones for financial transactions.

2.5 Results

Jones [JH2012] presented a survey under the topic: Do Business Students Practice Smartphone Security? The results of the survey showed students to be lax in their smartphone security with men more willing to engage in risky behaviors than women. The main limitation to this research is that the generalizability of the study is limited because the subject pool only included students in business classes at one university. In the present thesis we examine a similar survey but on members of the academic faculty of two universities (Athens and Aegean).

Shi et al. [SNJC2011] presented an approach that was built on the concept that most users are habitual in nature and are prone to performing similar tasks at a certain time of the day. The researchers collected a wide range of behavioral information such as location, communication, and usage of applications, in order to create a user profile. Their method is based on identification of positive events and boosting the authentication score when a \good" or habitual event is observed. The passage of time is treated as a negative event in that scores gradually degrade. One of the main caveats with this work is that it is trying to model what good geographic locations, phone calls, text messages, and website urls are. The data collected is highly intrusive in terms of privacy. They further modeled all good events as ones that are expected to be performed at a certain time of day, which is an assumption of habit that is not proven in the literature. What we want to see via our questionnaire is if users do perform similar tasks at a certain time of the day. In addition, through this approach we want to examine under what basis the user's profile can be created.

In the literature review we studied certain surveys which support that the password is not sufficient for the protection of mobile devices [CF2005], [AEG+2007], [KS2010], [CFS+2012], [KTH+2013]. Aviv et al. [AGM+2010] proved that mobile devices are vulnerable to smudge attacks. Finally, Clarke et al. [CFR+2002], Clarke et al. [CF2005], Karatzouni et al. [KFC2007] showed that users are willing to adopt alternative methods of authentication such as biometrics in order to protect their privacy on their devices.

3

Background of the study

In this chapter we will refer to the theoretical part that relates to CA and we will make a small recursion. In addition, we will briefly mention the techniques used in this thesis while their understanding is necessary for the reader prior to the presentation.

3.1 Biometric Characteristics

Biometrics can be categorized as either physiological (iris, fingerprint, DNA etc.) or behavioral (voice, gait, signature etc.).

Fig. 1. Morphological and Behavioral biometrics [B2015].

3.1.1 Morphological biometrics

Morphological biometrics are used in order to identify and/or verify a person by using one or more anatomical or biological characteristics, including (but not restricted to): fingerprints, palm prints, hand geometry, face, iris, retina, DNA.

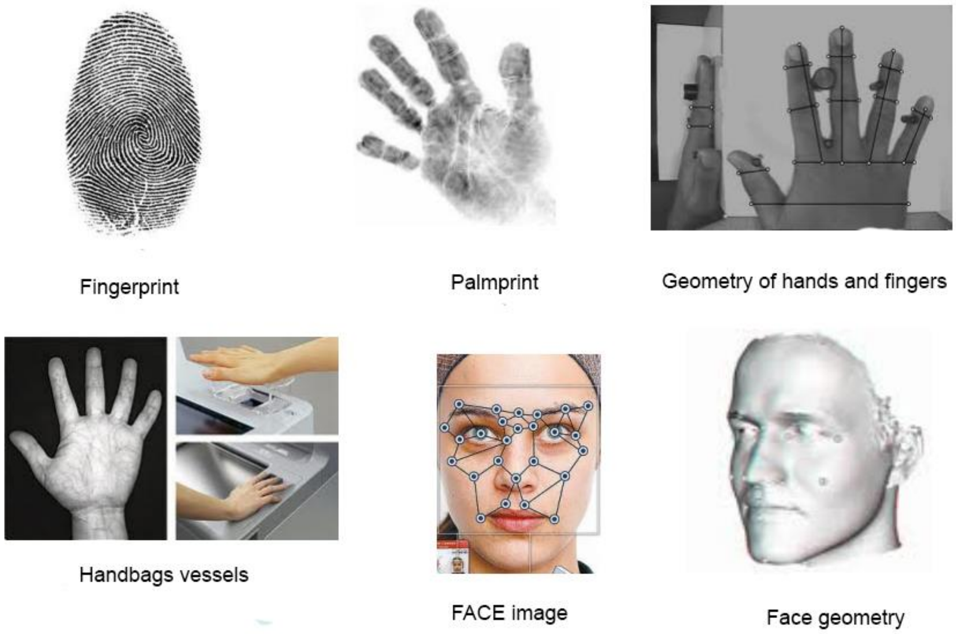


Fig. 2. Morphological biometrics [B2015].

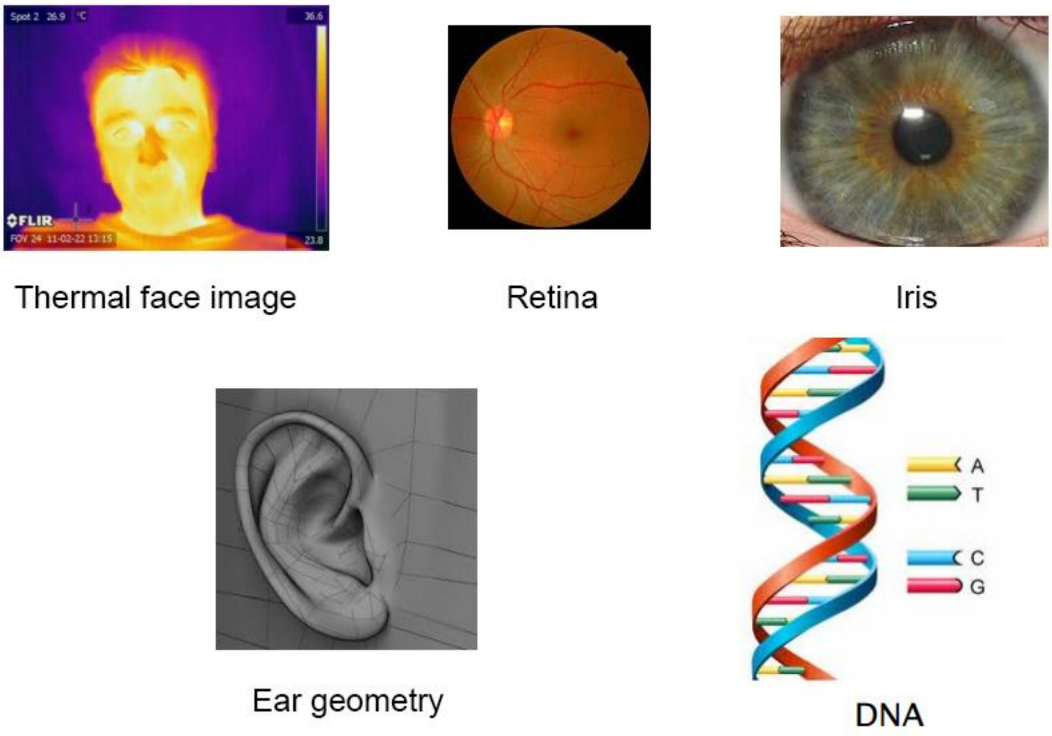


Fig. 3. Morphological biometrics [B2015].

3.1.2 Behavioral Biometrics

Behavioral biometrics use the behaviors of a person, which are characteristic, and are learned and acquired over time, in order to identify him. They can include: voice, signature, gait, keystroke.

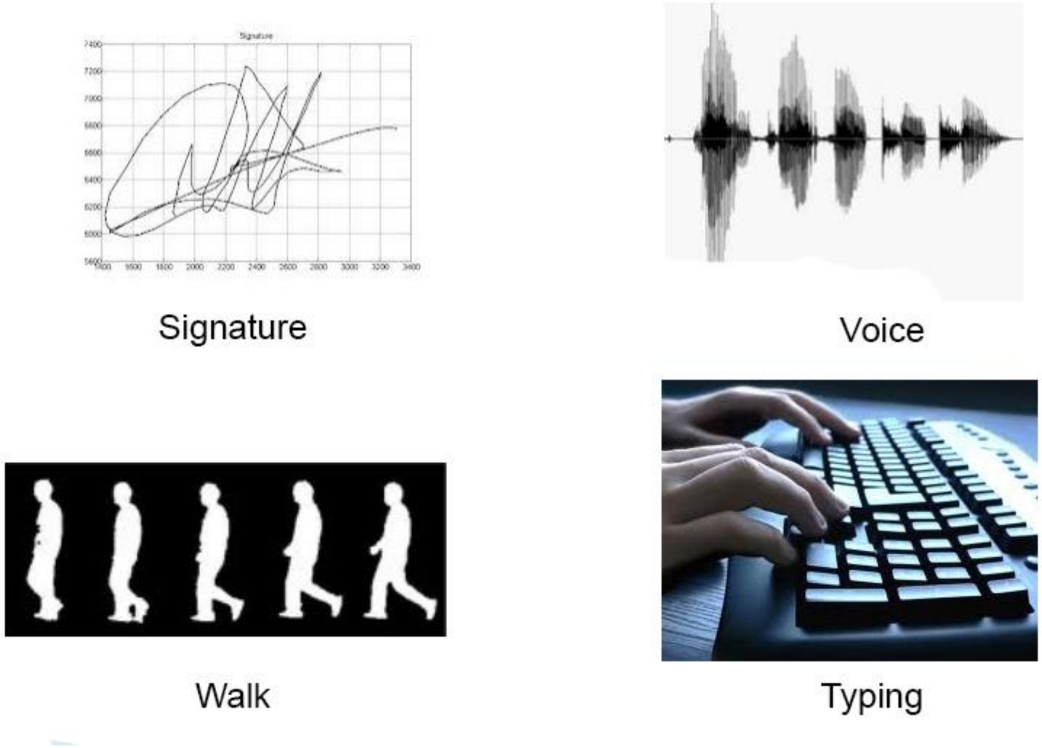
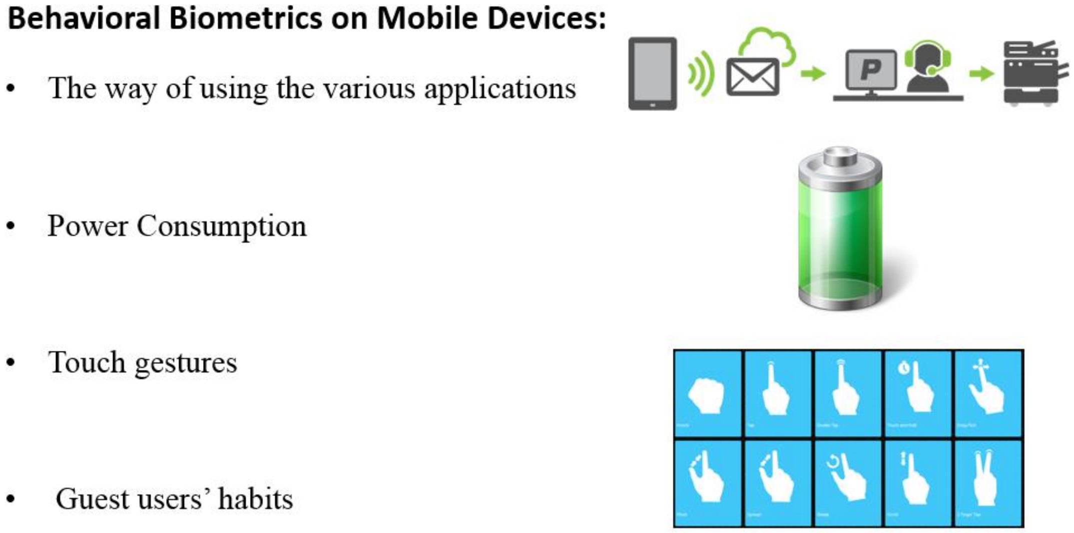


Fig. 4. Behavioral biometrics [B2015].


Behavioral Biometrics on Mobile Devices:

- The way of using the various applications
- Power Consumption
- Touch gestures
- Guest users' habits

Fig. 5. Behavioral biometrics

3.1.3 Assessment of Biometric Characteristics

According to Jain et al [JBP2006], a biometric system can be assessed by assessing the following properties of the Morphological or Behavioral characteristic on which it is based.

- **Universality**, which measures the degree to which the characteristic can be found in the majority of people.
- **Uniqueness**, which measures the degree to which the characteristic is unique among different people.
- **Permanence**, which measures the characteristic's resistance to change due to advancing age, illness and/or accidents.
- **Collectability**, which measures how easy and convenient it is to capture and measure the characteristic.
- **Performance**, which measures factors such as the speed and accuracy of the capturing of the characteristic.
- **Acceptability**, which measures peoples' willingness to accept a biometric system based on that characteristic.
- **Circumvention**, which measures how easy it is to use fraudulent techniques in order to fool a biometric system based on that characteristic.

It is important to emphasize here that there is no single answer as to how suitable a characteristic is for a biometric system. Each characteristic has different attributes and it should be assessed regarding the context and the application of the biometric system to be built. A characteristic that seems to be a poor candidate for one biometric system could be an excellent candidate for some other system. Following is a short description of the main biometric characteristics, as well as their assessment as to the degree to which each characteristic could be suitable for use in a biometrics Device Centric Authentication (DCA) system.

- **Fingerprint Recognition**: Fingerprints have been used as an identification method for many centuries, although they were first studied on a scientific basis in 1892 by Francis Galton [F1892]. Initially extracted by creating ink impressions on paper, fingerprint patterns are nowadays captured by fingerprint sensors, based on various technologies (optical, thermal, CMOS, ultrasonic, etc.). Fingerprints are unique to each person (even to identical twins) and they don't change due to advancing age, although they can be temporarily or permanently damaged due to accidents (burned, cut, etc.). Fingerprint scanners are very accurate and inexpensive and have recently been embedded in many mobile devices, thus making fingerprint recognition the most popular and publicly accepted biometrics authentication method. Fingerprint scanning will probably become an even more integral part of our lives, since over 50% of the smartphones are expected to have a fingerprint sensor by 2019 [RC2015]. Fingerprint recognition is a very easy (the user only has to use one thumb) and very fast (usually it takes less than a second) authentication method. Finally, although it is possible to fool a fingerprint verification system by using forged fingerprints, the required techniques are difficult and time consuming and newest scanners become less resilient to fraud. For all those reasons, we consider fingerprint recognition as the currently ideal method for the implementation of a biometrics DCA system.

- **Palm Print Recognition**: Palm print recognition is similar to the fingerprint recognition and it uses a pattern that typically comes from the butt of the palm and that contains lines, wrinkles and epidermal ridges. Palm prints are unique and universal and they don’t change over time. Although palm print recognition is widely used by police and forensics across the world for the identification of criminal subjects, its use as a system authentication method is extremely limited. This is mainly due to the fact that palm print scanners need to capture a larger area and are more expensive. As with fingerprints, palm prints are very difficult, although not impossible, to be forged. Although palm prints can be highly rated in most of the assessment properties, the big size of the scanners is a major hindrance to their use in a biometrics DCA system, especially when fingerprints provide similar functionality at a lower cost and with higher user acceptance.
- **Hand Geometry Recognition**: Hand geometry is the first and longest implemented biometrics authentication method, since hand geometry readers debuted in the market in the mid-1980s and have since been used, mainly for physical access control as well as for time and attendance records. Hand geometry readers measure the shape of a person’s hand and they have been installed at the entrances of nuclear power plants, restricted areas in airports, amusement parks (e.g. Disneyland), they were even used for the athletes’ entrance to the Olympic Village during the 1996 Olympic Games in Atlanta [GAO2002]. Hand geometry is a universal characteristic but it is not highly unique, comparing to other physiological characteristics. Although the hand geometry doesn’t usually change during an adult’s life, big changes occur during the growth period of children and it can be affected by accidents and certain illnesses, such as arthritis. Hand geometry systems are fast, easy to use and among the most acceptable verification systems. However, the required readers are big in size, and therefore cannot be embedded into mobile devices. Overall, although hand geometry recognition can have many practical applications, it is a very poor candidate when it comes to a biometrics DCA system.

- **Face Detection & Recognition**: Face detection and recognition systems use specialized algorithms in order to compare facial features from a subject's freshly taken photograph to those from an archived template. Face recognition is among the least reliable and effective biometric verification methods, since it can perform poorly under certain conditions, such as poor lightning, not clear or low resolution images, not neutral facial expressions, bizarre angles, dark skin colors etc. In addition, it is a method that can be forged relatively easily, by using someone else's printed photo or by playing a recorded video (face spoofing attacks) [EM2013]. Moreover, the identification / verification process becomes much more difficult when it comes to identical twins [SE2011]. On the other hand, face detection's main advantage is that it relies on a piece of H/W available in most modern mobile devices: a camera with some decent resolution. Therefore, face detection could play a role in a biometrics DCA system, but only supplementary, as an additional biometrics authentication factor.
- **Iris Recognition**: Iris recognition systems apply pattern recognition and analysis to images of a person's irides (the annular region of the eye bounded by the pupil and the sclera - white of the eye), in order to verify the identity of that person. Not only is iris a universal characteristic but also in 1985, ophthalmologists Leonard Flom and Aran Safi proposed that irides are unique for each person, and were later awarded a patent for the iris identification concept. In fact, irides are indeed unique, even those of identical twins. Furthermore, the iris doesn't change over time and is a very well protected internal organ. Iris recognition requires only a small sensor which could easily be embedded in any mobile device and the iris scanning can happen from a distance, making it a fairly acceptable verification method. As with fingerprint scanning, iris scanning can be a very fast and easy authentication method. On the other hand, iris scanners are still rather expensive and only recently did appear the first mobile devices with embedded sensors. As the technology evolves and iris scanners become more widely accessible, iris recognition is likely to become a very popular and effective biometric DCA system.
- **Retinal Scan**: Retinal scan is another eye recognition method, with many similarities to the iris recognition. It is universal, highly unique (even between both eyes of the same person) and is very hard to change or replicate. Retinal scanners can be easily embedded into any mobile device and are very accurate. However, retinal scanning requires the user to peep into the scanner's eye-piece, which makes the process somewhat inconvenient and hinders its acceptability among users. Overall, both iris and retinal scans make very good candidates for a biometrics DCA system, with retinal scans being more accurate but having lower user acceptance.

- **DNA Recognition**: Deoxyribonucleic Acid (DNA) is the most universal, unique and permanent biometric feature. However, and despite its wide recent use in forensics for the identification of criminals, DNA has serious drawbacks when it comes to user verification especially for a DCA system. First and foremost, there is currently no way to apply automatic and real-time verification using DNA, since DNA analysis requires various chemical methods and an expert's skills. For that reason, DNA is the characteristic with the lowest collectability among all the biometrics. Moreover, most users would be reluctant to provide their DNA, mainly due to privacy issues, therefore DNA recognition has a very low acceptability. Finally, DNA recognition has a very high circumvention, since it is fairly easy to steal a piece of DNA from a person and use it to authenticate as that person.

3.2 Continuous Authentication

In order to address the shortcomings of the entry-point authentication model, one of the approaches proposed in literature is called continuous authentication [CF2007]. Hereinafter, we will see a CA approach as presented in the book: "Continuous Authentication Using Biometrics: Data, Models and Metrics" [AT2011]. Continuous Authentication (CA) systems represent a new generation of security mechanisms that continuously monitor user behavior and use this as basis to re-authenticate periodically throughout a login session. The idea of continuous authentication emerged in the early 2000s, in part due to heightened security concerns brought about after 9/7. Interest in this technology has been increasing since then, both in academia and industry.

Continuous authentication represent a subclass of activity monitoring. The field of activity monitoring was originally investigated by Fawcett and Provost (1999) as a new class of Knowledge and Data Discovery problems (KDD), which consist of observing the behavior of a large number of entities or individuals with the purpose of detecting unusual events occurring requiring immediate actions. Activity monitoring applications greatly vary in terms of the kinds of data streams involved. Nonetheless, Fawcett and Provost have attempted in their study to provide a general and common representation for activity monitoring tasks. These tasks vary from fraud detection to intrusion detection, or news story monitoring systems. As a subclass of activity monitoring the field of application of CA is narrower and broadly fall under the category of intrusion detection.

3.2.1 Static vs. Continuous Authentication

Static authentication is a binary decision process consisting of three sub-processes: enrollment, presentation and evaluation (see figure 1). During the enrollment sub-process information is collected about the individual, processed and stored as a template or profile to be used subsequently as basis for authentication. The presentation sub-process is executed when an individual wants to use the system. The evaluation sub-process which is then triggered consists of comparing the presented authentication information against the stored profile for the claimed identity. The outcome of this process will be a match or non-match.

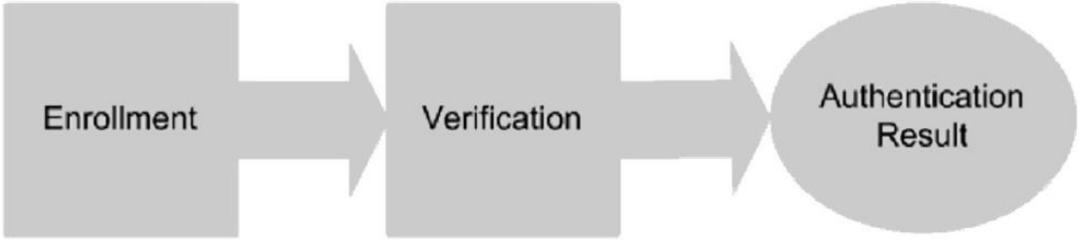


Figure 1. Static authentication process [AT2001].

Continuous authentication is a mechanism that checks the identity of an individual repeatedly for the entire duration of an authorized session. Static authentication provides assurance of the individual's identity only at the point of entry of a session. As the session progresses, assurance that the individual is who he claims to be can be given only through CA process. The CA process dynamically iterates the three steps involved in the static authentication process repeatedly throughout the session (see figure 2). Iterations can be performed randomly or at fixed time interval, or according to the occurrence of specific events.

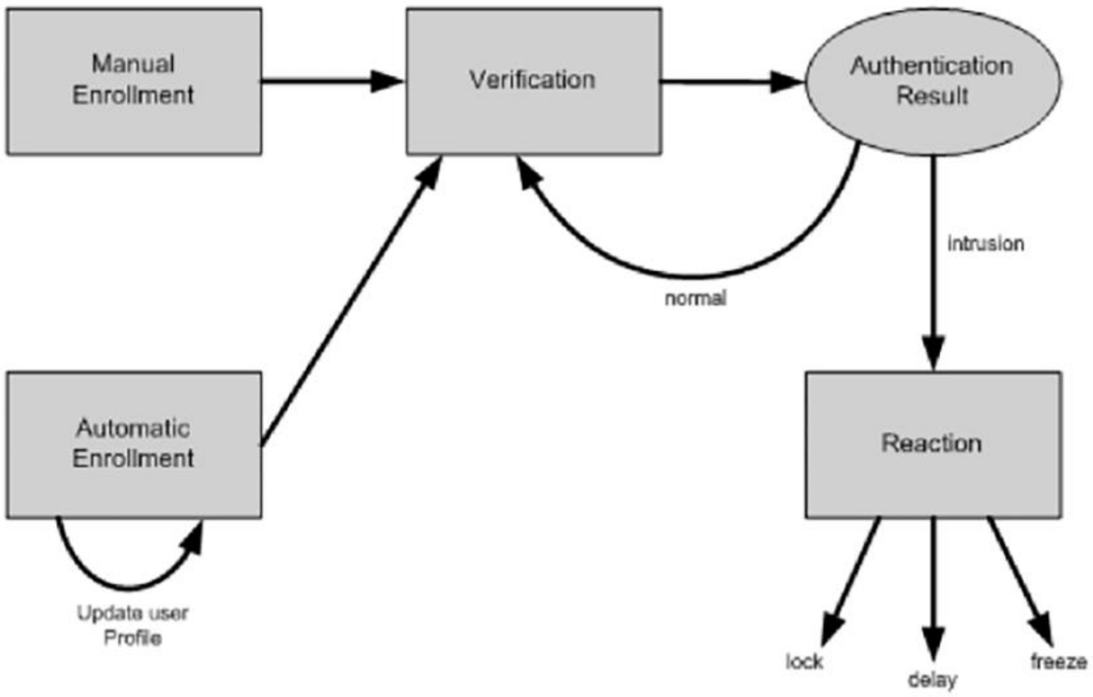


Figure 2. Continuous authentication process [AT2001].

Establishing an accurate user profile is a key prerequisite for successful continuous authentication. User profile in globally distributed networked environments may involve user knowledge and characteristics, access location, job characteristics, recourses used, workstations and transaction profiles. A key challenge is that the user profiles may be subject to constant changes over time in networked environments. This is referred to as behavior drift and may be dealt with using appropriate artificial intelligence techniques.

3.2.2 CA Entities

A CA system can be characterized primarily by two major entities: the sensor which is linked to a data source and the controller which implements the underlying data processing scheme. A typical CA system may involve one or several data sensor/ controller pair. Although the data source needs not to be a biometric, it is expected that it should have strong discriminative capability. So ideally, biometric data sources would be more appropriate.

Desirable characteristics for the data processing component include adaptive learning and the capability to handle behavior drift, and noisy and incomplete data and so on. Since such characteristics are typically in artificial intelligence (AI) techniques, we use a terminology reminiscer to AI to characterize CA data processing schemes. More specifically, we categorize broadly CA data processing schemes as either supervised or unsupervised.

The data processing scheme depends on the particular type of data (i.e. keystroke, voice, mouse fingerprint, etc.), but in any case, it must allow extracting from the data a profile for the user that uniquely characterizes his behavior. In the supervised model, the derivation of a user profile requires using sample data from both the individual (self) and other people (nonself), while in the unsupervised model only sample data from self is needed when building a user profile.

Using a supervised or unsupervised model will have a significant impact on the scope of the CA system. A supervised model may be used to discriminate between users in a closed setting, where CA data can be controlled for all the users. The main weakness of a supervised approach, however, is that all users' data must be collected before CA activity monitoring can be proceed and even so the approach may be hindered by the non-uniform class problem as the number of classes of users increases. When public access to hosts is not restricted, as in the case in many operational environments, the unsupervised model is more suitable for the CA process. In this case, we do not need the impostor's profile a priori in order to detect him. A normal profile is built for each authorized user during enrollment and compared against a current behavior to establish whether such behavior is genuine or intrusive.

3.2.3 CA Phases

A CA system can also be characterized in terms of the major phases involved in the CA process. There are two major phases in a typical CA process: enrollment and monitoring (see figure 3).

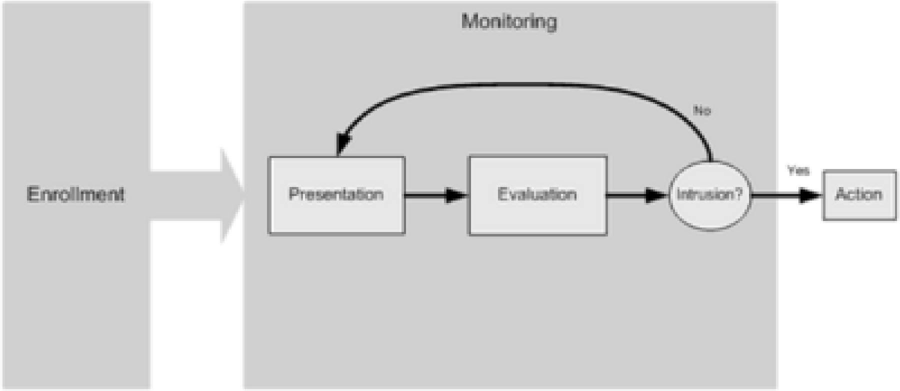


Figure 3. Continuous authentication phases [AT2001].

3.2.4 Enrollment Phase

The enrollment always precedes the monitoring phase. It is a critical phase during which individual user profiles also referred to as signatures are built from sample data collected from the user. The key questions that need to be answered prior or during this phase are the following:

1. What is the minimum amount of data needed to enroll a user?
2. How sound are the enrollment samples?
3. Do the enrollment process and sample collections require active participation of the user or are those transparent?

3.2.5 Monitoring Phase

The monitoring phase relies on the outcome of the enrollment phase to carry out the actual function of the CA activity. At the beginning of this phase, the user claims specific identity, for instance, by providing some user identification and/ or password. The profile of the claimed identity is considered the reference profile. The monitoring phase simply consists of comparing on a regular basis the monitored sample or data (received from the user) against the reference profile. In case of a non-match, an intrusion is reported, otherwise, the user behavior is considered as normal. The following two important questions need to be answered for this phase:

1. What is the length of the verification period?
2. Do the verification process and sample collections require active participation of the user or are those transparent?

3.3 Statistical Methods

3.3.1 Kolmogorov–Smirnov Test

In statistics, the Kolmogorov–Smirnov test (K–S test or KS test) is a nonparametric test of the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution (one-sample K–S test), or to compare two samples (two-sample K–S test). The Kolmogorov–Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples. The null distribution of this statistic is calculated under the null hypothesis that the samples are drawn from the same distribution (in the two-sample case) or that the sample is drawn from the reference distribution (in the one-sample case). In each case, the distributions considered under the null hypothesis are continuous distributions but are otherwise unrestricted.

The two-sample K–S test is one of the most useful and general nonparametric methods for comparing two samples, as it is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples.

The Kolmogorov–Smirnov test can be modified to serve as a goodness of fit test. In the special case of testing for normality of the distribution, samples are standardized and compared with a standard normal distribution. This is equivalent to setting the mean and variance of the reference distribution equal to the sample estimates, and it is known that using these to define the specific reference distribution changes the null distribution of the test statistic: see below. Various studies have found that, even in this corrected form, the test is less powerful for testing normality than the Shapiro–Wilk test or Anderson–Darling test [S1974]. However, other tests have their own disadvantages. For instance the Shapiro–Wilk test is known not to work well with many ties (many identical values).

3.3.1.1 Kolmogorov–Smirnov statistic

The empirical distribution function F_n for n iid observations X_i is defined as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{[-\infty,x]}(X_i)$$

where $I_{[-\infty,x]}(X_i)$ is the indicator function, equal to 1 if $X_i \leq x$ and equal to 0 otherwise.

The Kolmogorov–Smirnov statistic for a given cumulative distribution function $F(x)$ is

$$D_n = \sup_x |F_n(x) - F(x)|$$

where *sup* $_x$ is the supremum of the set of distances. By the Glivenko–Cantelli theorem, if the sample comes from distribution $F(x)$, then D_n converges to 0 almost surely in the limit when n goes to infinity. Kolmogorov strengthened this result, by effectively providing the rate of this convergence (see below). Donsker's theorem provides yet a stronger result.

In practice, the statistic requires a relatively large number of data points to properly reject the null hypothesis.

3.3.1.2 Kolmogorov distribution

The Kolmogorov distribution is the distribution of the random variable

$$K = \sup_{t\in[0,1]} |B(t)|$$

where *B*(*t*) is the Brownian bridge. The cumulative distribution function of *K* is given by [MTW2003].

$$\Pr(K \le x) = 1 - 2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2x^2} = \frac{\sqrt{2\pi}}{x}\sum_{k=1}^{\infty}e^{-(2k-1)^2\pi^2/(8x^2)}.$$

Both the form of the Kolmogorov–Smirnov test statistic and its asymptotic distribution under the null hypothesis were published by Andrey Kolmogorov [*K1933]*, while a table of the distribution was published by Nikolai Vasilyevich Smirnov [S1948]. Recurrence relations for the distribution of the test statistic in finite samples are available [MTW2003].

Under null hypothesis that the sample comes from the hypothesized distribution *F*(*x*),

$$\sqrt{n}D_n \xrightarrow{n\to\infty} \sup_t |B(F(t))|$$

in distribution, where *B*(*t*) is the Brownian bridge.

If *F* is continuous then under the null hypothesis $\sqrt{n}D_n$ converges to the Kolmogorov distribution, which does not depend on *F*. This result may also be known as the Kolmogorov theorem.

The *goodness-of-fit* test or the Kolmogorov–Smirnov test is constructed by using the critical values of the Kolmogorov distribution. The null hypothesis is rejected at level α if

$$\sqrt{n}D_n > K_\alpha,$$

where K_α is found from

$$\Pr(K \le K_\alpha) = 1 - \alpha.$$

The asymptotic power of this test is 1.

3.3.1.3 Test with estimated parameters

If either the form or the parameters of *F*(*x*) are determined from the data X_i the critical values determined in this way are invalid. In such cases, Monte Carlo or other methods may be required, but tables have been prepared for some cases. Details for the required modifications to the test statistic and for the critical values for the normal distribution and the exponential distribution have been published, [PH1972] and later publications also include the Gumbel distribution [SW1986]. The Lilliefors test represents a special case of this for the normal distribution. The logarithm transformation may help to overcome cases where the Kolmogorov test data does not seem to fit the assumption that it came from the normal distribution.

3.3.1.4 Discrete null distribution

The Kolmogorov–Smirnov test must be adapted for discrete variables [AE2011]. The form of the test statistic remains the same as in the continuous case, but the calculation of its value is more subtle. We can see this if we consider computing the test statistic between a continuous distribution $f(x)$ and a step function $g(x)$ that has a discontinuity at x_i. In other words, the limit $\lim_{x \to x_i} g(x)$, if it exists, is different from $g(x_i)$. Thus, when computing the statistic

$$\sup_x |g(x) - f(x)| = \max_i \left[\max \left(|g(x_i) - f(x_i)|, \lim_{x \to x_i} |g(x) - f(x_{i-1})| \right) \right],$$

it is unclear how to replace the limit, unless we know the limiting value of the underlying distribution.

In SAS, the Kolmogorov–Smirnov test is implemented in PROC NPAR1WAY [SASSTATISTICS]. The discretized KS test is implemented in the ks.test() function in the dgof package of the R project for statistical computing [AE2011]. In Stata, the command ksmirnov performs a Kolmogorov–Smirnov test [STATA].

Two-sample Kolmogorov–Smirnov test

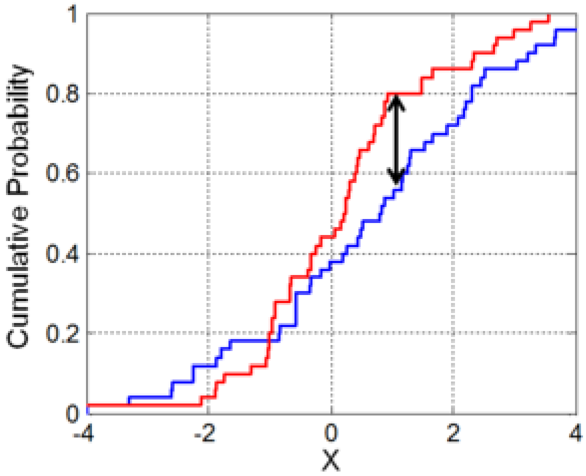


Illustration of the two-sample Kolmogorov–Smirnov statistic. Red and blue lines each correspond to an empirical distribution function, and the black arrow is the two-sample KS statistic.

The Kolmogorov–Smirnov test may also be used to test whether two underlying one-dimensional probability distributions differ. In this case, the Kolmogorov–Smirnov statistic is

$$D_{n,n'} = \sup_x |F_{1,n}(x) - F_{2,n'}(x)|,$$

where $F_{1,n}$ and $F_{2,n'}$ are the empirical distribution functions of the first and the second sample respectively, and sup is the supremum function.

The null hypothesis is rejected at level α if

$$D_{n,n'} > c(\alpha)\sqrt{\frac{n+n'}{nn'}}.$$

The value of $c(\alpha)$ is given in the table below for each level of α

α	0.10	0.05	0.025	0.01	0.005	0.001
$c(\alpha)$	1.22	1.36	1.48	1.63	1.73	1.95

Note that the two-sample test checks whether the two data samples come from the same distribution. This does not specify what that common distribution is (e.g. whether it's normal or not normal). Again, tables of critical values have been published [PH1972]. These critical values have one thing in common with the Anderson–Darling and Chi-squares, namely the fact that higher values tend to be more rare [M2014].

1. Setting confidence limits for the shape of a distribution function

While the Kolmogorov–Smirnov test is usually used to test whether a given $F(x)$ is the underlying probability distribution of $F_n(x)$, the procedure may be inverted to give confidence limits on $F(x)$ itself. If one chooses a critical value of the test statistic D_α such that $P(D_n > D_\alpha) = \alpha$, then a band of width $\pm D_\alpha$ around $F_n(x)$ will entirely contain $F(x)$ with probability $1 - \alpha$.

2. The Kolmogorov–Smirnov statistic in more than one dimension

A distribution-free multivariate Kolmogorov–Smirnov goodness of fit test has been proposed by Justel, Peña and Zamar [JPZ1997]. The test uses a statistic which is built using Rosenblatt's transformation, and an algorithm is developed to compute it in the bivariate case. An approximate test that can be easily computed in any dimension is also presented.

The Kolmogorov–Smirnov test statistic needs to be modified if a similar test is to be applied to multivariate data. This is not straightforward because the maximum difference between two joint cumulative distribution functions is not generally the same as the maximum difference of any of the complementary distribution functions. Thus the maximum difference will differ depending on which of $\Pr(x < X \wedge y < Y)$ or $\Pr(X < x \wedge Y > y)$ or any of the other two possible arrangements is used. One might require that the result of the test used should not depend on which choice is made.

One approach to generalizing the Kolmogorov–Smirnov statistic to higher dimensions which meets the above concern is to compare the cdfs of the two samples with all possible orderings, and take the largest of the set of resulting K–S statistics. In d dimensions, there are 2^d-1 such orderings. One such variation is due to Peacock [P1983] and another to Fasano and Franceschini [FF1987] (see Lopes et al. [LRH2007] for a comparison and computational details). Critical values for the test statistic can be obtained by simulations, but depend on the dependence structure in the joint distribution.

3.3.2 Levene's test

In statistics, Levene's test is an inferential statistic used to assess the equality of variances for a variable calculated for two or more groups [LH60]. Some common statistical procedures assume that variances of the populations from which different samples are drawn are equal. Levene's test assesses this assumption. It tests the null hypothesis that the population variances are equal (called *homogeneity of variance* or *homoscedasticity*). If the resulting *p*-value of Levene's test is less than some significance level (typically 0.05), the obtained differences in sample variances are unlikely to have occurred based on random sampling from a population with equal variances. Thus, the null hypothesis of equal variances is rejected and it is concluded that there is a difference between the variances in the population.

Some of the procedures typically assuming homoscedasticity, for which one can use Levene's tests, include analysis of variance and t-tests.

Levene's test is often used before a comparison of means. When Levene's test shows significance, one should switch to more generalized tests that is free from homoscedasticity assumptions (sometimes even non-parametric tests).

Levene's test may also be used as a main test for answering a stand-alone question of whether two sub-samples in a given population have equal or different variances.

3.3.2.1 Definition

The test statistic, *W*, is defined as follows:

$$W = \frac{(N-k)}{(k-1)} \frac{\sum_{i=1}^{k} N_i (Z_{i\cdot} - Z_{\cdot\cdot})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - Z_{i\cdot})^2},$$

where

- W is the result of the test,
- k is the number of different groups to which the sampled cases belong,
- N is the total number of cases in all groups,
- N_i is the number of cases in the ith group,
- Y_{ij} is the value of the measured variable for the$j$th case from the ith group,

- $$Z_{ij} = \begin{cases} |Y_{ij} - \bar{Y}_{i\cdot}|, & \bar{Y}_{i\cdot} \text{ is a mean of i-th group} \\ |Y_{ij} - \tilde{Y}_{i\cdot}|, & \tilde{Y}_{i\cdot} \text{ is a median of i-th group} \end{cases}$$

(Both definitions are in use though the second one is, strictly speaking, the Brown–Forsythe test – see below for comparison)

$Z_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{N_i} Z_{ij}$ is the mean of all Z_{ij},

$Z_{i\cdot} = \frac{1}{N_i}\sum_{j=1}^{N_i} Z_{ij}$ is the mean of the Z_{ij} for group i.

The significance of W is tested against $F(\alpha, k-1, N-k)$ where F is a quantile of the F-test distribution, with $k-1$ and $N-k$ its degrees of freedom, and α is the chosen level of significance (usually 0.05 or 0.01).

3.3.3 Kruskal–Wallis one-way analysis of variance

The Kruskal–Wallis test by ranks, Kruskal–Wallis H test [LS2015] (named after William Kruskal and W. Allen Wallis), or One-way ANOVA on ranks is a non-parametric method for testing whether samples originate from the same distribution [KW1952], [CF2009], [SC1988]. It is used for comparing two or more independent samples of equal or different sample sizes. It extends the Mann–Whitney U test when there are more than two groups. The parametric equivalent of the Kruskal-Wallis test is the one-way analysis of variance (ANOVA). A significant Kruskal-Wallis test indicates that at least one sample stochastically dominates one other sample. The test does not identify where this stochastic dominance occurs or for how many pairs of groups stochastic dominance obtains. Dunn's test [D1964] would help analyze the specific sample pairs for stochastic dominance.

Since it is a non-parametric method, the Kruskal–Wallis test does not assume a normal distribution of the residuals, unlike the analogous one-way analysis of variance. If the researcher can make the less stringent assumptions of an identically shaped and scaled distribution for all groups, except for any difference in medians, then the null hypothesis is that the medians of all groups are equal, and the alternative hypothesis is that at least one population median of one group is different from the population median of at least one other group.

3.3.3.1 Method

Rank all data from all groups together; i.e., rank the data from 1 to N ignoring group membership. Assign any tied values the average of the ranks they would have received had they not been tied.

The test statistic is given by:

$$H = (N-1)\frac{\sum_{i=1}^{g} n_i(\bar{r}_{i\cdot} - \bar{r})^2}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(r_{ij} - \bar{r})^2},$$

where:

- n_i is the number of observations in group i
- r_{ij} is the rank (among all observations) of observation j from group i
- N is the total number of observations across all groups
- $\bar{r}_{i\cdot} = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}$,
- $\bar{r} = \tfrac{1}{2}(N+1)$ is the average of all the r_{ij}.

If the data contain no ties the denominator of the expression for H is exactly and $(N-1)N(N+1)/12$ $\bar{r} = \frac{N+1}{2}$. Thus

$$\begin{aligned} H &= \frac{12}{N(N+1)}\sum_{i=1}^{g} n_i\left(\bar{r}_{i\cdot} - \frac{N+1}{2}\right)^2 \\ &= \frac{12}{N(N+1)}\sum_{i=1}^{g} n_i\bar{r}_{i\cdot}^2 - \ 3(N+1). \end{aligned}$$

The last formula only contains the squares of the average ranks.

A correction for ties if using the short-cut formula described in the previous point can be made by dividing H by $1 - \frac{\sum_{i=1}^{G}(t_i^3 - t_i)}{N^3 - N}$, where G is the number of groupings of different tied ranks, and ti is the number of tied values within group i that are tied at a particular value. This correction usually makes little difference in the value of H unless there are a large number of ties.

Finally, the p-value is approximated by $\Pr(\chi^2_{g-1} \geq H)$. If some n_i values are small (i.e., less than 5) the probability distribution of H can be quite different from this chi-squared distribution. If a table of the chi-squared probability distribution is available, the critical value of chi-squared, $\chi^2_{\alpha:g-1}$, can be found by entering the table at g − 1 degrees of freedom and looking under the desired significance or alpha level.

If the statistic is not significant, then there is no evidence of stochastic dominance between the samples. However, if the test is significant then at least one sample stochastically dominates another sample. Therefore, a researcher might use sample contrasts between individual sample pairs, or post hoc tests using Dunn's test, which (1) properly employs the same rankings as the Kruskal-Wallis test, and (2) properly employs the pooled variance implied by the null hypothesis of the Kruskal-Wallis test in order to determine which of the sample pairs are significantly different [D1964]. When performing multiple sample contrasts or tests, the Type I error rate tends to become inflated, raising concerns about multiple comparisons.

3.3.3.2 Exact probability tables

A large amount of computing resources is required to compute exact probabilities for the Kruskal-Wallis test. Existing software only provides exact probabilities for sample sizes less than about 30 participants. These software programs rely on asymptotic approximation for larger sample sizes. Exact probability values for larger sample sizes are available. Spurrier 2003 published exact probability tables for samples as large as 45 participants [S2003]. Meyer and Seaman (2006) produced exact probability distributions for samples as large as 105 participants [MS2006].

3.3.4 Chi-Square Test (χ^2 test).

A chi-squared test, also referred to as χ^2 test (or chi-square test), is any statistical hypothesis test in which the sampling distribution of the test statistic is a chi-square distribution when the null hypothesis is true. Chi-squared tests are often constructed from a sum of squared errors, or through the sample variance. Test statistics that follow a chi-squared distribution arise from an assumption of independent normally distributed data, which is valid in many cases due to the central limit theorem. A chi-squared test can then be used to reject the hypothesis that the data are independent.

Also considered a chi-square test is a test in which this is asymptotically true, meaning that the sampling distribution (if the null hypothesis is true) can be made to approximate a chi-square distribution as closely as desired by making the sample size large enough. The chi-squared test is used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories. Does the number of individuals or objects that fall in each category differ significantly from the number you would expect? Is this difference between the expected and observed due to sampling variation, or is it a real difference?

The Chi-squared Statistic is a measure of how similar two categorical probability distributions are. If the two distributions are identical, the chi-squared statistic is 0, if the distributions are very different, some higher number will result. The formula for the chi-squared statistic is:

$$\chi^2(C,E) = \sum_{i=A}^{i=Z} \frac{(C_i - E_i)^2}{E_i}$$

where C_A is the count (not the probability) of letter A, and E_A is the expected count of letter A. This page will describe the use of the chi-squared statistic for cryptanalysis. Ordinarily, statisticians use the chi-squared statistic for measuring the goodness of fit of data. Unlike statisticians, we make no assumptions about the distribution of our data, and draw no conclusions about the significance of the result. We simply use the method to suggest a possible decryption.

3.3.4.1 Examples of chi-square tests with samples

One test statistic that follows a chi-square distribution exactly is the test that the variance of a normally distributed population has a given value based on a sample variance. Such tests are uncommon in practice because the true variance of the population is usually unknown. However, there are several statistical tests where the chi-square distribution is approximately valid:

3.3.4.2 Pearson's chi-square test

Pearson's chi-square test, also known as the chi-square goodness-of-fit test or chi-square test for independence. When the chi-square test is mentioned without any modifiers or without other precluding context, this test is often meant (for an exact test used in place of χ^2, see Fisher's exact test).

3.3.4.3 Yates's correction for continuity

Using the chi-square distribution to interpret Pearson's chi-square statistic requires one to assume that the discrete probability of observed binomial frequencies in the table can be approximated by the continuous chi-square distribution. This assumption is not quite correct, and introduces some error.

To reduce the error in approximation, Frank Yates suggested a correction for continuity that adjusts the formula for Pearson's chi-square test by subtracting 0.5 from the difference between each observed value and its expected value in a 2×2 contingency table. This reduces the chi-square value obtained and thus increases its p-value.

3.3.4.4 Other chi-square tests

- Cochran–Mantel–Haenszel chi-squared test.
- McNemar's test, used in certain 2×2 tables with pairing.
- Tukey's test of additivity.
- The portmanteau test in time-series analysis, testing for the presence of autocorrelation.
- Likelihood-ratio tests in general statistical modelling, for testing whether there is evidence of the need to move from a simple model to a more complicated one (where the simple model is nested within the complicated one).

3.3.4.5 Chi-squared test for variance in a normal population

If a sample of size *n* is taken from a population having a normal distribution, then there is a result which allows a test to be made of whether the variance of the population has a pre-determined value. For example, a manufacturing process might have been in stable condition for a long period, allowing a value for the variance to be determined essentially without error. Suppose that a variant of the process is being tested, giving rise to a small sample of *n* product items whose variation is to be tested. The test statistic *T* in this instance could be set to be the sum of squares about the sample mean, divided by the nominal value for the variance (i.e. the value to be tested as holding). Then T has a chi-square distribution with n–1 degrees of freedom. For example if the sample size is 21, the acceptance region for T for a significance level of 5% is the interval 9.59 to 34.17.

3.4 Conclusions

From the tests presented above we eventually ended up using the Kolmogorov–Smirnov and the Kruskal – Walis tests. We used the Kolmogorov–Smirnov test to see if we have a normal distribution. Since we don’t have a normal distribution, in the cases we examined, and we had more than two groups to check, we applied the non-parametric Kruskal – Walis test. We examine if there is a statistically significant difference between the age groups in correlation to the variables.

4

Results: "Users' Attitudes on Mobile Devices: Can the users' practices protect their sensitive data?"

4.1 Introduction

Presently, smartphones are the most popular devices. They are used to do just about everything. As the amount of the available services rise, the amount of sensitive data stored on or processed by handheld devices rise as well. This enables privacy risk threats as the users tend not to take the necessary measures to protect their privacy. Even though individuals are highly concerned about their privacy they often reveal personal information. The purpose of this research is to find the users' trends that relate to their Attitudes regarding to the ensurance of their privacy.

4.2 Problem Analysis

In the first survey we examine if the users adopt some basic practices to protect their sensitive personal data themselves or if there is a need to further strengthen their protection. To this end, we used an original questionnaire, created for the needs of the specific survey. For the statistical analysis, our main variable is *age* because we wanted to evaluate the significance degree regarding users' Attitudes and Practices between different age groups. The target group of the survey are 204: students, employees and members of University of Athens and University of the Aegean. Our survey was conducted using in-person delivery technique with a multiple-choice questionnaire. It consists of four subsections and is formed as follows:

- Demographics
- Storage Practices
- PIN Practices
- Device Protection

Our first survey answers two main research questions:

- What are the Users' Attitudes on Mobile Devices?
- Can the users' practices protect their sensitive data?

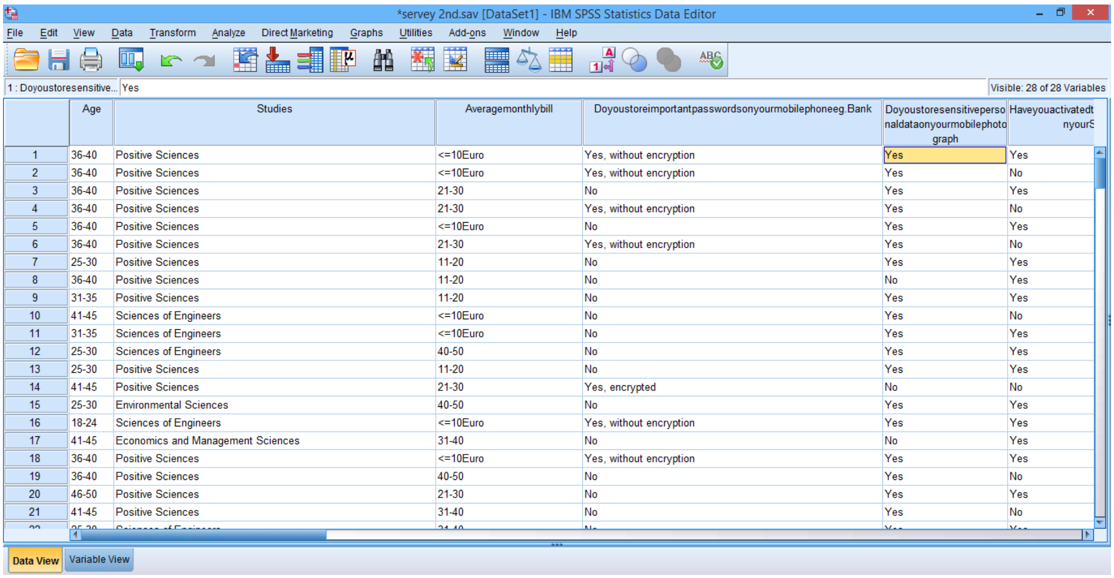
Finally, we seek the factors that influence the attitude of users with respect to their practices for the protection of personal data. We also search for the factors that affect the users' attitude in relation to the practices they follow so as to protect their personal data. To achieve this we searched, through statistical hypotheses, if the age and gender of users relate to their practices for the protection of their personal data.

Our first survey answers four main research hypotheses:

- First hypothesis: Does *age* correlate to users' practices concerning the storage of important passwords on their mobile phone?
- Second hypothesis: Does *age* correlate to the users' practices concerning the store sensitive personal data on their mobile (photographs / videos /voice recordings etc.)?
- Third hypothesis: Does *gender* correlate to the users' practices concerning the sharing of their PIN with third persons?

4.3 First Survey Features Encoding

In the first survey the data were collected with an original questionnaire, created for the needs of the specific survey. This questionnaire consists of 14 questions and the data were collected by members of two universities (Athens University and University of the Aegean). The results of the questionnaire were corresponded to variables and entered in an SPSS worksheet.


*servey 2nd.sav [DataSet1] - IBM SPSS Statistics Data Editor
File
Edit
View
Data
Transform
Analyze
Direct Marketing
Graphs
Utilities
Add-ons
Window
Help
1 : Doyoustoresensitive...
Yes
Visible: 28 of 28 Variables
Data View
Variable View

	Age	Studies	Averagemonthlybill	Doyoustoreimportantpasswordsonyourmobilephoneeg.Bank	Doyoustoresensitivepersonaldataonyourmobilephotograph	Haveyouactivatedtnyour$
1	36-40	Positive Sciences	<=10Euro	Yes, without encryption	Yes	Yes
2	36-40	Positive Sciences	<=10Euro	Yes, without encryption	Yes	No
3	36-40	Positive Sciences	21-30	No	Yes	Yes
4	36-40	Positive Sciences	21-30	Yes, without encryption	Yes	No
5	36-40	Positive Sciences	<=10Euro	No	Yes	Yes
6	36-40	Positive Sciences	21-30	Yes, without encryption	Yes	No
7	25-30	Positive Sciences	11-20	No	Yes	Yes
8	36-40	Positive Sciences	11-20	No	No	Yes
9	31-35	Positive Sciences	11-20	No	Yes	Yes
10	41-45	Sciences of Engineers	<=10Euro	No	Yes	No
11	31-35	Sciences of Engineers	<=10Euro	No	Yes	Yes
12	25-30	Sciences of Engineers	40-50	No	Yes	Yes
13	25-30	Positive Sciences	11-20	No	Yes	Yes
14	41-45	Positive Sciences	21-30	Yes, encrypted	No	No
15	25-30	Environmental Sciences	40-50	No	Yes	Yes
16	18-24	Sciences of Engineers	<=10Euro	Yes, without encryption	Yes	Yes
17	41-45	Economics and Management Sciences	31-40	No	No	Yes
18	36-40	Positive Sciences	<=10Euro	Yes, without encryption	Yes	Yes
19	36-40	Positive Sciences	40-50	No	Yes	No
20	46-50	Positive Sciences	21-30	No	Yes	Yes
21	41-45	Positive Sciences	31-40	No	Yes	No

Fig. 1. The SPSS worksheet before encoding.

4.4 Coding of data

The answers given by the respondents were encoded with numerical values and are presented in the following table.

Question	Variable	Coding

Gender	Sex	Male = 1
		Female = 2

Age	Age_groups	18-24 = 1 25-30 = 2 31-35 = 3 36-40 = 4 41-45 = 5 46-50 = 6
Average monthly bill	Avg_mnthly_bill	<=10Euro = 1 11-20 = 2 21-30 = 3 31-40 = 4 40-50 = 5

Question	Variable	Coding
Storage of important passwords (eg. Bank & alarm passwords) on mobile phone	Store_pwd_on_phone	No = 1 Yes, encrypted = 2 Yes, without encryption = 3

Storage of sensitive personal data on mobile (eg. Photographs / videos / voice recordings)	store_personal_data	No = 1
		Yes = 2

Activation of the PIN question on the SIM card	Pin_on_sim	No = 1
		Yes = 2

Existence of password on the Screen-Saver of mobile phone and frequency of change	Pwd_on_screen_saver	3 times a year = 1 I do not know if it has such an option = 2 More often = 3 Never = 4 Once a year = 5 The device does not have such an option = 6 Twice a year = 7

Protection of sensitive applications with a PIN or touch gesture	Protect_sensitive_app_with_pin_or_tg	No = 1
		Yes = 2

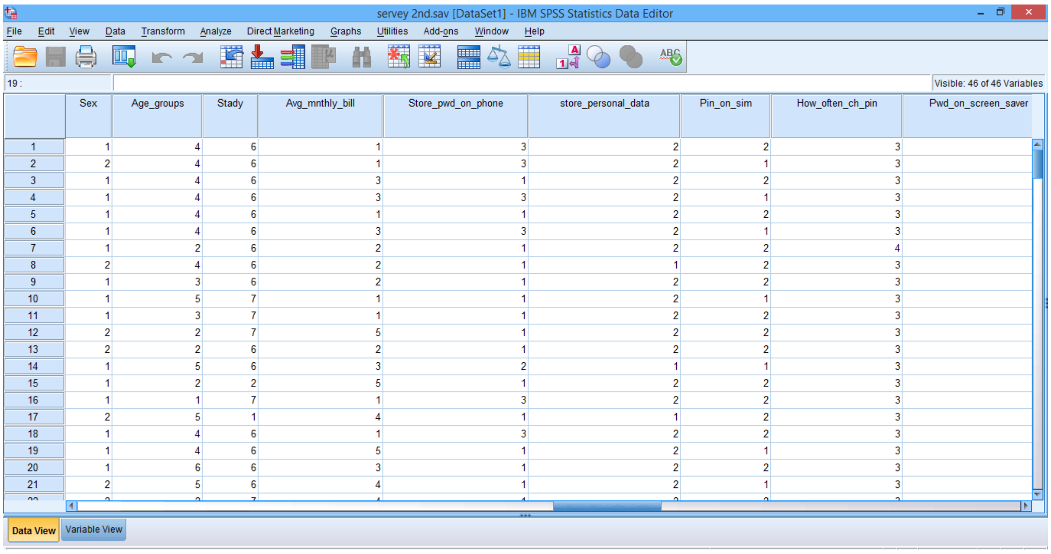
Question	Variable	Coding
Frequency of change of the PIN of the cash card	how_often_ch_pin_on_cashcard	3 times a year = 1 More often = 2 Never = 3 Once a year = 4 Twice a year = 5

Giving of pin to third persons	Given_your_pin	Yes = 1
		No = 2

Have you ever lost your device or has it ever been stolen	Stolen_your_device	Twice = 1
		Once = 2
		Never = 3
		More = 4
		3 times = 5

Have you ever left your device on a e.g. Coffee shop	left_your_device	Twice = 1
		Once = 2
		Never = 3
		More = 4
		3 times = 5

The SPSS worksheet after encoding:



	Sex	Age_groups	Stady	Avg_mnthly_bill	Store_pwd_on_phone	store_personal_data	Pin_on_sim	How_often_ch_pin	Pwd_on_screen_saver
1	1	4	6	1	3	2	2	3	
2	2	4	6	1	3	2	1	3	
3	1	4	6	3	1	2	2	3	
4	1	4	6	3	3	2	1	3	
5	1	4	6	1	1	2	2	3	
6	1	4	6	3	3	2	1	3	
7	1	2	6	2	1	2	2	4	
8	2	4	6	2	1	1	2	3	
9	1	3	6	2	1	2	2	3	
10	1	5	7	1	1	2	1	3	
11	1	3	7	1	1	2	2	3	
12	2	2	7	5	1	2	2	3	
13	2	2	6	2	1	2	2	3	
14	1	5	6	3	2	1	1	3	
15	1	2	2	5	1	2	2	3	
16	1	1	7	1	3	2	2	3	
17	2	5	1	4	1	1	2	3	
18	1	4	6	1	3	2	2	3	
19	1	4	6	5	1	2	1	3	
20	1	6	6	3	1	2	2	3	
21	2	5	6	4	1	2	1	3	

Fig. 2: The SPSS worksheet after coding with numerical values

4.5 Methodology

Our survey was conducted using in-person delivery technique, with a total of 204 participants that were requested to complete it anonymously and voluntarily. The target group of the survey is University of Aegean and University of Athens students, professors, and university members. We mostly choose members of the Academic faculty because they are more receptive to new technologies. They also understand better the technological evolution than externals.

A very useful evaluation method for surveying user's practices is the use of multiple-choice questionnaires [ACVS209]. This method was selected from other alternatives because it is more accurate and has a bigger degree of participation from the respondents.

The questionnaire is original and created for the needs of the specific survey.

It consists of six subsections and is formed as follows:

1. Demographics
2. Storage Practices
3. PIN Practices
4. Device Protection

We tried to formulate our questions in a fully understood way, in order to be answered and filled correctly. Also, 50% of the participants answered the questionnaire through an interview and the 50%, under instructions, via e-mail. The parts of the questionnaire follow a logical continuity and are clearly distinct, since we have used headings that indicate each group of questions.

For the statistical analysis, our main variable is *age* because we wanted to evaluate the significance degree regarding users' Attitudes and Practices between different age groups. Survey responses were analyzed using descriptive analysis, Crosstabs, Frequencies and Kruskal–Wallis test (p<.05) on SPSS.

4.6 Survey Results

Afterwards, the results are presented in full detail and an analysis and discussion of every issue is made:

4.6.1 Demographics

The participants were asked about their gender, age and field of studies. 55,9% of the participants where males and 44,1% where females.

Table 1. Gender

Gender	Frequency	Percent
1 Male	114	55,9%
2 Female	90	44,1%
Total	204	100,0%

Their ages are presented in Table 2:

Table 2. Group of Ages

Age_groups	Frequency	Percent
18-24	89	43,6 %
25-30	37	18,1 %
31-35	21	10,3 %
36-40	30	14,7 %
41-45	15	7,4 %
46-50	12	5,9 %
Total	204	100,0 %

They were studying Applied Sciences (42,2%) while a smaller percentage of the participants studied in another scientific area (19,6% Theoretical and 38,2% Technological).

Tale 3. Studies

Studies	Frequency	Percent
Economics and Management Sciences	8	3,9 %
Environmental Sciences	6	2,9 %
Health Sciences	8	3,9 %
Humanities	18	8,8 %
Other	19	9,3 %
Positive Sciences	86	42,2 %
Sciences of Engineers	59	28,9 %
Total	204	100,0 %

They also answered to the question concerning their monthly bill of their mobile phone:

Table. 4. Monthly bill

Age_groups	<=10Euro	11-20	21-30	31-40	40-50
18-24	68,5%	21,3%	9,0%	1,1%	
25-30	56,8%	16,2%	10,8%	8,1%	8,1%
31-35	28,6%	28,6%	28,6%	4,8%	9,5%
36-40	16,7%	23,3%	36,7%	10,0%	13,3%
41-45	33,3%	13,3%	20,0%	26,7%	6,7%
46-50	33,3%	16,7%	33,3%		16,7%
Total	50,0%	20,6%	17,6%	5,9%	5,9%

4.6.2 Storage Practices

In this subsection of questions the users answered about their storage practices. The results are as follows:

- Do you store sensitive personal data on your mobile device? (e.g. photographs/videos/conversations' recordings etc.).

Table. 5. Store sensitive personal data

Age_groups	No	Yes
18-24	15,7%	84,3%
25-30	16,2%	83,8%
31-35	9,5%	90,5%
36-40	10,0%	90,0%
41-45	46,7%	53,3%
46-50	41,7%	58,3%
Total	18,1%	81,9%

As we can see, in the results of our survey, 81,9% of the users do store sensitive personal data on their mobile devices such as: photographs, videos, conversations' recordings etc. In addition, a small percentage of the users 18.1% do not stores important sensitive personal data on their mobile devices.

- Do you store important passwords on your mobile device? (e.g. Bank passwords, Alarm passwords etc.)

Table. 6. Store important passwords

Age_groups	No	Yes, encrypted	Yes, without encryption
18-24	74,2%	12,4%	13,5%
25-30	81,1%	2,7%	16,2%
31-35	76,2%	9,5%	14,3%
36-40	50,0%		50,0%
41-45	86,7%	13,3%	
46-50	66,7%	8,3%	25,0%
Total	72,5%	8,3%	19,1%

As we can see, 72,5% of the users do not store important passwords on their mobile devices. In addition, a great percentage of the users 19.1% (as we will analyze below) stores important passwords such as bank PINs', alarm passwords etc. The percentage that actually stores important passwords on their device without encryption is 19,1% and only 8,3% encrypted.

Even though 8,3% answered "encrypted" in fact this percentage is much smaller since immediately after we set the question "which method of encryption do you use?" and only 1% knew an encryption method and used it on their device.

4.6.3 PIN practices

In this subsection of questions the users answered about the password practices they apply. The results are as follows:

- Have you activated the PIN question on your SIM card?

Table. 7. Activated the PIN question on your SIM card?

Age_groups	No	Yes
18-24	23,6%	76,4%
25-30	13,5%	86,5%
31-35	14,3%	85,7%
36-40	40,0%	60,0%
41-45	40,0%	60,0%
46-50	33,3%	66,7%
Total	25,0%	75,0%

We observe that 75% has activated the PIN question on their SIM card. But, as we can see in table 8 the vast majority (85%) never changes their PIN.

- How often do you change the PIN question on your mobile device?

Table. 8. Frequency of change of the PIN question?

Age_groups	Never	Once a year	Twice a year	3 times a year	More often
18-24	80,9%	16,9%			2,2%
25-30	91,9%	8,1%			
31-35	90,5%	9,5%			
36-40	86,7%	6,7%	3,3%	3,3%	
41-45	93,3%				6,7%
46-50	83,3%				16,7%
Total	85,8%	10,8%	0,5%	0,5%	2,5%

- Do you have a PIN on your mobile's phone Screen-Saver and how often do you change it?

Table. 9. Activated the PIN question on mobile's phone Screen-Saver?

Age_groups	Once a year	Twice a year	3 times a year	I do not know if it has such an option	The device does not have such an option	More often	Never
18-24	14,6%	11,2%	7,9%	9,0%	6,7%	15,7%	34,8%
25-30	16,2%	2,7%	5,4%	2,7%	13,5%	10,8%	48,6%
31-35	9,5%			4,8%	4,8%	4,8%	76,2%
36-40	10,0%	6,7%		43,3%		3,3%	36,7%
41-45				40,0%	6,7%		53,3%
46-50	8,3%		8,3%	41,7%	16,7%		25,0%
Total	12,3%	6,4%	4,9%	16,7%	7,4%	9,8%	42,6%

- Do you protect sensitive applications with a pin or touch gestures?

Table. 10. Protection of sensitive applications with a pin or touch gestures

Age_groups	No	Yes
18-24	69,7%	30,3%
25-30	78,4%	21,6%
31-35	61,9%	38,1%
36-40	90,0%	10,0%
41-45	93,3%	6,7%
46-50	83,3%	16,7%
Total	76,0%	24,0%

- How often do you change the PIN on your cash card?

Table. 11. Frequency of change of the PIN on your cash card?

Age_groups	3 times a year	More often	Never	Once a year	Twice a year
18-24	3,4%	3,4%	82,0%	7,9%	3,4%
25-30		5,4%	73,0%	18,9%	2,7%
31-35		9,5%	66,7%	19,0%	4,8%
36-40	3,3%	3,3%	90,0%		3,3%
41-45			80,0%	20,0%	
46-50		25,0%	50,0%	16,7%	8,3%
Total	2,0%	5,4%	77,9%	11,3%	3,4%

- Do you give your PIN to third persons?

Table. 12. Do you give your PIN?

Age_groups	**Yes**	**No**
18-24	20,2%	79,8%
25-30	21,6%	78,4%
31-35	9,5%	90,5%
36-40	40,0%	60,0%
41-45	20,0%	80,0%
46-50	25,0%	75,0%
Total	22,5%	77,5%

4.6.4 Device Protection

In this subsection of questions the users answered about how careful they are with their device. The results are as follows:

- Have you ever lost your phone or has it ever been stolen?

Table. 13. Lose or stealing of the device?

Age_groups	Twice	Once	Never	More	3 times
18-24	1,1%	22,5%	76,4%		
25-30	5,4%	18,9%	73,0%		2,7%
31-35	14,3%	19,0%	66,7%		
36-40	10,0%	30,0%	56,7%	3,3%	
41-45		6,7%	93,3%		
46-50		25,0%	58,3%	16,7%	
Total	4,4%	21,6%	72,1%	1,5%	0,5%

- Have you ever forgotten your device e.g. at a coffee shop?

Table. 14. Forget the Device.

Age_groups	Twice	Once	Never	More	3 times
18-24	5,6%	13,5%	78,7%	2,2%	
25-30	2,7%	10,8%	78,4%	2,7%	5,4%
31-35	9,5%	9,5%	81,0%		
36-40	13,3%	23,3%	63,3%		
41-45		13,3%	80,0%	6,7%	
46-50		25,0%	58,3%	16,7%	
Total	5,9%	14,7%	75,5%	2,9%	1,0%

4.7 Survey analysis Among the Ages group

Afterwards, by using the results we examined the trends of the users with respect to our basic variable, which is *age*. Our conclusions are presented below:

4.7.1 Ages versus Privacy

By examining the collected data of our questionnaire, we came up with some interesting trends among the participants. Moreover, most of the participants have lost their device, at least once. In addition, they save their cash cart pin without encryption. Lastly, they give their device to others.

- **In the age group 18-24** the 74,1% is careful and does not store at all important and other passwords such as Bank passwords, Alarm passwords etc. on their mobile phone. Nevertheless, a great percentage and in specific the 23,7% do stores important passwords without encryption and the 82, 6% never changes the PIN of their cash card. The 2,2% stores important passwords but uses encryption. The 84,3% stores sensitive personal data such as photographs, videos etc., and 15,7% does not. The PIN question in the SIM card is enabled by the 76,4% but the 80,1% never changes it. The 23,5% has lost their device at least once and the 20,2% gives their passwords to third persons.
- **In the age group 25-30** the 81% is careful and does not store important and other passwords on their mobile phone. The 16,2% do stores important passwords without encryption and of this percentage nobody ever changes the PIN of their cash card. The 2,8% stores important passwords but uses encryption. The 83,8% stores sensitive personal

data and 78,3% does not protect them with a PIN or a touch gesture. The 16,2% does not store sensitive data. The PIN question in the SIM card is not enabled by the 86,4% while the 91,9% never changes it. The 27% has lost their device at least once and the 26,1% gives their passwords to third persons.

- **In the age group 31-40** the 35,7% stores important passwords without encryption while the rest 64,3% does not store important passwords on their device. The 78,6% never changes the pin of their cash card. The 90,5% stores sensitive personal data and nobody protects them with a PIN or a touch gesture. The PIN question in the SIM card is enabled by the 73,8% and the 66,7% never changes it. The 38% has lost their device at least once and the 24% gives their passwords to third persons.
- **In the age group 41-50** the 30,6% stores important passwords without encryption while only 5,6 stores important passwords encrypted. The 88,9% never changes the pin of their cash card. The 63,9% stores sensitive personal data and the 58,3% does not protect them with a PIN or a touch gesture. The PIN question in the SIM card is enabled by the 61,1% and the 55,5 never changes it. The 27,8% has lost their device at least once and the 27,8% gives their passwords to third persons.

4.8 Hypotheses of Survey

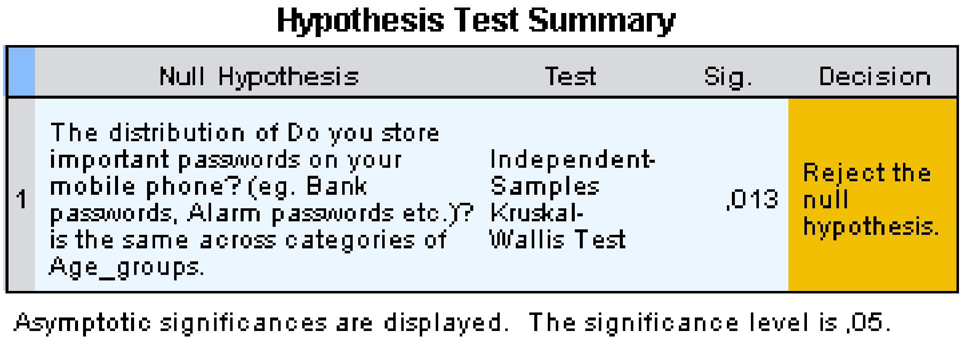
4.8.1 First Hypothesis

First hypothesis: Does age correlate to users' practices concerning the storage of important passwords on their mobile phone?

To check if this hypothesis applies we use the non-parametric Kruskal – Walis test since we don't have a normal distribution and we have more than two groups to check. We examine if there is a statistically significant difference between the age groups in correlation to the variable Store_important_password. Initially we set the null and the alternative hypothesis:

- H_0: The distribution of the variable Store_important_password is the same to all age groups.
- H_1 The distribution of the variable Store_important_password is not the same to all age groups.

The results of the test from the SPSS are presented in the following picture:



Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Do you store important passwords on your mobile phone? (eg. Bank passwords, Alarm passwords etc.)? is the same across categories of Age_groups.	Independent-Samples Kruskal-Wallis Test	,013	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

Since p0,013 <0.05 we reject the H_0. So, we observe that the users' age concerning the storage of important passwords on their mobile phones do correlates.

The median values for every age per category of answers is the following:

Do you store important passwords on your mobile phone? (e.g. Bank passwords, Alarm passwords, web browsers etc.)			
	no	**Yes, with encrypted**	**Yes, without encrypted**
median	25-30	18-24	31-35

From the median values we can see that younger ages 18-30 are more careful concerning the storage of personal data on their device. The age of 31-35 is on the borderline, since it is the median in the third category (yes, without encryption) of the first case of the survey. At this age it seems that they do not store PIN and passwords on their device. From the median and above though, i.e. at ages 36- 50, as is apparent from the descriptive statistics, the users do store important password such as Bank passwords, with no encryption.

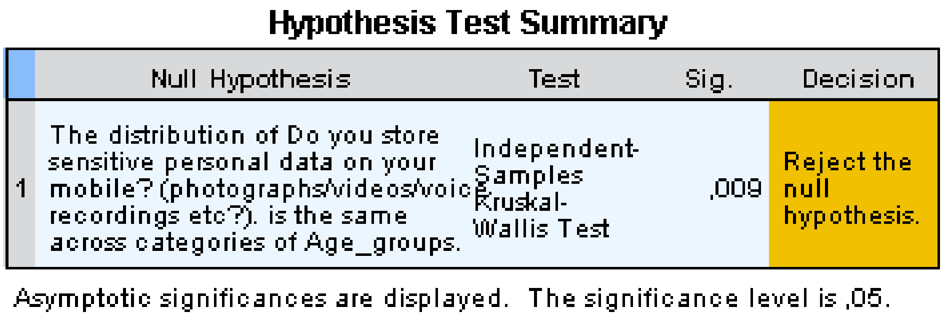
4.8.2 Second Hypothesis:

Second hypothesis: Does age correlate to the users' practices concerning the store sensitive personal data on their mobile (photographs / videos /voice recordings etc.)?

To check if this hypothesis applies we will use again the non-parametric Kruskal – Walis test since we don't have a normal distribution and we have more than two groups to examine. We will check if there is a statistically significant difference between the age groups in correlation to the variable Store_important_password. Initially, we set the null and the alternative hypothesis:

- H_0: The distribution of the variable Store_important_password is the same in all age groups.
- H_1 The distribution of the variable Store_personal_data is not the same in all age groups.

The results of the Kruskal – Walis test from the SPSS are presented in the following picture:



Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Do you store sensitive personal data on your mobile? (photographs/videos/voice recordings etc?). is the same across categories of Age_groups.	Independent-Samples Kruskal-Wallis Test	,009	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

Since there is a statistically significant difference, where p= 0.009 <0.05 we reject the H_0. So we observe that the users' age in correlation to the sensitive personal data storage on their mobile (photographs / videos /voice recordings etc.) do relates.

The median values for every category of answers per age is the following:

Do you store sensitive personal data on your mobile device? (e.g. photographs/videos/conversations' recordings etc.).		
	No	**Yes**
Median	25-30	25-30

From the median and above though, i.e. at ages 31-50, as is apparent from the descriptive statistics, the users do store sensitive personal data but the percentage gradually decreases in older ages, while it increases in the younger ones.

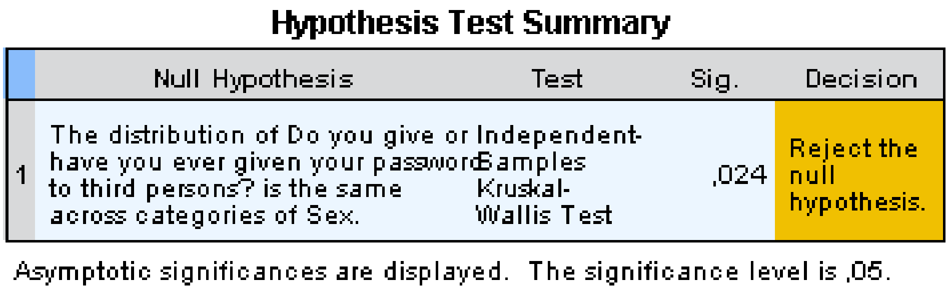
4.8.3 Third Hypothesis:

Third hypothesis: Does gender correlate to the users' practices concerning the sharing of their PIN with third persons?

To check if this hypothesis applies we will use again the non-parametric Kruskal – Walis test since we don't have a normal distribution and we have more than two groups to examine. We will check if there is a statistically significant difference between the age groups in correlation to the variable Store_important_password. Initially, we set the null and the alternative hypothesis:

- H_0: The distribution of the variable Given_your_pin is the same between the two genders.
- H_1 The distribution of the variable Given_your_pin is not the same between the two genders.

The results of the Kruskal – Walis test from the SPSS are presented in the following picture:



Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Do you give or have you ever given your password to third persons? is the same across categories of Sex.	Independent-Samples Kruskal-Wallis Test	,024	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

Since there is a statistically significant difference, where p= 0.024 <0.05 we reject the H_0. So we observe that the users' age in correlation to Given_your_pin do relates.

The median values for every category of answers per gender is the following:

Do you give your PIN to third persons?		
	yes	**no**
Median	Female	Male

We observe that the median for the variable *female* is in the answer "yes", while for the variable *male* the median is in the answer "no".

4.9 Conclusions

From the cases' results we noticed that the age factor affects the users' attitude concerning the storage of PIN and password and of sensitive personal data in general. Smaller age groups, as we have seen, seem to be more cautious in relation to older age groups, about the protection of their personal data. Generally, in all age groups of this category the percentage of users who do not follow any practices in order to protect their Pin and Passwords is about 19,1%.

In addition, we saw that the gender factor affects the users' attitude on whether they give their PIN to third persons. We see that females give their PIN to third persons more easily. The percentage of users that gives their PIN is 22,5%.

Finally, the results of this study, shows that while many smartphone users do take some security measures, a high percentage of them, 24%, still ignores potential risks.

5

Results: “Mobile Phones & Behavioral Modalities: Surveying users’ practices”

5.1 Introduction

Mobile phones are one of the most popular means of access to the internet. Users, via the telephone, connect to different services such as: Google, social networks, work accounts, banks accounts, etc. Those services, are many times, left open in their device. This enables risks, such as, loss or/and the violation of their personal data. In addition, in case of device theft after login, full access to sensitive data and applications may be fully granted. The purpose of this research is to analyze the most salient patterns characterizing user practices regarding certain behavioral modalities including: the way of using the various applications, power consumption, touch gestures and guest users’ habits. To this end, we used an original questionnaire, created for the needs of the specific survey, to examine whether we can find some trends among the users. This can give us a qualitative information, for the different behaviors / “characters” of users, in order to be used in further research regarding User’s Continuous Authentication.

5.2 Problem Analysis

In the second part of the research we analyze the most salient patterns characterizing user practices regarding certain behavioral modalities including: the way of using the various applications, power consumption, touch gestures and guest users' habits. To this end, we used an original questionnaire, created for the needs of the specific survey, to examine whether we can find some trends among the users. This can give us a qualitative information, for the different behaviors / "characters" of users, in order to be used in further research regarding User's Continuous Authentication.

Our survey answers one main research questions:

- What are the behavioral modalities among the users?
- Can analysis of Behavioral Modalities be utilized in the context of Continuous User Authentication?

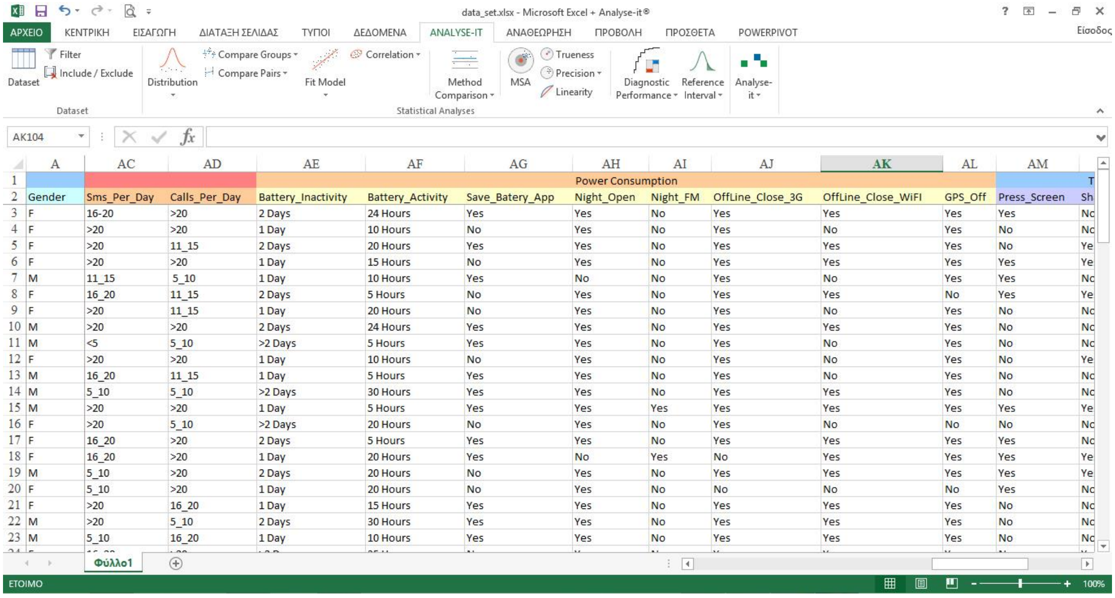
Shi et al. [SNJC2011] presented an approach that was built on the concept that most users are habitual in nature and are prone to performing similar tasks at a certain time of the day. The researchers collected a wide range of behavioral information such as location, communication, and usage of applications, in order to create a user profile. They further modeled all good events as ones that are expected to be performed at a certain time of day, which is an assumption of habit that is not proven in the literature.

What we want to see via our questionnaire is if users do perform similar tasks at a certain time of the day, confirming Shi's hypothesis. In addition, through this approach we want to examine under what basis the user's profile can be created.

We examine some statistical hypotheses concerning age in relation to the use of applications in a specific part of the day, the time period of use by the users and the correlation between Age and Communication. We will examine if there is a statistically significant difference between the age groups in correlation to the variables.

5.3 Second Survey Features Encoding

In the second survey the data were collected with an original questionnaire, created for the needs of the specific survey. This questionnaire consists of 49 questions and the data were collected by 100 members of Cyprus University of Technology. The results of the questionnaire were corresponded to variables and entered in an SPSS worksheet.



	A	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	
1							Power Consumption						T
2	Gender	Sms_Per_Day	Calls_Per_Day	Battery_Inactivity	Battery_Activity	Save_Batery_App	Night_Open	Night_FM	OffLine_Close_3G	OffLine_Close_WiFI	GPS_Off	Press_Screen	Sh
3	F	16-20	>20	2 Days	24 Hours	Yes	Yes	No	Yes	Yes	Yes	Yes	Nc
4	F	>20	>20	1 Day	10 Hours	No	Yes	No	Yes	No	Yes	No	Nc
5	F	>20	11_15	2 Days	20 Hours	Yes	Yes	No	Yes	Yes	Yes	No	Ye
6	F	>20	>20	1 Day	15 Hours	No	Yes	No	Yes	Yes	Yes	Yes	Ye
7	M	11_15	5_10	1 Day	10 Hours	Yes	No	No	Yes	No	Yes	Yes	Nc
8	F	16_20	11_15	2 Days	5 Hours	No	Yes	No	Yes	Yes	No	Yes	Ye
9	F	>20	11_15	1 Day	20 Hours	No	Yes	No	Yes	No	Yes	No	Nc
10	M	>20	>20	2 Days	24 Hours	Yes	Yes	No	Yes	Yes	Yes	No	Nc
11	M	<5	5_10	>2 Days	5 Hours	Yes	Yes	No	Yes	No	Yes	No	Nc
12	F	>20	>20	1 Day	10 Hours	No	Yes	No	Yes	No	Yes	No	Ye
13	M	16_20	11_15	1 Day	5 Hours	Yes	Yes	No	Yes	No	Yes	No	Nc
14	M	5_10	5_10	>2 Days	30 Hours	Yes	Yes	No	Yes	Yes	Yes	No	Nc
15	M	>20	>20	1 Day	5 Hours	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Ye
16	F	>20	5_10	>2 Days	20 Hours	No	Yes	No	Yes	No	No	No	Nc
17	F	16_20	>20	2 Days	5 Hours	Yes	Yes	No	Yes	Yes	Yes	Yes	Nc
18	F	16_20	>20	1 Day	20 Hours	Yes	No	Yes	No	Yes	Yes	Yes	Ye
19	M	5_10	>20	2 Days	20 Hours	No	Yes	No	Yes	Yes	Yes	Yes	Ye
20	F	5_10	>20	1 Day	20 Hours	No	Yes	No	No	No	No	Yes	Nc
21	F	>20	16_20	1 Day	15 Hours	Yes	Yes	No	Yes	Yes	Yes	No	Nc
22	M	>20	5_10	2 Days	30 Hours	Yes	Yes	No	Yes	Yes	Yes	No	Nc
23	M	5_10	16_20	1 Day	10 Hours	Yes	Yes	No	Yes	Yes	Yes	No	Nc

Fig. 1. The excel worksheet before encoding.

Our features were encoded with numerical values and we present the most significant ones.

In questions related to “Part of the Day”, the variables, for example Morning (M), Noon (N), etc. were coded with numerical values (1,2,3 etc.) as follows:

All_Day	: 1
Morning (M)	: 2
Noon (N)	: 3
Afternoon (A)	: 4
Night (Ni)	: 5

In the results of the questionnaire there also was a combination of “Parts of the Day”. They were coded with numerical values as follows:

- N_A: 1, M_N: 2, A_Ni: 3, M_A: 4, M_Ni: 5, M_N_A: 6, M_N_Ni: 7

The answers “Yes” and “No” were coded with numerical values as follows:

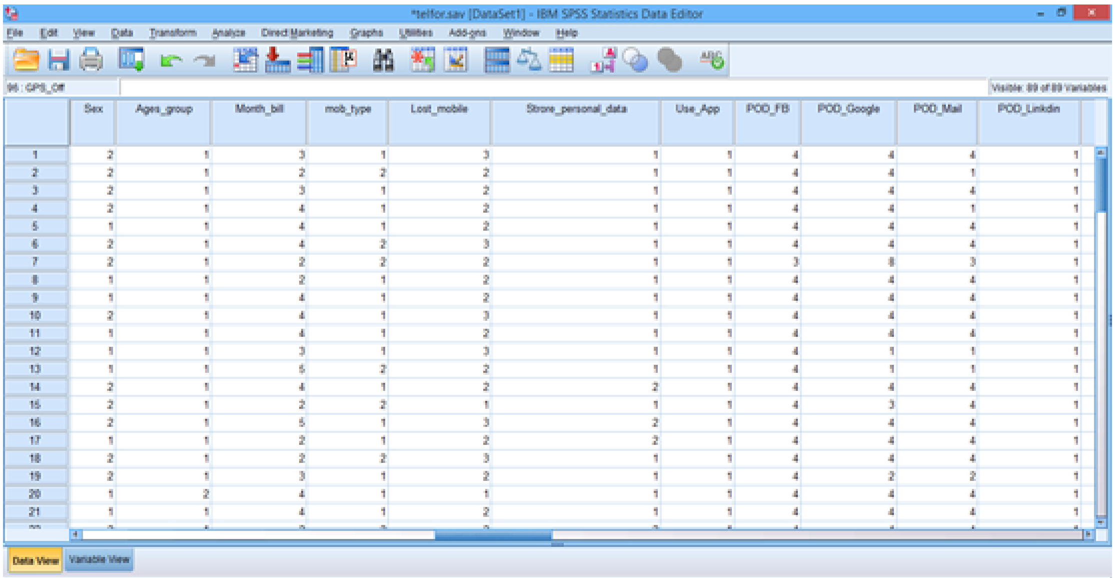
- Yes: 1, No: 2

Value 0 was given to the variables when there was no answer, while values from 1 to 5 were given to variables relevant to the applications’ hours of use.

In questions concerning “Power Consumption” answers were coded with numerical values as follows:

- <5: 1, 5-10: 2, 11-15: 3, 16-20: 4, >20 : 5
- 1 Day: 1, 2 D: 2, >2D: 3
- 5 Hours: 1, 10 H: 2, 15 H: 3, 20 H: 4, 24 H: 5, 30 H: 6, 35 H: 7

The SPSS worksheet after encoding:



	Sex	Ages_group	Month_bill	mob_type	Lost_mobile	Strore_personal_data	Use_App	POD_FB	POD_Google	POD_Mail	POD_Linkdin
1	2	1	3	1	3	1	1	4	4	4	1
2	2	1	2	2	2	1	1	4	4	1	1
3	2	1	3	1	2	1	1	4	4	4	1
4	2	1	4	1	2	1	1	4	4	1	1
5	1	1	4	1	2	1	1	4	4	4	1
6	2	1	4	2	3	1	1	4	4	4	1
7	2	1	2	2	2	1	1	3	8	3	1
8	1	1	2	1	2	1	1	4	4	4	1
9	1	1	4	1	2	1	1	4	4	4	1
10	2	1	4	1	3	1	1	4	4	4	1
11	1	1	4	1	2	1	1	4	4	4	1
12	1	1	3	1	3	1	1	4	1	1	1
13	1	1	5	2	2	1	1	4	1	1	1
14	2	1	4	1	2	2	1	4	4	4	1
15	2	1	2	2	1	1	1	4	3	4	1
16	2	1	5	1	3	2	1	4	4	4	1
17	1	1	2	1	2	2	1	4	4	4	1
18	2	1	2	2	3	1	1	4	4	4	1
19	2	1	3	1	2	1	1	4	2	2	1
20	1	2	4	1	1	1	1	4	4	4	1
21	1	1	4	1	2	1	1	4	4	4	1

Fig. 2. The SPSS worksheet after encoding.

5.4 *Methodology*

Our survey was conducted using in-person delivery technique, with a total of 100 participants that were requested to complete it anonymously and voluntarily. A very useful evaluation method for surveying user's practices is the use of multiple-choice questionnaires [ACS2009]. This method was selected from other alternatives because is more accurate and has a bigger degree of participation from the respondents.

The target group of the survey is Cyprus University of Technology (CUT) students, professors, university members and visitors. We mostly choose members of the Academic faculty because they are more receptive to new technologies. They also understand better the technological evolution than externals.

For the statistical analysis, our main variable is Age because we wanted to evaluate the significance degree regarding mobile phone's behavioral modalities between different age groups. What we want to do with the data is to see if there are any trends among the users.

The questionnaire is original and created for the needs of the specific survey. It consists of five subsections and is formed as follows:

1. Demographics.
2. The Applications' way of use.
3. Power Consumption.
4. Touch Gestures.
5. Guest Users

We tried to formulate our questions in a fully understood way, in order to be answered and filled correctly. Also, 80% of the participants answered the questionnaire through an interview and the 20%, under instructions, via e-mail. The parts of the questionnaire follow a logical continuity and are clearly distinct, since we have used headings that indicate each group of questions.

The size of the questionnaire we tried to be such as to allow the participants consent and grant its completion. We tried to formulate our questions in a fully understood way, in order to be answered and filled correctly. Also, we made sure that the structure and presentation of the questionnaire was simple, so as to encourage our collaboration with the participants. It is also necessary to note that after the completion of the questionnaire, we proceeded to its pilot implementation in order to diagnose any problems that could arise during the answering of questions. Finally, we composed an accompanying letter through which was disclosed to the participants the purpose of the investigation and stressed out that all given information would remain strictly confidential and would solely be used for research purposes.

Survey responses were analyzed using descriptive analysis, Crosstabs, Frequencies and Kruskal–Wallis test (p<.05) on SPSS.

The main limitation to this research is that the generalizability of the study is limited because the subject pool only included members of the academic faculty of CUT.

5.5 Survey Results

Afterwards, the results are presented in full detail and an analysis and discussion of every issue is made:

5.5.1 Demographics

The participants were asked about their gender, age and field of studies. 40% of the participants where males and 60% where females. Their ages are presented in Fig. 1:

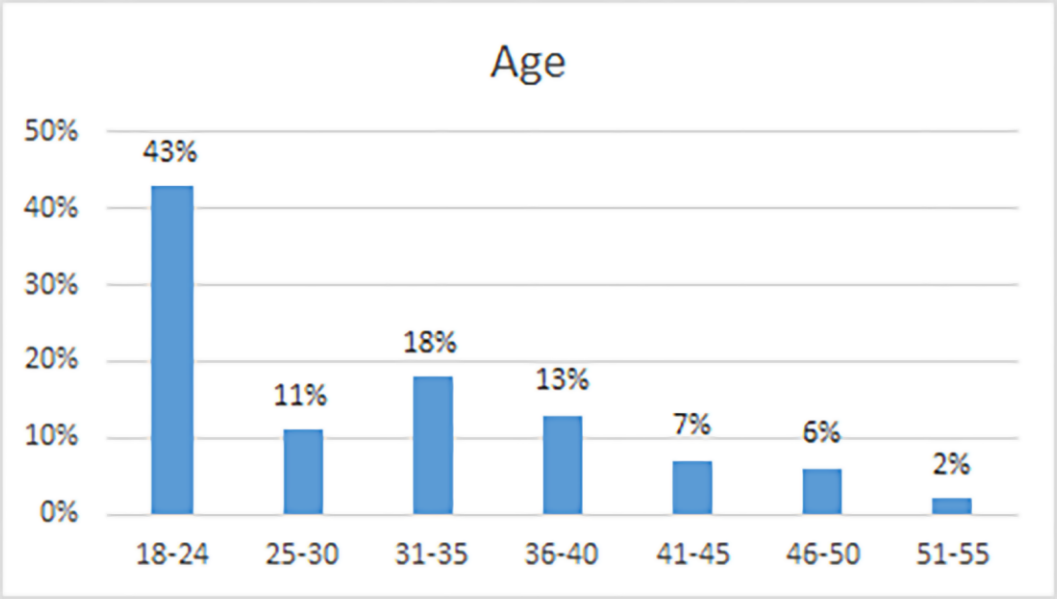


Fig. 1. Age of the participants

They were studying Applied Sciences (51%) while a smaller percentage of the participants studied in another scientific area (33% Theoretical and 16% Technological). 75% of the participants uses Android and 25% uses iPhone. Of interest are the answers to the question: "Have you ever lost your mobile phone, or has it ever been stolen?", 24% of the respondents answered: "once", 9% answered: "more than once", and 67% answered that "they have never lost their device". In the question: "Do you save sensitive personal data on your mobile phone?", 80% answered: "Yes", while none of them (100% answered: "No") encrypts them.

5.5.2 The Applications' way of use

In this subsection of questions the users answered about the use of certain applications of their device. It is important to mention that beyond the general results in the tables, there was no user observed to have the exact same behavior with anyone else over the use of applications. The results are as follows:

Table 1: Use of Applications

Choose the applications that you use:	***Yes***	***No***
Facebook	83%	17%
Google	96%	4%
E-mail	86%	14%
Linkedin	22%	78%
Youtube	88%	12%

Table 2: Part of the Day

Part of the day?	***No use***	***All Day***	***Combination***
Facebook	17%	61%	22%
Google	4%	75%	21%
E-mail	14%	69%	17%
Linkedin	78%	12%	10%
Youtube	12%	69%	19%

Table 3: Hours of Use

How many hours;	***0***	***1***	***2***	***3***	***4***	***5***
Facebook	17%	29%	20%	10%	6%	18%
Google	4%	61%	21%	9%	0%	5%
E-mail	14%	62%	17%	3%	0%	4%
Linkedin	78%	18%	2%	2%	0%	0%
Youtube	12%	49%	22%	8%	4%	5%

Table 4: Other Applications Behavioral Modalities

Questions	Yes	No
Do you use messenger?	72%	28%
Do you use dropbox?	48%	52%
Do you have priorities in the use of apps?	35%	65%
Many apps “running” at the same time?	63%	37%
Do you use “search” by using the mic?	9%	91%

The respondents also answered in some other questions in the subscale “The applications’ way of use”, which are the following: In the question “Do you open your e-mail by using the application’s icon or by the notifications?”, 29% answered “by the application’s icon”, 30% “by the notifications” and 41% “in both ways”. In questions about the writing language during the use of applications: “Do you write in Greeklish?, Greek?, English?”, the respondents answered either a single language or a combination such as Greek and English or even Greeklish. The results are as follows: Greek 67%, Greeklish 76%, English 48%. Lastly, the participants replied to a question concerning the number of SMS and Phone Calls they send - make per day. Their answers are presented in Fig. 2:

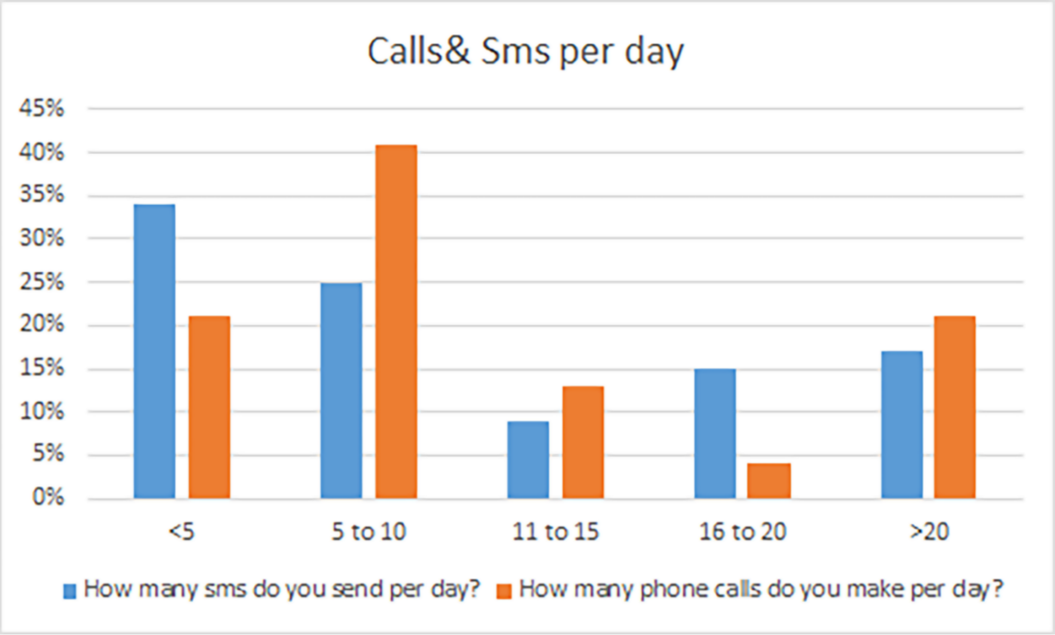


Fig. 2. Calls & SMS per day.

The results clearly show that there is consistency in the answers of the respondents. As shown in Table 1, where users answered if they use specific applications, we noticed that, for example 17% does not use Facebook, 4% does not use Google, etc. In tables 2 and 3 we observe that the corresponding values in columns “No Use” and “0” coincide exactly. For example, at table 2, the values from the column “No Use” are: 17% does not use Facebook and 4% does not use Google, etc. From these results we conclude that we have no random answers. In addition, we have no Bias since the answers are clear and precise.

5.5.3 Power Consumption

This subsection studies some behavioral modalities of the participants concerning Power Consumption. Initially, we wanted to learn how long the battery of their mobile phone lasts with no use. The answers we got were as follows: “One day”: 46%, “Two days”: 35%, “More than 2 days”: 18%. Then we asked them: “How many hours does it last with your usual use?”. The answers we took are presented in Fig. 3:

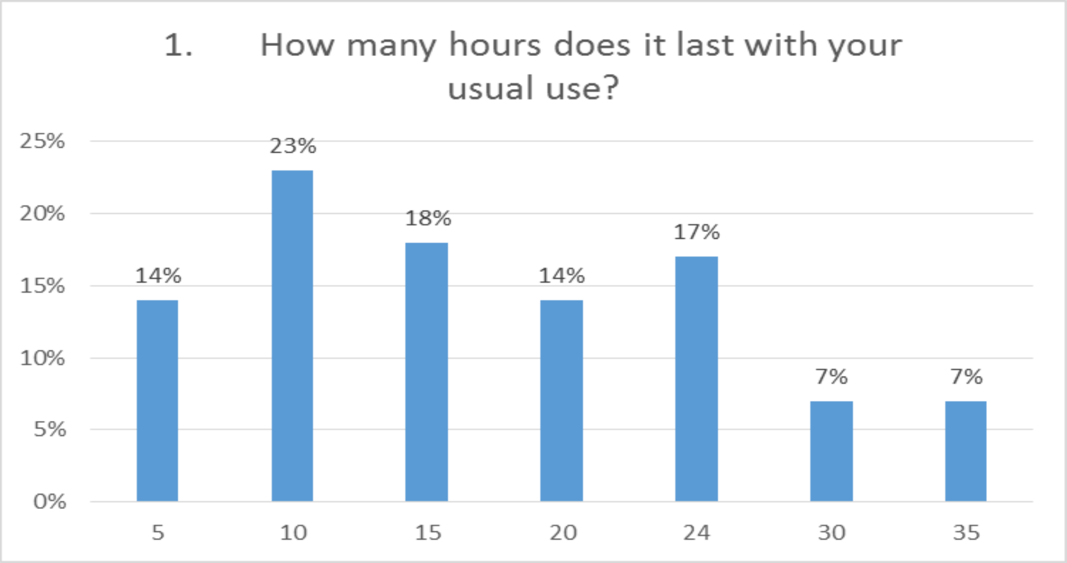


Fig. 3. Power Consumption.

As we can see in figure 3 the battery's life is the major weak spot of smartphones. The increased features and mainly the increased speed data access are to be "blamed" for that. The majority of phone batteries in the whole sample lasts 1 to 2 days maximum, with a 69% lasting even less than one day. 17% lasts one day and only 14% lasts more than one day.

Table 5: Other Power Consumption Modalities

Questions	Yes	No
Mobile phone 'on' during night?	89%	11%
'Flight mode' during night?	7%	93%
Do you turn off the '3G' when you are not connected?	82%	18%
Do you turn the 'wifi' off when you are not connected?	60%	40%
Do you turn the 'GPS' off?	80%	20%

5.5.4 Touch Gestures

This subsection studies some behavioral modalities of the participants concerning Touch Gestures.

Table 6: Touch Gestures Modalities

Questions	Yes	No
Do you press with strength the screen when it is not responding?	36%	64%
Do you shake your device when it is not responding quickly?	31%	69%
Wipe the screen after calling?	58%	42%

5.5.5 Guest Users

This subsection studies some behavioral modalities of the Guest Users. In the question "Do you give your mobile phone to ‘guest users’?”, 36% answered: “Yes” and 64% answered: “No”. Those who answered “Yes”, also answered to the question: “Does the use by others happen at specific hours or days?”, 25% answered: “Yes” and 75% answered: “No”. Then they were asked if the Guest Users use certain applications: 67% answered: “Yes” and 33% answered: “No”. Finally, they were asked: “Are guest users people you trust or close to you?”, 97% answered: “Yes” and only 3% answered: “No”.

5.6 Survey Analysis Among the Age Groups

Afterwards, by using the results we examined the trends of the users with respect to our basic variable, which is *age*. Our conclusions are presented below:

5.6.1 Behavioral Trends of the Age Groups

By examining the collected data of our questionnaire, we came up with some interesting trends among the participants. Most of them use the applications at any time during the day and they leave many apps running at the same time. Also, it seems that users are “always connected” even leaving their phone on when they sleep. The offline-airplane mode is rarely used, GPS is enabled only when needed, and WiFi is turned off when not in use. The use of Greeklish is very popular while the microphone search is not popular at all.

Afterwards we present some tables relating to age in correlation to the use of applications in a specific part of the day as well as the period of time that people use them. The variables as we saw in chapter 4 are as follows: All_Day, Morning (M), Noon (N), Afternoon (A), Night (Ni) and their combinations.

Table 7: Part of the Day Facebook

Age_Groups	A	A_Ni	All_Day	M_A	M_N	M_Ni	Ni	No_Use
18-24	2,2%	6,7%	84,4%	2,2%	2,2%			2,2%
25-30			45,5%			27,3%	18,2%	9,1%
31-35	11,1%	5,6%	55,6%			5,6%	5,6%	16,7%
36-40			46,2%			7,7%	7,7%	38,5%
41-45		14,3%	14,3%			14,3%		57,1%
46-50							25,0%	75,0%
56-60			100,0%					
Total	3,0%	5,0%	62,0%	1,0%	1,0%	6,0%	5,0%	17,0%

Table 8: Part of the Day Google

Age_Groups	A	A_Ni	All_Day	M	M_N_A	M_Ni	N	Ni	No_use
18-24	4,4%	4,4%	77,8%	2,2%			4,4%		6,7%
25-30			63,6%			9,1%		27,3%	
31-35	11,1%		83,3%		5,6%				
36-40			84,6%	7,7%				7,7%	
41-45	14,3%	14,3%	57,1%						14,3%
46-50			75,0%					25,0%	
56-60		100,0%							
Total	5,0%	5,0%	75,0%	2,0%	1,0%	1,0%	2,0%	5,0%	4,0%

Table 9: Part of the Day Mail

Age_Groups	A	A_Ni	All_Day	M	M_N_A	M_N_Ni	M_Ni	N	Ni	No_use
18-24	2,2%	4,4%	71,1%	4,4%			2,2%	2,2%		13,3%
25-30	9,1%		63,6%	9,1%					9,1%	9,1%
31-35	11,1%		77,8%		5,6%					5,6%
36-40	7,7%		61,5%	15,4%		7,7%				15,4%
41-45			57,1%							42,9%
46-50			50,0%						25,0 %	25,0%
56-60			100,0%							
Total	5,0%	2,0%	69,0%	5,0%	1,0%	1,0%	1,0%	1,0%	2,0%	14,0%

Table 10: Part of the Day Linkedin

Age_Groups	All_Day	N	Ni	No_use
18-24	2,2%	2,2%		95,6%
25-30	18,2%		27,3%	54,5%
31-35	33,3%		5,6%	61,1%
36-40	23,1%		15,4%	61,5%
41-45			14,3%	85,7%
46-50				100,0%
56-60			100,0%	
Total	12,0%	1,0%	9,0%	78,0%

Table 11: Part of the Day Youtube

Age_Groups	A	A_Ni	All_Day	M_Ni	N_A	Ni	No_use
18-24	4,4%	4,4%	82,2%		2,2%	2,2%	4,4%
25-30			63,6%	9,1%		9,1%	18,2%
31-35	11,1%		72,2%			11,1%	5,6%
36-40			61,5%			15,4%	23,1%
41-45	14,3%	14,3%	42,9%				28,6%
46-50			25,0%			25,0%	50,0%
56-60	100,0%						
Total	7,0%	3,0%	69,0%	1,0%	1,0%	7,0%	12,0%

Table 12: Hours of Use Facebook

Age_Groups	0	1	2	3	4	5
18-24		24,4%	17,8%	13,3%	8,9%	35,6%
25-30	9,1%	63,6%		18,2%	9,1%	
31-35	22,2%	27,8%	44,4%	5,6%		
36-40	38,5%	15,4%	30,8%	7,7%	7,7%	
41-45	71,4%	28,6%				
46-50	75,0%	25,0%				
56-60		50,0%				50,0%
Total	18,0%	29,0%	20,0%	10,0%	6,0%	17,0%

Table 13: Hours of Use Google

Age_Groups	0	1	2	3	5
18-24	6,7%	53,3%	22,2%	13,3%	4,4%
25-30		72,7%	18,2%		9,1%
31-35		66,7%	16,7%	11,1%	5,6%
36-40		61,5%	30,8%	7,7%	
41-45	28,6%	71,4%			
46-50		100,0%			
56-60			100,0%		
Total	5,0%	61,0%	21,0%	9,0%	4,0%

Table 14: Hours of Use Mail

Age_Groups	0	1	2	3	5
18-24	13,3%	73,3%	11,1%	2,2%	
25-30	9,1%	45,5%	27,3%		18,2%
31-35	5,6%	61,1%	27,8%		5,6%
36-40	15,4%	53,8%	15,4%	15,4%	
41-45	28,6%	42,9%	14,3%		14,3%
46-50	25,0%	75,0%			
56-60		50,0%	50,0%		
Total	13,0%	63,0%	17,0%	3,0%	4,0%

Table 15: Hours of Use Linkedin

Age_Groups	0	1	2	3
18-24	95,6%	2,2%	2,2%	
25-30	54,5%	36,4%		9,1%
31-35	61,1%	38,9%		
36-40	61,5%	30,8%		7,7%
41-45	85,7%		14,3%	
46-50	100,0%			
56-60		100,0%		
Total	78,0%	18,0%	2,0%	2,0%

Table 16: Hours of Use Youtube

Age_Groups	0	1	2	3	4	5
18-24	4,4%	35,6%	28,9%	11,1%	8,9%	11,1%
25-30	18,2%	45,5%	18,2%	18,2%		
31-35	5,6%	83,3%	11,1%			
36-40	23,1%	30,8%	38,5%	7,7%		
41-45	28,6%	71,4%				
46-50	50,0%	50,0%				
56-60		100,0%				
Total	12,0%	49,0%	22,0%	8,0%	4,0%	5,0%

Table 17: SMS per day

Age_Groups	<5	>20	11-15	16-20	5-10
18-24	6,7%	26,7%	13,3%	33,3%	20,0%
25-30	27,3%	27,3%	9,1%		36,4%
31-35	77,8%				22,2%
36-40	61,5%		15,4%	7,7%	15,4%
41-45	57,1%	14,3%			28,6%
46-50	50,0%				50,0%
56-60					100,0%
Total	34,0%	16,0%	9,0%	16,0%	25,0%

Table 18: Calls per day.

Age_Groups	5-10	<5	>20	11-15	16-20
18-24	26,7%	11,1%	37,8%	17,8%	6,7%
25-30	45,5%	27,3%	9,1%	9,1%	9,1%
31-35	44,4%	33,3%	11,1%	11,1%	
36-40	69,2%	23,1%		7,7%	
41-45	42,9%	42,9%	14,3%		
46-50	50,0%	25,0%		25,0%	
56-60	100,0%				
Total	41,0%	21,0%	21,0%	13,0%	4,0%

- **The age group 18-24** tends to mostly use Facebook, Google and Youtube at any time of the day and their battery doesn't last long. They prefer Greeklish to Greek and English, they make many phone calls and send many SMS.
- **The age group 25-30** even though tends to also use Facebook, Google and Youtube, at any time of the day, the duration time of their use is less than the previous age group. They also use e-mail and Linkedin at certain parts of the day. Greeklish are popular but they also use Greek and English. They send many SMS but phone calls are less. Finally their battery lasts longer.
- **The age group 31-35** mostly uses Google and e-mail and devotes more time in using these applications than the previous ones. They mostly use Greek and English, they make more phone calls and send less SMS. They are more green users since they consume less battery.

- **The age groups 36-40, 41-45, 46-50, 51-55 and 56-60** share the same trends so we put them all together. The most popular applications here are e-mail and Google, while the rest of the applications are used at night and for a short period of time. The use of Greeklish becomes again popular even though the use of Greek and English does not disappear. They send very few SMS and make very few phone calls. They are also green users since they don't consume too much battery, compared with younger age groups.

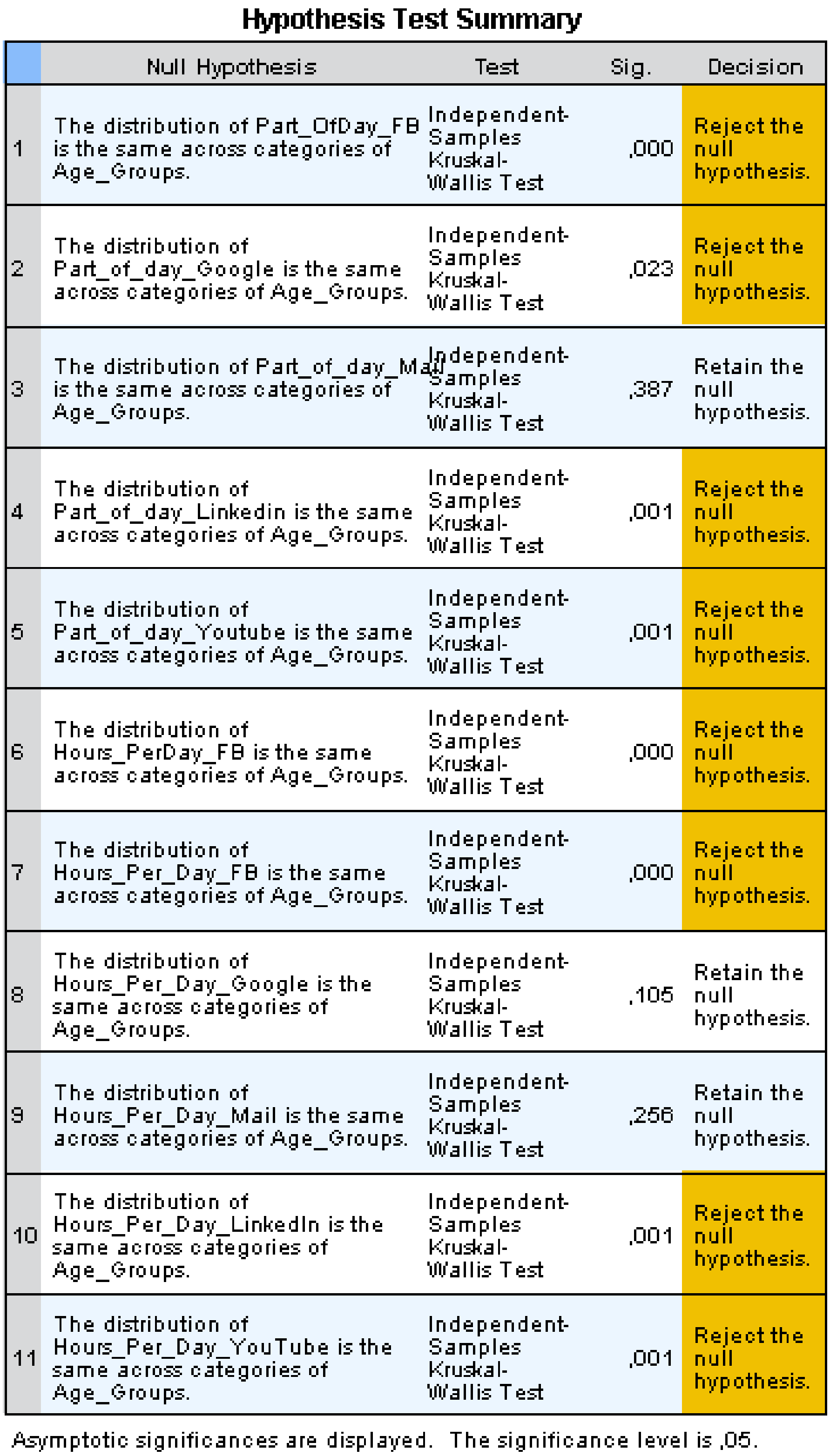
5.7 Hypotheses of Survey

As we saw in chapter 4 Shi et al. [SNJC2011] presented an approach that was built on the concept that most users are habitual in nature and are prone to performing similar tasks at a certain time of the day. The researchers collected a wide range of behavioral information such as location, communication, and usage of applications, in order to create a user profile. What we want to see via our hypotheses is if users do perform similar tasks at a certain time of the day, confirming Shi's hypothesis, and if there is statistically significant difference between the age groups. In addition, through this approach we want to examine under what basis the user's profile can be created.

Afterwards we examine some hypotheses concerning age in relation to the use of applications in a specific part of the day, the time period of use by the users and the correlation between Age and Communication. To check if these hypotheses apply we will use the non-parametric Kruskal – Walis test since we don't have a normal distribution and we have more than two groups to check. We will examine if there is a statistically significant difference between the age groups in correlation to the variables 1, 2, 3, …,12 which are presented below. Initially we set the null and the alternative hypothesis:

- H_0: The distribution of the variable 1, 2, 3,…,12 is the same to all age groups.
- H_1 The distribution of the variable 1, 2, 3, …,12 is not the same to all age groups.

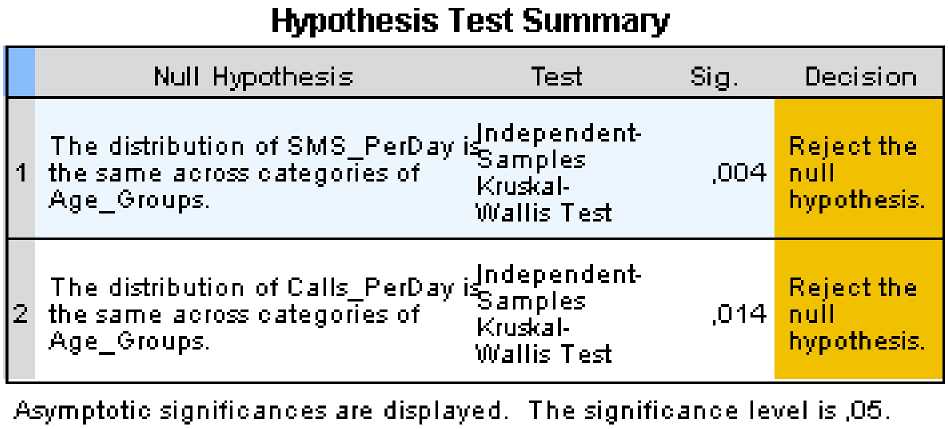
The results of the test from the SPSS are presented in the following picture:



Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Part_OfDay_FB is the same across categories of Age_Groups.	Independent-Samples Kruskal-Wallis Test	,000	Reject the null hypothesis.
2	The distribution of Part_of_day_Google is the same across categories of Age_Groups.	Independent-Samples Kruskal-Wallis Test	,023	Reject the null hypothesis.
3	The distribution of Part_of_day_Mail is the same across categories of Age_Groups.	Independent-Samples Kruskal-Wallis Test	,387	Retain the null hypothesis.
4	The distribution of Part_of_day_Linkedin is the same across categories of Age_Groups.	Independent-Samples Kruskal-Wallis Test	,001	Reject the null hypothesis.
5	The distribution of Part_of_day_Youtube is the same across categories of Age_Groups.	Independent-Samples Kruskal-Wallis Test	,001	Reject the null hypothesis.
6	The distribution of Hours_PerDay_FB is the same across categories of Age_Groups.	Independent-Samples Kruskal-Wallis Test	,000	Reject the null hypothesis.
7	The distribution of Hours_Per_Day_FB is the same across categories of Age_Groups.	Independent-Samples Kruskal-Wallis Test	,000	Reject the null hypothesis.
8	The distribution of Hours_Per_Day_Google is the same across categories of Age_Groups.	Independent-Samples Kruskal-Wallis Test	,105	Retain the null hypothesis.
9	The distribution of Hours_Per_Day_Mail is the same across categories of Age_Groups.	Independent-Samples Kruskal-Wallis Test	,256	Retain the null hypothesis.
10	The distribution of Hours_Per_Day_LinkedIn is the same across categories of Age_Groups.	Independent-Samples Kruskal-Wallis Test	,001	Reject the null hypothesis.
11	The distribution of Hours_Per_Day_YouTube is the same across categories of Age_Groups.	Independent-Samples Kruskal-Wallis Test	,001	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.



Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of SMS_PerDay is the same across categories of Age_Groups.	Independent-Samples Kruskal-Wallis Test	,004	Reject the null hypothesis.
2	The distribution of Calls_PerDay is the same across categories of Age_Groups.	Independent-Samples Kruskal-Wallis Test	,014	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

Since $p < 0.05$ we reject the H_0. Generally we see that the H_0 is rejected except for the variable concerning Google in “Part of Day” and the variables concerning Google and Mail in “Hour per Day”. Therefore the users, depending on age, are performing similar tasks at a certain time of the day. Our survey confirms Shi’s observation that users are prone to performing similar tasks at a certain time of the day.

5.8 Conclusions

Shi et al. [SNJC2011] presented an approach that was built on the concept that most users are habitual in nature and are prone to performing similar tasks at a certain time of the day. The researchers collected a wide range of behavioral information such as location, communication, and usage of applications, in order to create a user profile. Our survey confirms Shi’s observation that users are prone to performing similar tasks at a certain time of the day and that in this way a general users’ profile could be created but with some additional parameters analyzed in final conclusions, in chapter 8.

6

Experimental biometric data collection process via mobile smartphones

6.1 Introduction

In the third part of the research we present an experimental biometric data collection process via mobile smartphones. In the present experiment we recorded modalities of movement imprinting the user's walk patterns. We present the Data Collection Architecture by which we can collect the biometric data of the users, the way and type of storage and the Data Preparation for introduction to machine learning algorithms. Our methodology imprints the modalities of movement by the accelerometer and gyroscope sensors, in total 10 volunteers participated. The procedure was designed in such a way so as to collect data from every participant for three sessions of ten minutes each with a break of 5 minutes for instructions. The session recorded three sequences of 10 minute each while the participant: walked and hold the device on his hand, walked and had the device on his pocket, was running and had the device on his pocket. These sessions were repeated for two days so as to effectively capture the biometric behavior of the user. This gives us a total of 60 minutes' real use data of the smartphone for each user.

6.2 Data Collection Architecture

We will present the Data Collection Architecture by which we can collect the behavioral biometrics of the users and in specific the users walk patterns.

6.2.1 Methodology

Our methodology imprints the modalities of movement by the accelerometer and gyroscope sensors, in total 10 volunteers participated. Our methodology records modalities of movement imprinting the user's walk patterns, by the sensors of the accelerometer and the gyroscope. For the modality of movement, the readings were recorded by using the Sensor Kinetics Pro tool. The sensor of the accelerometer measures the acceleration in SI (m / s^2) units along the local [X, Y, Z] axes of the device and the sensor of the gyroscope measures the degree of rotation in SI (rad / s) units around the local axes of the device.

For the uniformity of measurements the participating volunteers used the same mobile phone device, and in particular the HTC Sensation XE Z715E, Android version 4.4.4 (Build number KTU84P), and the collection of data was done in a uniform and controllable environment.

In the present experiment we recorded modalities of movement imprinting the user's walk patterns. The procedure was designed in such a way so as to collect data from every participant for three sessions of ten minutes each with a break of 5 minutes for instructions.

In both sessions we recorded with the accelerometer and the gyroscope, sequences of 10 minutes while the participant:

- Walked and hold the device on his hand.
- Walked and had the device on his pocket.
- Was running and had the device on his pocket.

In both sessions we had a maximum recording rate of 50 samples per second. These sessions were repeated for two days so as to effectively capture the biometric behavior of the user. This gives us a total of 60 minutes' real use data of the smartphone for each user.

6.2.2 About the Sensors of Mobile Devices

Modern mobile devices, and in particular smartphones and tablets have a rich selection of built in sensors. The first popular sensor was the accelerometer followed by the gyroscope and the magnetometer. Combination (or fusion) of these sensors provide more sensing capabilities like rotation sensors, gravity sensor and the linear acceleration sensor. We used the Accelerometer and the gyroscope by using the Sensor Kinetics Pro Application [R2015].

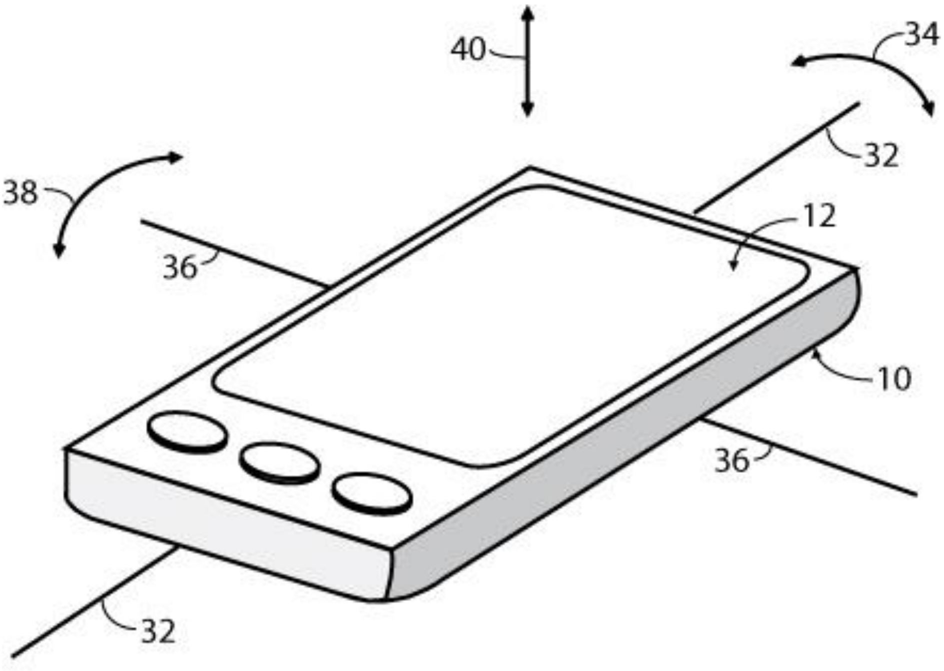


Fig.1. Schematic structure of movements [R2015].

6.2.3 The Accelerometer

An accelerometer is a sensor for testing the acceleration along a given axis. When a physical body accelerates at a certain direction, it becomes subject to a force equal to:

$$F=ma \qquad (1)$$

In accordance with Newton's Second Law. In this formula, m is the mass, a is the acceleration. Therefore, accelerometers are built on the principle of measuring the force exerted on a test body of a known mass along a given axis. The following drawing schematically shows the structure of an accelerometer.

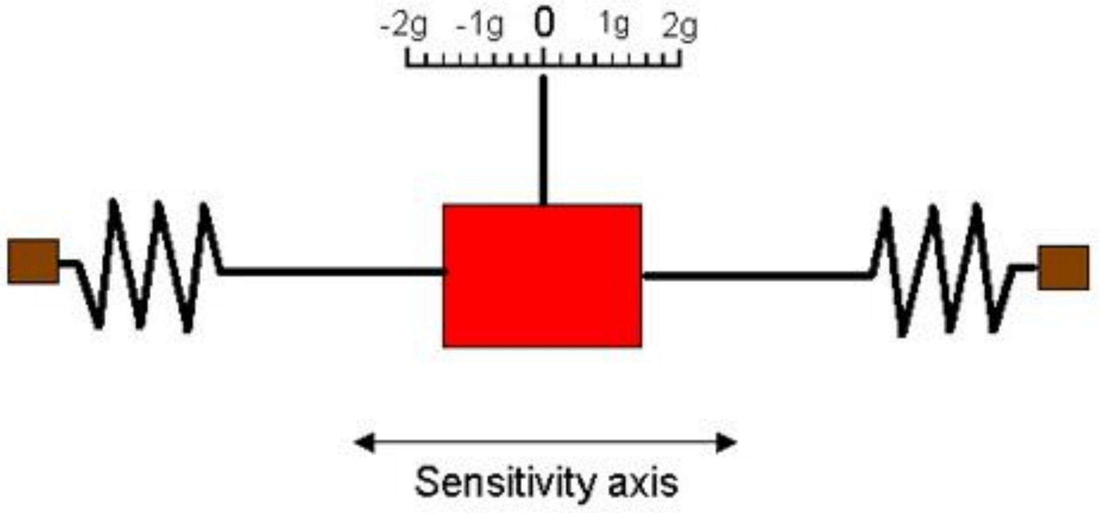


Fig. 2. Schematic structure of an accelerometer [R2015].

In Newton's day, accelerometers where built using a test mass (shown in red) held at rest with springs and having a scale showing the acceleration along the sensitivity axis. Note that the unit g is equal to the acceleration subject to all bodies at the surface of the earth due to gravity, and is equal to about 9.8 meter/second^2. The same gravity is the acceleration that translates our body mass to a weight we can measure when we stand on a scale.

In the early 1950s, accelerometers were used in inertial navigation systems, and their structure has been modernized to include an easy electronic interface, and to replace the springs with magnetic forces. In the early 1990s, a new generation of MEMS devices integrated the accelerometer into a single silicon structure.

With modern MEMS technology, the sensors are easily included in miniature electronic boards, like inRotoView.

The accelerometer can detect movement based on double integration of the measured acceleration and addition of the initial position and speed. However, since the Earth exerts a gravity acceleration on all bodies, we can also use the accelerometer to measure tilt.

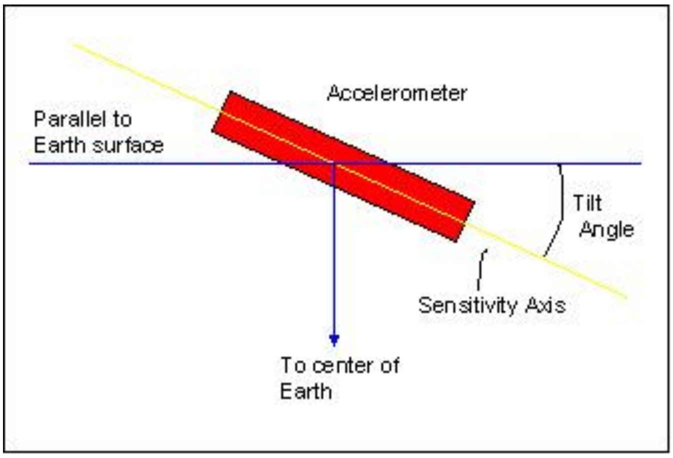


Fig. 3. The accelerometer measure tilt [R2015].

When the sensitivity axis points directly to the center of the Earth, it measures 1g (assuming no additional hand acceleration in this direction). When the accelerometer sensitivity axis lies parallel to the surface of the Earth, it measures 0 acceleration. The actual tilt angle may be inferred with the following formula:

Tilt Angle=ArcSin (measured acceleration / 1g).

When using an accelerometer to measure the tilt of a hand-held device, the movements of the hand create additional accelerations components which distort the exact calculation of the tilts. Therefore, RotoView NLDR algorithms are used to allow easy and intuitive view navigation, as you can experiment with this development system.

The gyroscope sensor is becoming more common in modern smartphones, and it complement the accelerometer with its ability to measure rotations directly.

6.2.4 The gyroscope

The gyroscope sensor measures rotational velocity along the Roll, Pitch and Yaw axes. It depends on the property of rotating mass as illustrated in the following schematic drawing of the classical mechanical gyroscope.

Gyroscope is second most popular sensor in today's smartphone, after the accelerometer sensor.

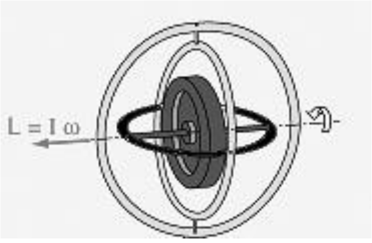


Fig. 4. Schematic structure of a gyroscope [R2015].

Of course, like the accelerometer, modern gyroscope sensors for mobile devices utilize MEMS technology contained in a tiny electronic package. The same tiny package may include both the gyroscope and accelerometer (and sometimes even the magnetometer).

Fig. 5. MEMS tiny electronic package [R2015].

6.2.5 Data Collection Tool

8.2.5.1 Sensor Kinetics app (Accelerometer)

Sensor Kinetics displays real time charts for the three axes of the accelerometer embedded in your phone. The charts can be viewed in either portrait or landscape mode.

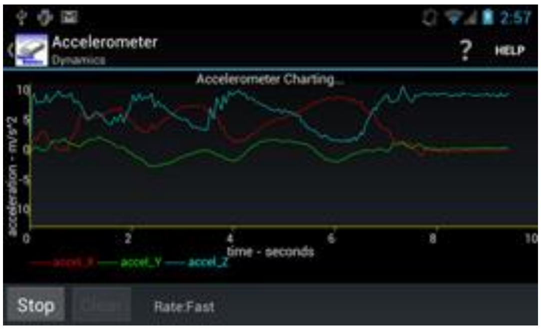


Fig. 6. Sensor Kinetics app (Accelerometer) [R2015].

Students can conduct interesting experiments while measuring accelerations and gravity effects with their phone's built in accelerometer.

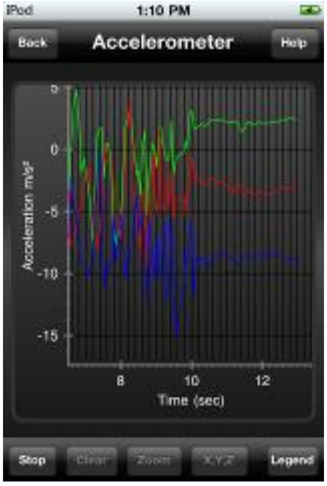


Fig. 7. Sensor Kinetics app (Accelerometer) [R2015].

The accelerometer readings are in m/s² and the are measured along the X,Y, and Z axes. It is possible to use the three axes measurement to infer rotations with Euler methods, but results are influenced by lateral movements of the accelerometer. Modern smartphones use fusion algorithm to combine results from the accelerometer, magnetometer and gyroscope to achieve precise measurements of linear acceleration, gravity and rotations.

6.2.5.1 Sensor Kinetics app (Gyroscope).

Sensor Kinetics displays real time charts for the gyroscope. It shows the rotation rates along the pitch, roll and yaw (azimuth) axes. The charts can be viewed in either portrait or landscape mode.

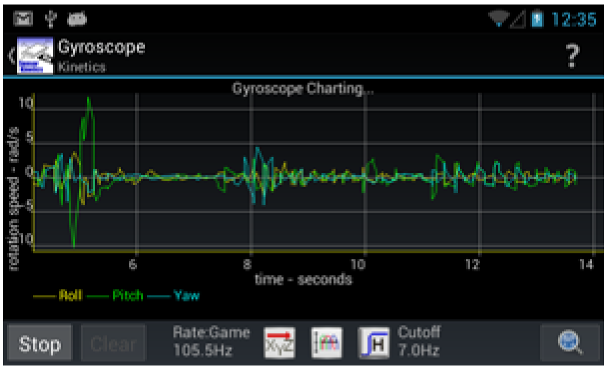


Fig. 8. Sensor Kinetics app (Gyroscope) [R2015].

Students can conduct interesting experiments while measuring rotations with their phone's built in gyroscope.

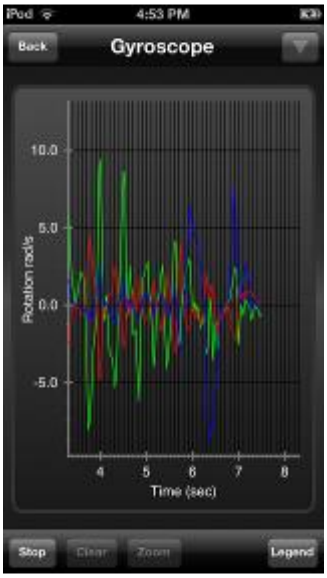


Fig. 9. Sensor Kinetics app (Gyroscope) [R2015].

The gyroscope reading in the main and summary screens are displayed in radians per second. The chart view integrates the rotational velocities to obtain graphs of the roll, pitch, and yaw of the device.

6.2.6 Experimental Design

The various hardware components available on a smartphone include: touchscreen, accelerometer, gyroscope, voltage sensor, current sensor, and battery. Each of the components has device drivers, which report sensory statistics to the kernel. For the movement modality, readings were recorded using the Sensor Kinetics Pro tool. For our analysis, we gathered movement readings from both accelerometer and gyroscope sensors.

When performing a study with volunteer participants, the results obtained depend strongly on the quality of the data collected. It is vital to understand any sources that can cause potential variance in the data for a specific user and to retain data in a uniform format using uniform devices. While our profile generation algorithms do not require such precautions, this step is needed in order to compare the datasets and evaluate the performance fairly.

To achieve uniformity of measurements, we used the same device (HTC Sensation XE Z715E) for all users who volunteered for this study. Further, all data collections were performed on Android version 4.4.4 (Build number KTU84P). Studying the effects of collecting data across different smartphone models or software versions was not attempted. We also did not use any tablet devices.

The procedure was designed in such a way so as to collect data from every participant for three sessions of ten minutes each and the collection of data was done in a uniform and controllable environment. In particular, in a pedestrian street of Athens, where the traffic was limited and there was enough space allowing the freedom of movements. Each user was asked to walk or run for ten minutes in every session with a break of 5 minutes for instructions. The users were asked to walk or run as usual, while they could interact with passers-by. All our volunteer participants were aged between lower 18s and upper 60s. Some of our participants were not regular smartphone users.

In the first session we recorded with the accelerometer and the gyroscope, one sequence of 10 minute while the participant walked and hold the device on his hand. In the second session we recorded one sequence of 10 minute while the user walked and had the device on his pocket. In the third session we recorded one sequence of 10 minute while the user was running and had the device on his pocket. In both sessions we had a maximum recording rate of 50 samples per second. These sessions were repeated for two days so as to effectively capture the biometric behavior of the user. This gives us a total of 60 minutes' real use data of the smartphone for each user.

As part of our experimental protocol, after the completion of the experimental work, the biometric patterns that were collected, will be destroyed.

6.2.7 Data Collection

We proceeded to the collection of biometrics, by taking samples of the waveforms of the signals that were recorded by sampling 50 samples per second by using the Accelerometer and Gyroscope.

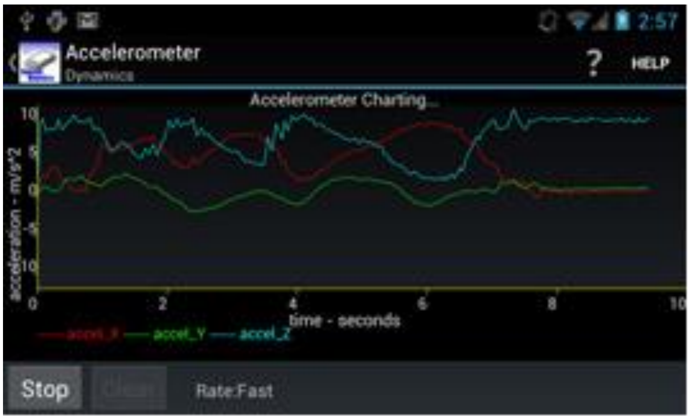
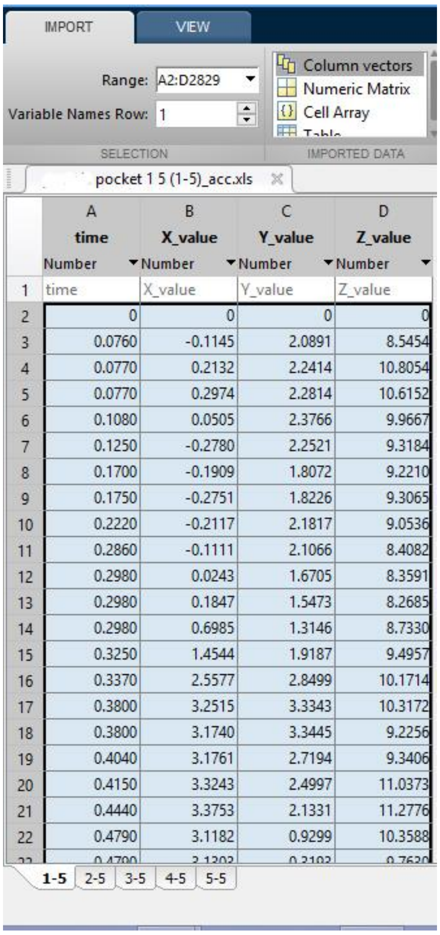


Fig. 10. The waveforms of the Accelerometer.

Then we proceeded to converting the data recorded, and the values of time and of the X,Y,Z axes were given to corresponding variables: time, X_value, Y_value, Z_value. The data were stored in CSV files and inserted to Matlab. The results are presented in figure 11.



Range: A2:D2829

Variable Names Row: 1

	A time Number	B X_value Number	C Y_value Number	D Z_value Number
1	time	X_value	Y_value	Z_value
2	0	0	0	0
3	0.0760	-0.1145	2.0891	8.5454
4	0.0770	0.2132	2.2414	10.8054
5	0.0770	0.2974	2.2814	10.6152
6	0.1080	0.0505	2.3766	9.9667
7	0.1250	-0.2780	2.2521	9.3184
8	0.1700	-0.1909	1.8072	9.2210
9	0.1750	-0.2751	1.8226	9.3065
10	0.2220	-0.2117	2.1817	9.0536
11	0.2860	-0.1111	2.1066	8.4082
12	0.2980	0.0243	1.6705	8.3591
13	0.2980	0.1847	1.5473	8.2685
14	0.2980	0.6985	1.3146	8.7330
15	0.3250	1.4544	1.9187	9.4957
16	0.3370	2.5577	2.8499	10.1714
17	0.3800	3.2515	3.3343	10.3172
18	0.3800	3.1740	3.3445	9.2256
19	0.4040	3.1761	2.7194	9.3406
20	0.4150	3.3243	2.4997	11.0373
21	0.4440	3.3753	2.1331	11.2776
22	0.4790	3.1182	0.9299	10.3588

Fig. 11. The data in Matlab.

6.3 Data Preparation

6.3.1 Feature Engineering

The records were divided into small sequences of time. We divided the sequences as follows:

- The records of the first session, while the participant walked and hold the device on his hand, were divided to 10 sequences of one minute.
- The records of the second session, while the user walked and had the device on his pocket, were divided to 10 sequences of one minute.
- The records of the third session, while the user was running and had the device on his pocket, were divided to 10 sequences of one minute.

6.3.2 Normalization

The last stage of data preparation is normalization so as to reduce the dimension of data in values between -1, 1. The data can now be introduced to machine learning algorithms.

6.4 Conclusion

In chapter 6 we proposed a method of behavioral biometric collection and particularly a method of user walk patterns collection. The use of sensor kinetics pro application was quite simple and reliable. From the results we noticed that the users’ waveforms are significantly different so we believe that the creation of a user profile based on the specific biometrics will be possible.

7

Conclusions

7.1 Summary and Findings

7.1.1 First Survey Conclusions

In the results of the first survey we saw the users' attitudes for the protection of their Personal Data and how these are affected by factors such as age. In the first hypothesis of the survey, on if the users store important PIN and password on their device, we observed that there is a statistically significant correlation. Younger people, mostly of the age 18-30, seem to be more careful for ensuring their personal data. More specifically, in the age group 18-24 they avoid storing on their device bank Pins and important passwords at a percentage of 74,2%, while in the age 25-30 at a percentage of 81,1%. These age groups encrypt their data at a percentage of 13,5 and 16,2 respectively. The age of 31-35 is on the borderline, since it is the median in the third category (yes, without encryption) of the first case of the survey. At this age it seems that they do not store PIN and passwords on their device. From the median and above though, i.e. at ages 36- 50, as is apparent from the descriptive statistics, the users do store important password such as Bank passwords, with no encryption. At the same time, even though all those sensitive data are exposed, 77.9% of the respondents never change their PIN in their cash card. Generally in all age groups of this category the percentage of users who do not follow any practices in order to protect their Pin and Passwords is about 19,1%.

In the second hypothesis of our first survey we also saw that the age factor affects the users' attitude in relation to the storage of sensitive personal data on their device. We already have seen in the results of our survey, 81,9% of the users stores sensitive personal data on their mobile device. Moreover, most of the participants, 28%, have lost their device, at least once. Finally, they give their Pin in third persons in a percentage of 22,5% and, in the third hypothesis, we saw that females give their PIN to third persons more easily.

Answering to our research question which is: "Can the users' practices protect their sensitive data?" we observe, that generally the users' majority is interested in the protection of their personal data. The younger age groups seem to take some extra steps for their protection that do not appear at older ages. But the measures taken by the users in general are not sufficient to protect them. Most of them for example protect their personal data by a PIN and they use a PIN or a touch gesture in order to protect individual elements (such as photographs, sms, telephone directory etc.) preventing access to third parties. But as it emerged from the literature review the use of current PIN-based authentication or touch gestures is problematic [CF2005], [AEG+2007], [KS2010], [CFS+2012], [KTH+2013], because there is no protection after the Pin is entered. In addition, it is not sufficient since the devices are vulnerable to smudge attacks [AGM+2010].

The results of this study, show that while many smartphone users do take some security measures, a high percentage 24% of them still ignore potential risks.

From all the above results we firmly believe that there is a need for the amplification of users' personal data protection via a Continuous Authentication System with biometrics & Behavioral modalities. Besides, a great number of studies, as the one of Clarke et al. [CFR + 2002], presented their findings on the views of the subscribers concerning the need for security in mobile devices. The users were positive to alternative identity control methods, such as the fingerprint scanning and the voice recognition. In addition, the results of a survey which was also conducted by Clarke et al. [CF2005], showed that 83% of the participants are willing to accept some form of biometric Authentication on their device.

7.1.2 Second Survey Conclusions

Many ways of user's biometrics behavioral Authentication have been proposed. There is a large corpus of published research works that individually use behavioral modalities [SKK2012], [SCF2008]. Shi et al. [SNJC2011] presented an approach that was built on the concept that most users are habitual in nature and are prone to performing similar tasks at a certain time of the day. The researchers collected a wide range of behavioral information such as location, communication, and usage of applications, in order to create a user profile. Our survey confirms Shi’s observation that users are prone to performing similar tasks at a certain time of the day and that in this way a general users’ profile could be created but with some additional parameters analyzed below. The results of our second survey have provided strong evidence that every user has his own distinctive and unique way of use of his mobile devices, regardless of the general similar tasks seen above, as well as, the age-specific trends we have elaborated upon in our analysis. For instance, it turns out that there is no example of distinct users that use exactly the same set of applications at exactly the same times of the day, with no single differentiation between them. In addition, users also differentiate in their use of applications or the magnitude of use. Differentiations exist even in the way that data are stored e.g., by using dropbox or by storing them in the phone’s memory, as well as in the writing language, in the way of search, in the number of SMS and phone calls, in the order of priority and in the way they invoke the applications, and whether they let them “run” at the same time. These observations provide strong evidence that the behavior of each user can be profiled on the basis of their application usage patterns; on this basis, we believe user Authentication will be feasibly effected via a System of Continuous Authentication.

7.1.3 Conclusions of Experiment Procedure and Future Research

In chapter 6 we proposed a method of behavioral biometric collection and particularly the users’ walk patterns. The session recorded three sequences of 10 minute each while the participant: walked and hold the device on his hand, walked and had the device on his pocket, was running and had the device on his pocket. The use of the sensor kinetics pro application was quite simple and reliable. From the results we noticed that the users’ waveforms differ significantly so we believe that the creation of a user profile based on the specific biometrics will be possible. This of course we will be able to confirm it through our further research.

In the near future we will create a multi-modal Continuous Authentication Model. It will be based on the modalities we searched in the present thesis and in biometrics relating to User's Walk Patterns that were collected in the present thesis as well. The Continuous Authentication model Using User's Walk Patterns is the first that we started creating.

8

Bibliography

[AT2011]. Ahmed Awad E. Ahmed, Issa Traore. Continuous Authentication Using Biometrics: Data, Models and Metrics. Publisher: IGI Global. ISBN: 9781613501290. Release Date: September 2011.

[ALZ2014] Androulidakis, I., Levashenko, V., Zaitseva, E.: Smart phone users: Are they green users? 10th International Conference on Digital Technologies. IEEE (DT2014).

[ACBS2009] Androulidakis, I., Christou, V., Bardis, N., Stilios, I.: Surveying users' practices regarding mobile phones' security features, Electrical And Computer Engineering Series, Proceedings of the 3rd WSEAS international conference on European computing conference table of contents, WSEAS 2009.

[BZJ+2014] Bo, C., Zhang, L., Jung, T., Han, J., Li, X.-Y., Wang, Y.: Continuous user identification via touch and movement behavioral biometrics. Performance Computing and Communications Conference (IPCCC), 2014 IEEE International. pp. 1{8.IEEE (2014).

[CF2005] Clarke N., L., Furnell, S., M.: Authentication of users on mobile telephones – A survey of attitudes and practices. Computers & Security (2005) 24, 519e527, Elsevier.

[CFR+2002]. N.L. Clarke, S.M. Furnell, P.M. Rodwell, P.L. Reynolds. Acceptance of Subscriber Authentication Methods For Mobile Telephony Devices. Computers & Security Volume 21, Issue 3, 1 June 2002, Pages 220–228.

[KFC2007]. S.Karatzouni, S.M.Furnell, N.L.Clarke and R.A.Botha. Perceptions of User Authentication on Mobile Devices. In Proceedings of the 6th Annual ISOnEworld Conference, April 11-13, 2007, Las Vegas, NV.

[D1999] Dillman, D. A. Mail and Internet Surveys: The Tailored Design Method, John Wiley & Sons, 2nd edition, November 1999.

[FBM+2013] Frank, M., Biedert, R., Ma, E., Martinovic, I., Song, D.: Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. Information Forensics and Security, IEEE Transactions on. 8, 136{148 (2013).

[FCKB2008] Furnell, S., Clarke, N., Karatzouni, S.: Beyond the PIN: Enhancing user authentication for mobile devices. Computer Fraud & Security, Volume 2008, Issue 8, August 2008,Elsevier.

[SCF2008] Saevanee, H., Clarke, N., Furnell, S,. M.: "Multi-modal behavioural biometric authentication for mobile devices, " in Information Security and Privacy Research, ser. IFIP Advances in Information and Communication Technology, D. Gritzalis, S. Furnell, and M. Theoharidou, Eds. Springer Berlin Heidelberg, 2012, vol. 376, pp. 465-474.

[SKK2012] Seo, H., Kim, E. and Kim, H., K.: "A novel biometric identification based on a users input pattern analysis for intelligent mobile devices, "International Journal of Advanced Robotic Systems, 2012.

[SNJC2011] Shi, E., Niu, Y., Jakobsson, M., Chow, R.: Implicit authentication through learning user behavior. Information Security. pp. 99{113. Springer (2011).

[G2007] G. A. v. Graevenitz, "Biometric Authentication in Relation to Payment Systems and ATMs," Datenschutz und Datensicherheit, vol. 31, no. 9, pp. 681-683, 2007.

[I2015] Iritech, inc, "Biometric identification Boosting Automotive Security," 18 September 2015. Online Available: http://www.iritech.com/blog/biometric-automotive-0915/. [Accessed 08 January 2016].

[S22015]"List of all Eye Scanner (Iris, Retina Recognition) Smartphones," 08 December 2015. Online Available: http://webcusp.com/list-of-all-eye-scanner-iris-retina-recognition-smartphones/. [Accessed 08 January 2016].

[S2015] Shams, "List of All Fingerprint Scanner Enabled Smartphones," 25 December 2015. Online Available: http://webcusp.com/list-of-all-fingerprint-scanner-enabled-smartphones/. [Accessed 08 January 2016].

[O2015] R. O'Neil, "Mobile Biometrics Market Analysis," Biometrics Research Group, Inc, 2015.

[K2015] S., Kokolakis, 2015. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. Computers & Security.

[K2015] Kokolakis, S., 2015. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. Computers & Security.

[ACBS2009] Androulidakis, I., Christou, V., Bardis, N., Stilios, I., (2009). Surveying users' practices regarding mobile phones' security features, Electrical And Computer Engineering Series, Proceedings of the 3rd WSEAS international conference on European computing conference table of contents, WSEAS.

[K2001] Kumar, N., (2011). Password In Practice: An Usability Survey. Journal of Global Research in Computer Science. Volume 2, No. 5.

[TAHO2009] Tamviruzzaman, M., Ahamed, S., I., Hasan, C., S., and Obrien, C., (2009). Epet: When cellular phone learns to recognize its owner. In Proceedings of the 2Nd ACM Workshop on Assurable and Usable Security Configuration, ser. SafeConfig 09, pp. 13-18.

[CF2005] Clarke N., L., Furnell, S., M., (2005). Authentication of users on mobile telephones – A survey of attitudes and practices, Computers & Security 24, 519e527, Elsevier.

[FC2008] Furnell, S., Clarke, N., Karatzouni, S., (2008). Beyond the PIN: Enhancing user authentication for mobile devices. Computer Fraud & Security, Volume 2008, Issue 8, Elsevier.

[SCF2012] Saevanee, H., Clarke, N., Furnell, S,. M., 2012. Multi-modal behavioural biometric authentication for mobile devices. In Information Security and Privacy Research, ser. IFIP Advances in Information and Communication Technology, D.

Gritzalis, S. Furnell, and M. Theoharidou, Eds. Springer Berlin Heidelberg, vol. 376, pp. 465-474.

[SP2012] Sujithra, M., Padmavathi, G., (2012). A Survey on Mobile Device Threats, Vulnerabilities and their Defensive Mechanism. International Journal of Computer Applications (0975 – 8887) Volume 56– No.14.

[SSM2009] Shye, A., Scholbrock, B., Memik, G., (2009). Into the wild: studying real user activity patterns to guide power optimizations for mobile architectures. Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture. pp. 168{178.

[FBM+2013] Frank, M., Biedert, R., Ma, E., Martinovic, I., Song, D., (2013). Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. Information Forensics and Security, IEEE Transactions on. 8, 136{148.

[BZJ+2014] Bo, C., Zhang, L., Jung, T., Han, J., Li, X.-Y., Wang, Y., (2014): Continuous user identification via touch and movement behavioral biometrics. Performance Computing and Communications Conference (IPCCC), IEEE

[CF2005]. N.L. Clarke, S.M. Furnell. 2005. Authentication of users on mobile telephones – A survey of attitudes and practices. Computers & Security Volume 24, Issue 7, October 2005, Pages 519–527. ELSEVIER.

[AEG+2007]. Shane Ahern, Dean Eckles, Nathaniel S. Good, Simon King,Mor Naaman, Rahul Nair. 2007. Over-exposed?: privacy patterns and considerations in online and mobile photo sharing. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Pages 357-366. Publisher ACM New York, USA.

[KS2010]. Stan Kurkovsky, Ewa Syta. 2010. Digital natives and mobile phones: A survey of practices and attitudes about privacy and security. 2010 IEEE International Symposium on Technology and Society (ISTAS). 7-9 June 2010. Conference Location : Wollongong, NSW. Page(s): 441 - 449. Print ISBN: 978-1-4244-7777-7.

[AGM+2010]. Aviv, A.J., Gibson, K., Mossop, E., Blaze, M., Smith, J.M.: Smudge attacks on smartphone touch screens. Proceedings of the 4th USENIX conference on Offensive technologies. pp. 1{7. USENIX Association (2010).

[CFS+2012]. Erika Chin, Adrienne Porter Felt, Vyas Sekar, David Wagner. 2012. Measuring user confidence in smartphone security and privacy. Proceedings of the

Eighth Symposium on Usable Privacy and Security. Article No. 1. ISBN: 978-1-4503-1532-6NY, Publisher ACM New York, USA.

[KTH+2013]. Mark J. Keith, Samuel C. Thompson, Joanne Hale, Paul Benjamin Lowry, Chapman Greer. 2013. Information disclosure on mobile devices: Re-examining privacy calculus with actual user behavior. International Journal of Human-Computer Studies Volume 71, Issue 12, December 2013, Pages 1163–1173. ELSEVIER.

[KCH2010]. Dong-Ju Kim, Kwang-Woo Chung, Kwang-Seok Hong. 2010. Person authentication using face, teeth and voice modalities for mobile device security. Published in: IEEE Transactions on Consumer Electronics (Volume:56 , Issue: 4). Page(s): 2678 - 2685. ISSN : 0098-3063. Date of Publication : November 2010.

[FBM+2013]. Frank, M., Biedert, R., Ma, E., Martinovic, I., Song, D.: Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. Information Forensics and Security, IEEE Transactions on. 8, 136{148 (2013).

[FZS2013]. Tao Feng, Xi Zhao ; Weidong Shi. 2013. Investigating Mobile Device Picking-up motion as a novel biometric modality. Published in: IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS). Page(s): 1 - 6. Conference Location : Arlington, VA.. Date of Conference: Sept. 29 2013-Oct. 2 2013.

[BZJ+2014]. Bo, C., Zhang, L., Jung, T., Han, J., Li, X.-Y., Wang, Y.: Continuous user identification via touch and movement behavioral biometrics. Performance Computing and Communications Conference (IPCCC), 2014 IEEE International. pp. 1{8. IEEE (2014).

[KWM2010]. Kwapisz, J.R., Weiss, G.M., Moore, S.A.: Cell phone-based biometric identification. Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on. pp. 1{7. IEEE (2010).

[SNJ+2011]. Shi, E., Niu, Y., Jakobsson, M., Chow, R.: Implicit authentication through learning user behavior. Information Security. pp. 99{113. Springer (2011).

[RQS+2012]. Riva, O., Qin, C., Strauss, K., Lymberopoulos, D.: Progressive authentication: deciding when to authenticate on mobile phones. Proceedings of the 21st USENIX Security Symposium (2012).

[CF2007]. Clarke, N.L., Furnell, S.M.: Advanced user authentication for mobile devices. Computers & Security. 26, 109{119 (2007).

[ZTQ+2010]. Zhang, L., Tiwana, B., Qian, Z., Wang, Z., Dick, R.P., Mao, Z.M., Yang, L.: Accurate online power estimation and automatic battery behavior based power model generation for smartphones. Proceedings of the eighth IEEE/ACM/IFIP international conference on Hardware/software codesign and system synthesis. pp. 105{114. ACM (2010).

[MMS+2012]. Murmuria, R., Medsger, J., Stavrou, A., Voas, J.M.: Mobile Application and Device Power Usage Measurements. 2012 IEEE Sixth International Conference on Software Security and Reliability (SERE). pp. 147{156 (2012).

[SSM2009]. Shye, A., Scholbrock, B., Memik, G.: Into the wild: studying real user activity patterns to guide power optimizations for mobile architectures. Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture. Pp. 168{178. ACM (2009).

[MSB+2015]. Rahul Murmuria , Angelos Stavrou, Daniel Barbará, Dan Fleck. Continuous Authentication on Mobile Devices Using Power Consumption, Touch Gestures and Physical Movement of Users. Research in Attacks, Intrusions, and Defenses. Volume 9404 of the series Lecture Notes in Computer Science. pp 405-424. Date: 12 December 2015.

[MBK+2013]. Muslukhov, I., Boshmaf, Y., Kuo, C., Lester, J., Beznosov, K.: Know your enemy: the risk of unauthorized access in smartphones by insiders. Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services. pp. 271{280. ACM (2013).

[KBS2009]. Karlson, A.K., Brush, A.J., Schechter, S.: Can I borrow your phone?: understanding concerns when sharing mobile phones. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1647{1650. ACM (2009).

[JBP2006] A. Jain, R. Bolle and S. Pankanti, Biometrics: Personal Identification in Networked Society, New York: Springer, 2006.

[G1892] F. Galton, Finger Prints, London: McMillan, 1892.

[RC2015] Research Capsule, "Fingerprint Sensors Market in Smart Mobile Devices 2012-2019," Research Capsule, Inc., 2015.

[GAO2002] GAO, "Technology Assessment Using Biometrics for Border Security," DIANE Publishing, 2002.

[EM2013] N. Erdogmus and S. Marcel, "Spoofing 2D Face Recognition Systems with 3D Masks," in Idiap Research Institute, Darmstadt, 2013.

[SE2011]. Strickland, "Can Biometrics ID an Identical Twin?," 11 March 2011. [Online]. Available: http://spectrum.ieee.org/computing/software/can-biometrics-id-an-identical-twin. [Accessed 07 January 2016].

[S1974]. Stephens, M. A. (1974). "EDF Statistics for Goodness of Fit and Some Comparisons".Journal of the American Statistical Association (American Statistical Association) 69 (347): 730–737. doi:10.2307/2286009. JSTOR 2286009.

[MTW2003]. Marsaglia G, Tsang WW, Wang J (2003). "Evaluating Kolmogorov's Distribution"Journal of Statistical Software 8 (18): 1–4.

[K1933]. Kolmogorov A (1933). "Sulla determinazione empirica di una legge di distribuzione".G. Ist. Ital. Attuari 4: 83–91.

[S1948]. Smirnov N (1948). "Table for estimating the goodness of fit of empirical distributions".Annals of Mathematical Statistics 19: 279–281. doi:10.1214/aoms/1177730256.

[PH1972]. Pearson, E. S. and Hartley, H. O., eds. (1972). Biometrika Tables for Statisticians 2. Cambridge University Press. pp. 117–123, Tables 54, 55. ISBN 0-521-06937-8.

[LH60] Levene, Howard (1960). "Robust tests for equality of variances". In Ingram Olkin, Harold Hotelling, et alia. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford University Press. pp. 278–292.

[SW1986]. Shorack, Galen R.; Wellner, Jon A. (1986). Empirical Processes with Applications to Statistics. Wiley. p. 239. ISBN 047186725X.

[AE2011]. Arnold, Taylor B.; Emerson, John W. (2011). "Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions" (PDF). The R Journal 3 (2): 34–39. [SAS STATISTICS].
http://support.sas.com/documentation/cdl/en/statug/68162/HTML/default/viewer.htm #statug_npar1way_toc.htm

[STATA]. stata.com- ksmirnov — Kolmogorov–Smirnov equality-of-distributions test

[M2014]. Mehta, S. (2014) Statistics Topics ISBN 978-1499273533.

[JPZ1997]. Justel, A.; Peña, D.; Zamar, R. (1997). "A multivariate Kolmogorov–Smirnov test of goodness of fit". Statistics & Probability Letters 35 (3): 251–259. doi:10.1016/S0167-7152(97)00020-5.

[P19863]. Peacock J.A. (1983). "Two-dimensional goodness-of-fit testing in astronomy". Monthly Notices of the Royal Astronomical Society 202: 615–627.Bibcode:1983MNRAS.202..615P. doi:10.1093/mnras/202.3.615.

[FF1987]. Fasano, G., Franceschini, A. (1987). "A multidimensional version of the Kolmogorov–Smirnov test". Monthly Notices of the Royal Astronomical Society 225: 155–170. Bibcode: 1987MNRAS.225.155F. doi:10.1093/mnras/ 225.1.155. ISSN 0035-8711.

[LRH2007]. Lopes, R.H.C., Reid, I., Hobson, P.R. (April 23–27, 2007). The two-dimensional Kolmogorov–Smirnov test (PDF). XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research. Amsterdam, the Netherlands.

[LS15]. Laerd Statistics. Kruskal-Wallis H Test using SPSS Statistics, Laerd Statistics 2015.

[KW1952]. Kruskal; Wallis (1952). "Use of ranks in one-criterion variance analysis". Journal of the American Statistical Association 47 (260): 583–621. doi:10.1080/01621459.1952.10483441.

[CF2009]. Corder, Gregory W.; Foreman, Dale I. (2009). Nonparametric Statistics for Non-Statisticians. Hoboken: John Wiley & Sons. pp. 99–105. ISBN 9780470454619.

[SC1988]. Siegel; Castellan (1988). Nonparametric Statistics for the Behavioral Sciences (Second ed.). New York: McGraw–Hill. ISBN 0070573573.

[D1964]. Dunn, Olive Jean (1964). "Multiple comparisons using rank sums". Technometrics 6 (3): 241–252. doi:10.2307/1266041.

[S2003]. Spurrier, J. D. (2003). "On the null distribution of the Kruskal–Wallis statistic". Journal of Nonparametric Statistics 15 (6): 685–691. doi:10.1080/10485250310001634719.

[MS2006]. Meyer, Seaman (2006). "Expanded tables of critical values for the Kruskal-Wallis H statistic". Paper presented at the annual meeting of the American

Educational Research Association, San Francisco. Available for download at http://faculty.virginia.edu/kruskal-wallis

[B2015] I, Biperis, 2015. Information Systems Security. Available for download at http://goo.gl/qsd0M1

[R2015] RotoView Group, 2015. Sensor Kinetics app. Available on: http://www.rotoview.com/sensor_kinetics.htm

9

Appendix

9.1 Questionnaire Chapter 6:

9.1.1 Demographics

- Gender: Male (M), Female (F)
- Age: (A: 18-24, C: 25-30, D: 31-35, E: 36-40, F: 41-45, G: 46-50, H: 51- 55, I: 56-60)
- Studies: (A: Humanities, B: Medical Sciences, C: Law, D: Engineering-Computers, E: Positive Sciences, F: Economics-Business Management, G: Other)
- Average monthly bll: (A<=10 Euro, B: 11-20, C: 21-30, D: 31-40, E: >40)
- Type of device? *Android, iphone*

9.1.2 Storage Practices

- Do you store sensitive personal data on your mobile device? (e.g. photographs/videos/conversations’ recordings etc.).

- Do you store important passwords on your mobile device? (e.g. Bank passwords, Alarm passwords etc.)

9.1.3 Pin Practices

- Have you activated the PIN question on your SIM card?
- How often do you change the PIN question on your mobile device?
- Do you have a password on your mobile's phone Screen-Saver and how often do you change it?
- Do you protect sensitive applications with a pin or touch gestures?
- How often do you change the PIN on your cash card?
- Do you give your PIN to third persons?

9.1.4 Device Protection

- Have you ever lost your phone or has it ever been stolen?
- Have you ever forgotten your device e.g. at a coffee shop?

9.2 Questioner Chapter 7:

9.2.1 Demographics

- Gender: Male *(M) Female (F)*
- Age? *(A: 18-24, C: 25-30, D: 31-35, E: 36-40, F: 41-45, G: 46-50, H: 51- 55, I: 56-60)*
- Studies: *(A: Humanities, B: Health Sciences, C: Law, D: Sciences of Engineers, E: Positive Sciences, F: Economics and Management Sciences, G: Other, H: Environmental Sciences)*
- Average monthly bill? *(A<=10 Euro,, B: 11-20, C: 21-30, D: 31-40, E: >40)*
- Type of device? *Android, iphone*
- Have you ever lost your mobile phone, or has it ever been stolen?",
- Do you save sensitive personal data on your mobile phone?
- If Yes, do you encrypt them?

9.2.2 The application's way of use

- Choose the applications that you use: facebook, google, e-mail, linkedin, youtube

The time of the day?	*morning*	*noon*	*afternoon*	*evening*
facebook				
google				
e-mail				
linkedin				
youtube				

How many hours per day?	*1*	*2*	*3*	*4*	*5*
facebook					
google					
e-mail					
linkedin					
youtube					

- Do you use dropbox?
- Do you use messenger?
- Do you write in greeklish?, Greek?, English?
- Do you have priorities in the use of applications? (e.g. first e-mails afterwards facebook)
- Do you open your e-mail by using the application' s icon or by the notifications?
- Do you leave many applications "running" at the same time?
- Do you use "search" by using the microphone?
- How many sms do you send per day?
- How many phone calls do you make per day?

9.2.3 Power Consumption

1. How long does the battery of your mobile phone last with no use?

2. How many hours does it last with your usual use? 5, 10, 15, 20, 24, 30
3. Do you leave your mobile phone ‘on’ during night?
4. Do you put your device on ‘flight mode’ during night?
5. Do you turn off the ‘3G’ when you are not connected?
6. Do you turn the ‘wifi’ off when you are not connected?
7. Do you turn the ‘GPS’ off?

9.2.4 Touch Gestures

1. Do you press with strength the screen of your mobile phone when it is not responding
2. Do you shake your mobile phone when it is not responding quickly?
3. Do you usually wipe the screen of your mobile phone after calling?

9.2.5 Guest Users:

1. Do you give your mobile phone to ‘guest users’?
2. If Yes, does the use by others happen at specific hours or days?
3. Are guest users people you trust or close to you?