Security and Usability of a Personalized User Authentication Paradigm: Insights from a Longitudinal Study with Three Healthcare Organizations

ARGYRIS CONSTANTINIDES

University of Cyprus, Cyprus & Cognitive UX LTD, Cyprus

MARIOS BELK

University of Cyprus, Cyprus & Cognitive UX GmbH, Germany

CHRISTOS FIDAS

University of Patras, Greece

ROY BEUMERS

Zuyderland Medical Center, Netherlands

DAVID VIDAL

Hospital Clinic Barcelona, Spain

WANTING HUANG

Accenture B.V., Netherlands

JULIANA BOWLES

University of St. Andrews, UK

THAIS WEBBER

University of St. Andrews, UK

AGASTYA SILVINA

University of St. Andrews, UK

ANDREAS PITSILLIDES

University of Cyprus, Cyprus & University of Johannesburg, South Africa (Visiting Professor)

This paper proposes a user-adaptable and personalized authentication paradigm for healthcare organizations, which anticipates to seamlessly reflect patients' episodic and autobiographical memories to graphical and textual passwords aiming to improve the security strength of user-selected passwords and provide a positive user experience. We report on a longitudinal study that spanned over three years in which three public European healthcare organizations participated in order to design and evaluate the aforementioned paradigm. Three studies were conducted (n=169) with different stakeholders: *i)* a verification study aiming to identify existing authentication practices of the three healthcare organizations with diverse stakeholders (n=9); *ii)* a patient-centric feasibility study during which users interacted with the proposed authentication system (n=68); and *iii)* a human guessing attack study focusing on vulnerabilities among people sharing common experiences within location-aware images used for graphical passwords (n=92). Results revealed that the suggested paradigm scored high with regards to users' likeability, perceived security, usability and trust, but more importantly it assists the creation of more secure passwords. On the downside, the suggested paradigm introduces password guessing vulnerabilities by individuals sharing common experiences with the end-users. Findings are expected to scaffold the design of more patient-centric knowledge-based authentication mechanisms within nowadays dynamic computation realms.

2691-1957/2022/1-ART1 $15.00
http://dx.doi.org/10.1145/3564610

CCS CONCEPTS • Applied computing~Life and medical sciences~Health care information systems • Human-centered computing~Human computer interaction (HCI)~Empirical studies in HCI

Additional Keywords and Phrases: Knowledge-based User Authentication, Graphical Passwords, Usability, Security, Feasibility User Study, Human Guessing Attack Study.

1 INTRODUCTION

User authentication is an essential security task within modern healthcare systems, which is performed daily by millions of patients across the world. The healthcare domain entails intrinsic characteristics and requirements, which position the user authentication process in a unique perspective. This is mainly accredited to the fact that healthcare organizations deploy different policies for a variety of end-user categories (*e.g.*, medical staff, patients, external caregivers, etc.), depending on the context of use, as well as the users' profiles. Given that sensitive information can be accessed online from patients and shared among medical care staff, healthcare environments increase the risk for information leaks and entail several challenges from a security, privacy and legal perspective [16, 31, 55, 93].

In order to safeguard health-related information, organizations deploy a variety of user authentication schemes, which can be either based on: *i)* a secret known by the user *(knowledge-based authentication)*, such as, a textual password, graphical password [13, 87, 109]; *ii)* a specific object owned by the user *(token-based authentication)*, such as, a smart card, smartphone, hardware token [39, 67, 68, 73, 77] ; *iii)* specific biometric information about the user *(biometric-based authentication)*, such as, her/his fingerprint, face, voice, physiological signals [14, 103]; and *iv)* a combination of the aforementioned factors *(multi-factor authentication)* [84].

Nonetheless, the aforementioned authentication schemes entail varying strengths and weaknesses with regards to security, privacy and usability from the end-user's perspective [15], but as well, with regards to costs and maintenance aspects from the healthcare organization's perspective. For example, current password policies create user frustration (*e.g.*, when users forget the passwords, or reset the passwords frequently due to strict policies) [32, 51, 98], biometric authentication schemes entail privacy threats (*e.g.*, biometric data could be used in impersonation attacks, the data cannot be revoked in case they are compromised or leaked) [43, 101], unusable password policies increase maintenance costs (*e.g.*, password resets increase labor costs of an organization) [98, 104], and in the case of password data breaches, such events negatively affect the organization's reputation and trust, lead to penalties by the corresponding health agencies [38], and may even threaten human life [31].

Furthermore, the literature reveals that healthcare organizations still rely on knowledge-based user authentication (*e.g.*, passwords), and research suggests that it will continue to prevail in the next decades [70] even in combination with other approaches (*e.g.*, token-based, biometric-based). In this respect, researchers have attempted to provide alternative knowledge-based authentication methods through graphical authentication, in which users either draw a secret gesture on the screen *(drawmetric authentication)* or select regions of images *(locimetric authentication)* [13, 28, 88]. Locimetric user authentication approaches have gained popularity in recent years, with popular examples including Android's pattern lock for unlocking smartphone devices, and Microsoft's Windows 10™ Picture Gesture Authentication (PGA) for unlocking conventional computers. Graphical user authentication research is

motivated as follows: *i)* it leverages on the fact that visual information is better recalled than textual information according to the picture superiority effect [80]; and *ii)* it can be easily adapted to ubiquitous environments due to its natural user interaction by clicking or drawing on regions of an image [42].

In this context, bearing in mind that: *i)* user interactions in the healthcare domain are characterized by high security standards; *ii)* users prefer seamless authentication policies, which should be able to adapt to different technology and contextual factors [12, 25, 28, 73]; and *iii)* authentication policies of textual passwords have become non-user friendly and hence non-secure [98], there is an urgent need to elaborate on novel user authentication paradigms that improve the current state-of-the-art. Therefore, our work is primarily driven by our vision to increase security of user-selected secrets and simultaneously provide a positive user experience through a seamless and adaptable user authentication paradigm, which will allow to transition from current "one-size-fits-all" authentication systems to user-adaptable and personalized authentication systems.

In this paper we aim to introduce a novel patient-centric authentication paradigm, coined *DuoPass*, which has been derived based on a longitudinal study that spanned over three years during which we verified, implemented, and evaluated the suggested approach in the healthcare domain, by following a User-Centered Design (UCD) approach. **Table A1** in the Appendix depicts our research methodology. We next present a literature review on state-of-the-art user authentication practices in healthcare environments, which is subsequently triangulated with diverse stakeholders of three European healthcare organizations. Consequently, we elaborate on the conceptual design of the suggested paradigm. We then present the user evaluation of DuoPass in terms of security strength, memorability, and user experience. Then, we report on a human guessing attack study that investigated vulnerabilities among people sharing common experiences within location-aware images used for graphical passwords. We conclude the paper with a discussion on the main findings, implications, and limitations of this work.

2 USER AUTHENTICATION RESEARCH AND PRACTICE IN HEALTHCARE ORGANIZATIONS

2.1 Literature Review

2.1.1 Search Strategy, Paper Selection and Eligibility Criteria

We examined papers from the ACM Digital Library and IEEE Xplore and used the following keywords in our queries: authentication; password; biometric; locimetric; drawmetric; healthcare; health. We have reviewed forty papers based on our inclusion criteria, which have been published during the period: 01/01/2015-01/06/2021.

2.1.2 Review Outcome

Several literature reviews on user authentication research and practices in the healthcare domain exist, *e.g.*, Jayabalan and O'Daniel [56], and Fernández-Alemán et al. [41] reported a study on authentication factors in Electronic Health Records; Mason et al. [74], Fatima et al. [40], and Okoh and Awad [79] reviewed the current state in biometric authentication in healthcare environments; Schwartze et al. [90]

conducted a systematic literature review on authentication systems for securing clinical documentation workflows; and Kumar et al. [65] that conducted a review on user authentication in gadget-free healthcare environments.

The healthcare domain embraces unique constraints and characteristics [64] that are related to different access control scenarios that need to be supported not only from a healthcare staff perspective but also from a patient-centric approach. Starting from the medical staff (*e.g.*, doctors, nurses, caregivers, etc.), there are numerous scenarios in which stakeholders interact with medical systems that are deployed on heterogeneous devices and within different contexts of use [38, 108]. For example, when doctors visit patients within the hospital, they typically have access to the patients' records through a smartphone or tablet device [7], whereas accessing more controlled places like surgery rooms, intensive care rooms, etc., these necessitate a multi-factor and/or biometric-based authentication approach. From a patients' perspective, access to services and data is limited to a Web-based solution in which patients access their personal health records using a textual password [38], in combination with a one-time password as a second layer for authentication. These authentication methods are analyzed in the next section under the following perspectives: security, privacy, usability, memorability, user experience, user acceptance, and trust.

The majority of healthcare organizations currently employ traditional textual password solutions [61, 63, 110]. However, textual password schemes have known security issues and are constantly becoming less usable and less memorable due to strict password policies [38, 98]. In order to address memorability issues of using multiple textual passwords in different services within the hospital, healthcare organizations deploy Single Sign-On (SSO) solutions that allow the end-users to enter their password credentials once in the beginning, and then access several independent services within their organization, without requiring them to re-enter their credentials. However, studies have shown that while SSO is effective in administrative contexts of medical staff, it creates difficulties in collaborative contexts due to the strict password policies of the healthcare organization, *e.g.*, medical staff need to frequently logout and login when they change location in the hospital, or after a small period of inactivity [38, 49, 76]. For example, according to Heckle and Lutters [49], given that the clinical staff are frequently changing locations in the hospital when taking care of patients, they are required to continuously login and logout from the system (*e.g.*, due to system timeouts, different access control depending on the system used in the corresponding location, walking away from the screen, etc.). In addition, studies have shown that medical staff may typically utilize SSO features within their network, however, they also utilize different textual password credentials for accessing systems that are off the network of their healthcare organization.

Furthermore, the literature reveals other proposals for improving password security and usability, *e.g.*, through mutual authentication or two-way authentication [52, 66], group-based authentication for assisting the access and sharing of electronic health records among trusted members [69], approaches suggesting practical recommendations for the creation of strong and usable passwords that combine minimum-strength, minimum-length and blocklist requirements [98], providing guidance and feedback during password creation [91], and proposing alternative mechanisms, such as, graphical user authentication schemes, which require from users to draw secret gestures on an image, or select a sequence of images as their secret key [1, 11, 13, 28]. In addition, healthcare environments entail unique

constraints and characteristics with regards to the end-users' activities, workflows, and context of use. For example, given that patients and medical staff access data and services from different locations (*e.g.*, within the hospital, or through a trusted location), hence, several works have proposed location-based authentication approaches in healthcare environments [56, 72].

In the past years, a variety of directives and regulations have been proposed that require the deployment of two-factor or multi-factor authentication solutions in healthcare environments aiming to add multiple layers of security in the user authentication process (*e.g.*, US Health Insurance Portability Accountability Act (HIPAA), European Union Agency for Network and Information Security (ENISA)). Multi-factor authentication solutions typically combine a knowledge-based authentication method (*e.g.*, a textual password), along with a token-based authentication method (*e.g.*, smart card, one-time password sent to the users' smartphone, etc.) [66]. Token-based authentication utilizing smart cards is one of the most common methods for multi-factor authentication in healthcare systems with numerous proposals in the literature [44, 48, 56, 60, 66, 71]. Works have also proposed three-factor authentication combining textual passwords, smart cards and biometric technology for increased security and usability [33, 35, 57], and the work in [5] proposed a remote patient mutual authentication scheme using smart cards and elliptic curve cryptography. Furthermore, with the advent of smartphone technology in recent years, token-based authentication is achieved with the usage of the user's smartphone device that acts as a trusted token for multi-factor authentication [1, 45, 56, 92].

Finally, biometric-based authentication is constantly gaining market share aiming to provide increased usability for accessing medical records without compromising the patients' privacy and security [79]. Biometric technologies are typically based on information about the users' physical characteristics (*e.g.*, fingerprint, iris, face, voice, etc.) and/or behavioral characteristics (*e.g.*, typing patterns, interaction and engagement patterns, etc.) [54]. Numerous biometric-based authentication schemes have been proposed that retrieve the users' biometric characteristics either: *i)* through their interaction device (smartphone, laptop, etc.) [46, 111]; or *ii)* through surroundings within smart environments, such as, smart healthcare, automated monitoring, smart manufacturing, etc. [47, 65]. Various biometric-based approaches have been proposed in the literature for granting access to medical records by utilizing voice acoustics' analysis and audio-visual identity verification [94], physiological signals analysis (*e.g.*, photoplethysmogram signals) [23, 112], face and voice analysis [75], hand geometry analysis [78], hand gesture spatial interaction analysis based on fingertips and joints [53], and periocular-based analysis [74].

2.2 Triangulating Results of Current State-of-the-Art with Healthcare Organizations

This section presents a user survey aiming to assess current user authentication practices at three European healthcare organizations in order to validate results and manifest the current literature on user authentication in the healthcare domain. Based on the analysis of the literature review, we formed the main research topics, which were further verified based on a mixed evaluation method that embraced semi-structured interviews with relevant stakeholders (security officers, department managers, doctors) of the healthcare organizations.

2.2.1 Participating Healthcare Organizations and Stakeholders

Stakeholders from three public European healthcare organizations, which support thousands of patients and users annually, have participated in the survey: *i)* Zuyderland Medical Center[1], Sittard, the Netherlands; *ii)* Hospital Clinic Barcelona[2], Barcelona, Spain; and *iii)* Western General Hospital within NHS Lothian[3], Edinburgh, Scotland. A total of nine (9) individuals participated in the user survey with varying roles in the aforementioned organization, *i.e.*, Chief Information Security Officers, Enterprise Architects, IT Department Managers, Security Experts, Doctors, Project Managers. Each stakeholder participated in a semi-structured interview that lasted for approximately forty-five (45) minutes each. Participation in the interviews was voluntary and could be cancelled at any time.

2.2.2 Procedure

A series of semi-structured interviews was conducted with key stakeholders from the participating healthcare organizations. The interviews were split in two parts. In Part A, participants were initially guided to an online consent form and each one read and agreed to participate. Participants were then introduced to the survey, its purpose and objectives. In Part B, we conducted an initial profiling (approximately five (5) minutes) of the participants asking questions that relate to the participant's background and position in the organization, with the aim to understand the background of the interviewee and the context of her/his answers. Then we discussed around two main topics: *Topic 1 - User Authentication Policy* (approximately twenty (20) minutes), which was focused on eliciting details about the user authentication policy and procedures of the organization (*e.g.*, how the policy was derived, since when the policy is valid); and *Topic 2 - Technical Details and Workflows* (approximately twenty (20) minutes), which was focused on eliciting details with regards to technical and security matters of the currently applied user authentication scheme and policy (*e.g.*, what is the current password complexity of the applied authentication policy, which is the maximum number of days a password may be used).

2.2.3 Highlights of Participants' Responses

Responses of the interviewees and the organizations were anonymized. All organizations reported that their user authentication policy is based on current industry standards and best practices, and that they primarily apply textual passwords as their core means for authentication. The main password policy is based on a widely applied policy, *i.e.*, a textual password with a minimum length of eight (8) characters containing no part of the user's real name or username, and including a minimum of one uppercase, one lowercase, one symbol and one numeric character. The policy had variations across organizations in terms of character type and their combination.

Furthermore, one out of three organizations employed multi-factor authentication in certain scenarios: *i)* when accessing the patients' database, medical staff uses an RFID badge, combined with a four-digit PIN code; and *ii)* when access the healthcare system from outside the organization's network, medical staff and patients are required to login with their textual password, combined with a second factor for authentication based on a one-time password and/or a push notification that is sent on a third-party

[1] *Zuyderland Medical Center, https://www.zuyderland.nl/english*
[2] *Hospital Clinic Barcelona, https://www.clinicbarcelona.org/en*
[3] *Western General Hospital, https://www.nhslothian.scot*

mobile application. In this respect, one participant stated that: *"Active directory is your first line of defense so that's why when you are internally it's ok and you login with your badge, so your badge is your second factor. If you are outside of the hospital that's also possible, you can get a remote reader at home, so you get an SMS or via the Microsoft app, you can authenticate and get your session" ~ Security Expert.* The same organization applies variations in the policy depending on the role of the user as well as the context of use. For example, exceptions for policies can be requested by end-users. Also, doctors may use their RFID badge to enter the emergency room. In this respect, one participant stated that: *"People can request exceptions on a policy and then we look at the case and decide whether we can change the policy" ~ Security Expert.*

We further asked participants whether their organization considers investing and deploying biometric technology, the majority reported that they have considered this technology, however, this has not been implemented yet due to increased costs and known security and privacy issues within biometric technologies. For instance, facial recognition was given a trial by one organization, however, there were problems in some cases. One participant stated that: *"The problem is that if you go a little away from the screen, or two persons are standing, one person is close and one is standing behind the screen, the system did not know which one is the user" ~ Security Expert.*

Moreover, interviews with end-users (*e.g.*, administrators, doctors) reveal that a high number of users expressed complaints on the authentication policy, *e.g.*, *"There are complaints about the complexity of the passwords, the amount of passwords they have to use, changing the passwords, so it's not a very nice picture" ~ Administrator.* Another participant reported that the users easily forget their passwords, either due to holidays or due to the frequent password changing, so there is often the need to reset them via the helpdesk, *e.g.*, *"They have problems to remember and sometimes they have to put the password in a post-it and the password is not hidden from the public when they are working in their desk" ~ Security Expert*. Several users of the organization also stated that they must remember and use more than one password, a factor that renders the authentication process harder to complete. In addition, the Web browser of these systems does not allow saving the password for the organization, *e.g.*, *"It feels like quite a large number. I would say at least 10 [passwords]" ~ Doctor.* Finally, users in general feel like they are putting a lot of effort to remember passwords and need to login several times per day. Some participants stated that they are more than willing to change their current authentication scheme, as long as it applies across multiple systems and it is not too complicated to be used, *e.g.*, *"I certainly will be willing to change as long as it is applied across multiple systems. But if it's a new authentication type that's different for each system then that would cause problems" ~ Doctor.*

Finally, we asked participants to provide details about the *"perfect authentication scheme"* and a wish list for *"better passwords"*. The majority responded that they would like to have a secure system that respects their privacy and usability. One participant responded that *"I would really like to leave our employees free and choosing what mechanism they want, the only concern is the level of security and its usability" ~ Manager.* Interest was expressed on the deployment of two-factor authentication methods utilizing the users' smartphones. Another participant was very interested in the integration of alternative and usable authentication schemes, however, concerns relate to the increased complexity and cost of applying new policies and systems in the organization's production line, *e.g.*, *"There are many procedures in order to make*

small changes. It is very difficult to implement. We are now testing another user authentication but this takes a lot of time and it will take as much time to implement it" ~ Security Expert.

3 RESEARCH MOTIVATION AND METHOD

3.1 Research Motivation

Based on the aforementioned analysis, we conclude that healthcare organizations still rely on traditional knowledge-based authentication approaches, and specifically, on textual passwords and/or location-aware approaches (*e.g.*, RFID, VPN). This is based on several reasons, *i.e.*, due to increased implementation and maintenance costs, due to immaturity of new authentication approaches, as well as known security and privacy issues of new user authentication paradigms (*e.g.*, biometrics), which negatively affect wide adoption of such technologies. Simultaneously, healthcare organizations' experts are aware that textual passwords negatively affect usability and security aspects due to complex policies, and therefore seek for novel and easy-to-adapt knowledge-based user authentication approaches as alternative solutions in order to avoid affecting the users' familiarity and existing practice.

Furthermore, the analysis revealed that: *a)* a plethora of user authentication methods (knowledge-, token-, biometric-based) has been introduced for healthcare environments, each one having its own strengths and weaknesses with regards to security, privacy and user experience; *b)* it is estimated that knowledge-based authentication mechanisms will continue to prevail in the next decades [70], even in combination with other approaches (*e.g.*, token-based) or as fallback mechanisms, hence, new approaches need to partially rely on existing textual password approaches in order to support the technology transition of users; *c)* user authentication in healthcare environments entails a mixture of unique constraints and challenges related to the location and context in which interaction takes place [38]; and *d)* evidence has shown that user's preference and task performance varies depending on the user (*e.g.*, age, abilities) and the context of use (*e.g.*, interaction device, screen size), suggesting that any specific solution might not please everyone [73].

Bearing in mind that user authentication in healthcare environments is performed by users with varying profiles, in different contexts of use and on multiple heterogeneous devices, this paper investigates whether end-users would benefit from a flexible and personalized user authentication solution that would adapt and personalize different authentication mechanisms (graphical and textual) depending on their context of interaction, aiming to achieve a viable balance between security and usability [11, 13, 27, 29, 30, 59]. Our work is primarily driven by our vision to combine graphical and textual password mechanisms based on a new *"Single-Secret Two Reflections" (SS2R)* user authentication paradigm, which allows us to move from current generic "one-size-fits-all" authentication systems towards flexible, user-adaptable and personalized authentication systems [12, 28]. The aim is to provide a viable and flexible authentication solution, by following state-of-the-art practices in the healthcare domain, and applicable within current healthcare organizations.

3.2 Research Method

The research work adopted a UCD methodology throughout the entire research, design, and development process. Multiple design iterations and a significant amount of evaluation have been incorporated into the research work, with the active participation of end-users with the aim of improving the framework design. The key idea of applying a UCD approach was to partially move our focus away from the technical issues of security towards understanding the users and developing new approaches for offering personalized solutions within the healthcare domain. The research adopted a three-phase methodological approach as follows:

***Phase A*:** The first phase involved the literature review on state-of-the-art user authentication research in the healthcare domain. To verify and triangulate the literature, we conducted semi-structured interviews with diverse key stakeholders (n=9) of three European healthcare organizations, including Chief Information Security Officers, Enterprise Architects, IT Department Managers, Security Experts, Doctors, and Project Managers. This phase lasted 6 months.

***Phase B*:** The second phase involved the design and development of DuoPass authentication system, which is based on the "Single-Secret Two Reflections" paradigm by following a UCD approach. With regards to design factors, we considered security factors, usability and user experience factors, adaptation and personalization, as well as security and usability key performance indicators. As part of this phase, we also set the key performance indicators that would be adopted for the evaluation study of the DuoPass authentication system, which included password guessability, password creation efficiency, memory time, login time, and users' perceived security, usability, trust, and likeability. This phase lasted 12 months.

***Phase C*:** The third and final phase involved the user evaluation with participants of three European healthcare organizations, during which we recorded users' interactions with the suggested DuoPass approach vs. a state-of-the-art authentication approach, aiming to evaluate its security, memorability and user experience. We conducted a patient-centric feasibility study during which users interacted with the proposed authentication system (n=68), and a human guessing attack study (n=92) focusing on vulnerabilities among people sharing common experiences within location-aware images used for graphical passwords. This phase lasted 11 months.

Table A1 in the Appendix depicts our research methodology.

4 A FLEXIBLE AND PERSONALIZED LOCIMETRIC USER AUTHENTICATION PARADIGM IN HEALTHCARE

In this section we propose a user authentication method, coined *DuoPass*, which is based on a novel *"Single-Secret Two Reflections" (SS2R)* authentication paradigm. We first provide details on the underlying theory and conceptual design of the approach. We further present the prototype designs and describe how we addressed security and usability aspects during the design of DuoPass.

4.1 Conceptual Design based on the Dual Coding Theory

User Scenario: From Location-based Memories towards Location-aware Passwords. Consider a scenario in which a patient, Emma, visits her hospital for her weekly checkup at her doctor. Emma drives

with her car through the entrance of the hospital and then parks her car. She further walks from the car park through the hospital's garden, enters the building and goes to the reception hall. She then registers at the reception hall in which she confirms her appointment with her doctor. She is then asked to wait for fifteen minutes until her appointment. During these fifteen minutes, Emma walks to the hospital's cafeteria and orders a coffee and croissant until her appointment. Emma completes the checkup with her doctor, receives a prescription of medication and then leaves the hospital and drives back home.

During Emma's visit at the hospital, she created several real-life memories within the hospital (*e.g.*, walk through the garden, visit at the cafeteria, appointment with the doctor). Based on the dual coding theory [80, 96], Emma encrypted a series of visual and verbal stimuli within her long-term memory [6, 9], and more specifically with the episodic, semantic and autobiographical memories [95, 102, 106], which entail information about certain events experienced in an individual's life-time and the corresponding semantic information describing these events. Furthermore, according to the dual coding theory, the human brain consists of a *visual cognitive sub-system*, which is utilized by the human brain during processing, representation and recall of imagery information, as well as a *verbal cognitive sub-system*, which is utilized by the human brain during processing, representation and recall of verbal information [80]. For example, information such as the word *"cappuccino"* is represented in the human mind as a visual representation of a cappuccino coffee cup, as well as the word *"cappuccino"*. During recall, individuals retrieve and process both representations simultaneously or separately.

4.2 DuoPass Authentication Paradigm

DuoPass aims to leverage on the dual coding theory based on a novel *"Single-Secret Two Reflections"* authentication paradigm, by enabling patients to create a single conceptual secret leveraging upon their personal *location-based memories* they have built through their interactions in certain locations within the hospitals, and further reflect the secret on a graphical and/or textual password key. For creating the graphical password key, DuoPass presents *location-aware images* that depict image content of a certain location of a hospital, in which the patient had prior interaction with. In addition, DuoPass provides an additional option to the patient to create a textual password key that may be then utilized interchangeably with the graphical password based on user's preference. Our Web-based solution intentionally includes a textual password as an option to avoid changing the current state-of-the-art practice in the healthcare domain, and a method in which users are familiar with. Hence, we anticipate that DuoPass will be more easily transferable from the current state-of-the-art towards the new suggested approach, providing the option to users to switch to their preferred authentication type (graphical or textual).

Graphical Passwords. The graphical password mechanism is based on cued-recall graphical authentication mechanisms [13], which ask users to draw secret gestures on a background image that acts as a cue. For its implementation, we follow design and development guidelines of Microsoft's PGA mechanism [58], deployed in Windows 8™ and 10™, which allows users to draw three types of gestures on the background image: taps (clicks), lines and circles. Free line gestures are automatically converted into one of the three allowed gestures. To process the gestures, the mechanism creates a grid of the image containing 100 squares (segments) on the longest side, and then divides the shortest side by the same scale. Rounding is not applied to any decimal segments and the mechanism allows .25 segments size overflow at the rightmost side of the image. The approach of creating a grid of squares allows for storing the gestures

based on their segment position on the grid rather than the coordinates in pixels. The following data is stored: for taps, the (x, y) coordinates of a point, for lines the (x, y) coordinates of the starting and ending point, and for circles the (x, y) coordinates of the center, the radius and the directionality (clockwise/counter-clockwise). The credentials are represented as a 7-tuple alphanumeric string (*e.g.*, <g, x_1, y_1, x_2, y_2, r, d>), which consists of the gesture's type, location, and other attributes (*e.g.*, radius and directionality in case of circles) [114], hashed using a hash function (*e.g.*, sha256), and securely stored similar to text-based passwords.

Textual Passwords. DuoPass follows state-of-the-art security metrics and authentication policies with regards to the implementation of textual passwords [19, 62]. The textual password keys rely on a basic 16-character password policy, allowing the creation of dictionary words with no composition requirements, which is more usable and as secure as traditional complex 8-character policies [62] (NIST predicts that both policies generate 30 bits of security entropy [19]).

In this context, DuoPass allows users to create a secret graphical and/or a textual password. During graphical password composition, DuoPass deploys images depicting popular sceneries of the hospital (*e.g.*, garden, reception hall, cafeteria, etc.). The user is asked to select an image of her preference and then create a graphical password by drawing secret gestures on certain regions of the image based on the experience she had with the depicted content in the image. For example, based on the aforementioned user scenario, a conceptual secret derived from Emma's episodic memory and experiences at the hospital would be: *"the cappuccino I drank at the hospital"*. Emma would reflect this secret on the graphical password by selecting for example a coffee cup and the exact table she sat for having her coffee in the hospital's cafeteria. As a next step, DuoPass also allows users to create a textual password by asking the patient to reflect the conceptual-based graphical secret as a textual representation by articulating the secret, *e.g.*, the textual version of the secret would be *"CappuccinoIDrankAtTheHospitalsCafeteria"*.

Hence, the "Single-Secret Two Reflections" paradigm extends existing works in knowledge-based user authentication based on the dual coding theory aiming to: *a)* enhance security by enabling users to select regions on an image that are familiar to the users and not to the attackers; *b)* to enhance memorability through ownership, and prior experience and knowledge of each single user; and *c)* to support user authentication adaptability since users can choose their preferred way to login based on their needs and context of use. For example, users that are on the move might prefer to login through touch-based graphical password input on the tablet device, while users that are in the office might prefer to login through a textual password input on the conventional desktop computer.

5 FEASIBILITY STUDY

The goal of the feasibility study is three-fold: *i)* compare the security strength of graphical passwords when users create a graphical password based on a location-aware image *vs.* non-location-aware image; *ii)* compare the memorability aspects when users create a graphical password based on a location-aware image *vs.* non-location-aware image; and *iii)* elicit the users' perceived security, usability, memorability, trust, and likeability towards the DuoPass paradigm.

5.1 Research Questions

We investigated the following research questions. The aim of *Research Questions RQ_1, RQ_2 and RQ_3* is to compare the suggested personalized and location-aware graphical password scheme of DuoPass with the state-of-the-art approaches in graphical password authentication. In addition, after analyzing quantitatively the observed effects, we investigate in RQ_4 the perceived security, usability, memorability, trust, and likeability towards the DuoPass approach, and in RQ_5 we investigate which of the authentication types of DuoPass (graphical *vs.* textual) the users prefer for authentication.

RQ_1: Is there a significant improvement in security strength of the selected graphical passwords between the DuoPass condition *(experimental group)* and the state-of-the-art condition *(control group)*?

RQ_2: Is there a significant difference in graphical password entry efficiency between the DuoPass condition *(experimental group)* and the state-of-the-art condition *(control group)*?

RQ_3: Is there a significant improvement in memorability between the DuoPass condition *(experimental group)* and the state-of-the-art condition *(control group)*?

RQ_4: Do end-users score positively with regards to perceived security, usability, memorability, trust, and likeability towards the DuoPass paradigm?

RQ_5: Which authentication type (graphical *vs.* textual) do users prefer for authentication?

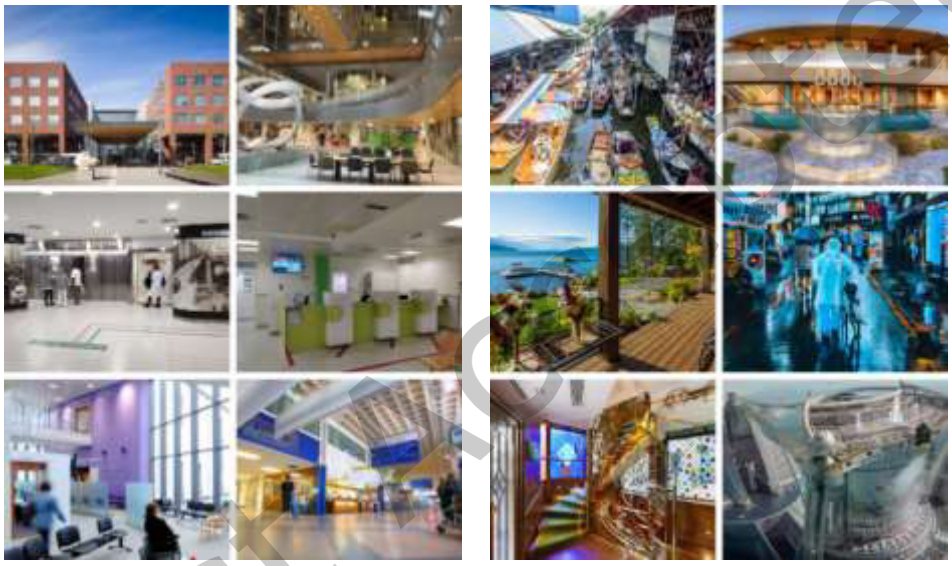
5.2 Image Sets – Location-aware and Non-location-aware Image Semantics

We created two image sets in order to control the image semantics and consequently investigate the research questions as follows: *i) location-aware image set (experimental group):* this image set included images that depicted content relevant to the participants' hospital (*e.g.*, hospital cafeteria, reception hall, front yard, etc.), which was related to their location-based experiences and memories created during their visits at the hospital; and *ii) non-location-aware image set (control group):* this image set included images that depicted generic content that was not relevant to the users (*e.g.*, sceneries from landscapes, people, etc.) in order to control the participants' familiarity with the image content. Both image sets followed existing research, which revealed that end-users typically choose images depicting sceneries [4, 18, 59, 83, 86, 113, 114].

Furthermore, bearing in mind that the complexity of an image and the number of Points-of-Interest (PoI – regions of an image that attract the users' attention) affect the security strength of user-created graphical passwords [28, 59], we carefully selected images that had similar content complexity and number of PoIs for both user groups (experimental and control). This was achieved by applying saliency maps and saliency filters [28, 81] to detect salient regions on the images, entropy estimators [20] to calculate the image complexity, and computer vision techniques to detect PoIs. **Figure 1** illustrates a subset of the images used in the study. **Table 1** illustrates the means of image complexity and mean number of PoIs for each image set.

5.3 Procedure and Participants

For investigating the research questions, a between-subjects study design was conducted in which we formed two groups of users, *i.e.*, the *experimental group* that used an authentication system, including location-aware images based on the suggested DuoPass paradigm, and the *control group* that used an authentication system, including non-location aware images based on current state-of-the-art authentication approaches in graphical user authentication. Specifically, the experimental group included patients from three different hospitals, which received location-aware images (*i.e.*, image content depicting sceneries from their hospitals) (**Figure 1 left**), while the control group included end-users in a non-healthcare context, which received non-location-aware images that depicted generic content that was not familiar to them (**Figure 1 right**). To avoid bias, we provided a set of six images for each group during password composition, and participants chose one image to create their secret and eventually graphical password.



Location-aware images **Non-location-aware/generic images**

Figure 1. Location-aware images depicting sceneries from the mainstream areas at the patients' hospitals (left), and non-location-aware images depicting generic and abstract content that was not familiar to the patients (right).

A total of 68 individuals (36 in the experimental group; 32 in the control group), ranging in age from 20 to 60, were recruited and split in the two groups. To assure that users in both groups were motivated to use secure passwords, we applied the user authentication task in the frame of an online service. Users were asked to perform specific tasks (*e.g.*, access a specific service and view information) that first required them to login. This way we did not explicitly ask the participants to login to keep the authentication task as a secondary task of interaction, and hence increase ecological validity. All the individuals participated voluntarily and provided their consent that their interactions would be recorded anonymously in the context of an experimental research study. Also, the participants could opt out of the study any time they liked.

Table 1: Means of image complexity and number of PoIs for each image set.

	Control		Experimental	
	Mean	**St. Dev.**	**Mean**	**St. Dev.**
Complexity in bits	7.53	.15	7.47	.14
Number of PoI Regions	6.77	.62	7.11	.73

The experiment was split in two phases. In *Phase A* (Day 0), participants were introduced to the assigned authentication system, completed a questionnaire on demographics and then created and confirmed their password key. Users then completed a short task within the service, requiring them to first login. *Phase B* was performed on Day 1, Day 3 and Day 6 after Phase A. In all sessions, we asked participants to complete a task in the online service, which was only accessible through login, during which they had to recall their password key and access the service through the assigned authentication system.

5.4 Data Metrics

Graphical password strength: we measured graphical password strength based on an accredited password guessability metric [113, 114], which is calculated based on the number of guesses required to crack the users' passwords. Based on existing approaches that have applied this metric [28, 59], we similarly implemented and applied a brute-force attack model that considers PoIs (*i.e.*, regions on an image that attract the users' attention), starting from segments covering the PoI segments, then checking the neighboring segments, and finally checking the rest of the segments.

Password composition time: password composition time is calculated as the time required to create the graphical password, starting from the time the image is illustrated until the end-user successfully completes the password composition task.

Memorability: for measuring memorability we used memory time [97], which is the greatest length of time between a password creation and a successful password login using the same password.

Users' perceived security, memorability, trust and likeability: at the end of the experiment we asked participants from the experimental group on aspects that relate to perceived security, memorability, trust and likeability of the proposed paradigm. We also measured usability aspects by utilizing the System Usability Scale (SUS) [17], which is a widely applied instrument for measuring password usability.

5.5 Analysis of Results

5.5.1 Security strength between the control and experimental group (RQ_1)

To investigate $\boldsymbol{RQ_1}$, we ran two security analyses to investigate whether there are differences in security strength of the user-created graphical passwords between the control and experimental user groups. The first analysis compared password guessability that was based on a naïve brute-force attack, while the second analysis compared password guessability that was based on the PoI-assisted brute-force attack. **Figure 2 left** illustrates the means of password guessability among user groups, as assessed by the naïve and the PoI-assisted brute-force attack model. For the naïve brute-force attack, we ran an independent

samples t-test to determine whether the two user groups (control *vs.* experimental) generated different password strengths in terms of password guessability. The assumption of homogeneity of variances was not violated, as assessed by Levene's test for equality of variances (p=.075). There were no significant outliers in the data, as assessed by inspection of boxplots, and data were normally distributed, as assessed by Shapiro-Wilk's test (p>.05). Results revealed significant differences with a mean difference of 22 million guesses (95% CI, -5.7 million to 1.28 million), t(66)=-1.261, p=.021. In particular, user-chosen graphical passwords of the experimental group required 53 million guesses to crack, while for the control group 31 million guesses.

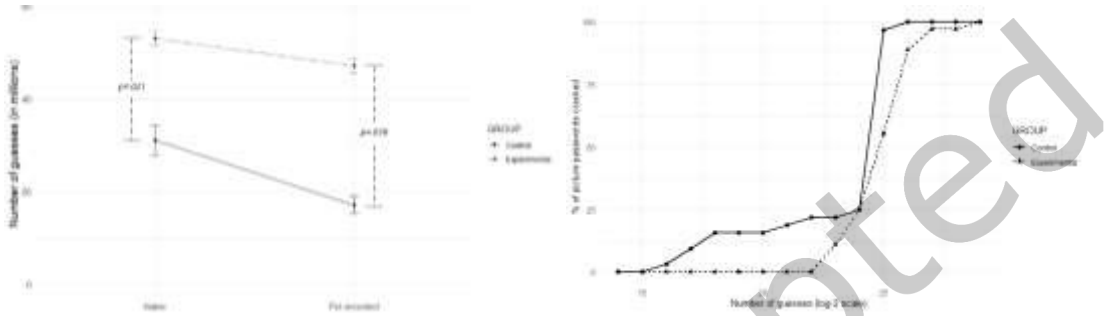


Figure 2. Means of password guessability among user groups (control *vs.* experimental), as assessed by the naïve brute-force attack model (left bars) and the PoI-assisted (right bars) brute-force attack model (left); percentage of graphical passwords cracked, as assessed by the PoI-assisted brute-force attack model (right).

For the PoI-assisted brute-force attack, we ran a Welch t-test to determine whether the two user groups (control *vs.* experimental) generated different password strengths in terms of password guessability, due to the assumption of homogeneity of variances being violated, as assessed by Levene's test for equality of variances (p=.048). There were no significant outliers in the data, as assessed by inspection of boxplots, and data were normally distributed, as assessed by Shapiro-Wilk's test (p>.05). Results revealed significant differences with a mean difference of 30 million guesses (95% CI, -6 million to -346 thousand), t(36.165)=-2.140, p=.039. In particular, user-chosen graphical passwords of the experimental group required 47 million guesses to crack, while those of the control group required 17 million guesses to crack. **Figure 2 right** illustrates the percentage of passwords cracked indicating that users from the control group exhibited higher percentage of passwords cracked than users from the experimental group. The percentage of graphical passwords cracked reached 100% for the control group within 2^{26} guesses, and for the experimental group within 2^{29} guesses.

To further verify the security strength of the created passwords, we took an extra step to analyze the users' individual gestures with respect to PoI regions (*i.e.*, regions that attract the users' attention and are prone to automated guessing attacks) [28]. To identify the PoIs of each image, we followed a semi-automated image analysis approach by applying saliency maps and saliency filters [81] to detect salient regions on the images and computer vision techniques to detect PoIs [28]. **Figure 3 left** illustrates an example of a hospital image used and its corresponding salient regions as detected through the image analysis (**Figure 3 right**).

A Mann-Whitney U test was run to determine if there were differences in number of PoI selections between the control and experimental group. Distributions of values for the two groups were similar, as assessed by visual inspection. Median number of PoI selections for the control group (1.72) was statistically significantly higher compared to the experimental group (1.53), U=224.000, z=-4.561, p<.001, using an exact sampling distribution for U [36]. We further ran an independent-samples t-test, with the user group (control *vs.* experimental) as the independent variable, and the proportion of gestures falling into PoI regions as the dependent variable. The analysis (**Figure 4**) revealed that users of the experimental group made a lower proportion of selections falling into PoI regions (0.45 ± 0.04) than users of the control group (0.71 ± 0.04), a statistically significant difference of 0.26 ± 0.05 (95% CI, .15 to .37), t(65)=4.93, p<.001. Also, the effect size (Hedge's g = 1.112 [50]) indicates a large effect since it is greater than 0.8 [24].

Figure 3. Regular image depicting the entrance of a hospital (left); and its corresponding salient regions as detected through the image analysis (right).

To further investigate whether individuals, who share common experiences within the sceneries depicted in the location-aware images, tend to create similar passwords when they use the same image during password creation, we first split the participants from the experimental group into subgroups based on the image they used. In the sample of the experimental group (n=36), one out of nine images was not used by any participant. From the remaining eight images that were selected by participants, two images were selected by only one participant, and six images were selected by more than one participant, thus, forming six subgroups of participants.

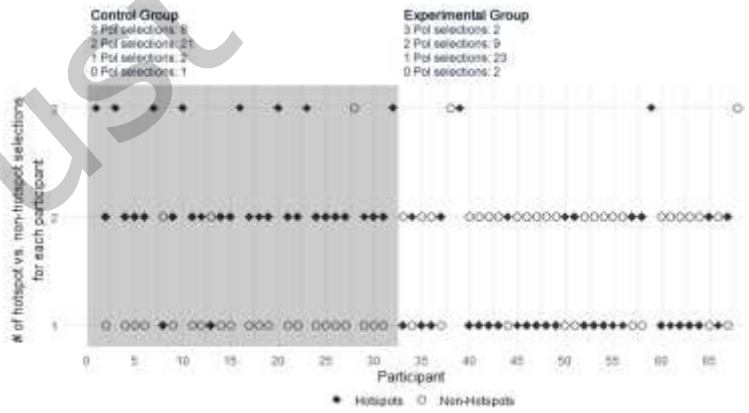


Figure 4. Proportion of user selections falling into PoI regions for each user group indicating that users of the experimental group made a lower proportion of selections falling into PoI regions compared to the control group.

Given that the implementation of DuoPass graphical password mechanism takes into consideration the order and the type of gestures (*e.g.*, circles are more complex than simple taps but less complex than lines[4]), in order to understand the similarities of users' password selections, we have disregarded the order and the type of the gestures and rather focused on the positions of the password selections. To do so, we simplified the gesture type as follows: For circles, we disregarded the radius and the directionality and kept only the center of the circle as a *x*, *y* segment, while for lines, we considered only the *x*, *y* segment of the starting point of the line. **Table 2** summarizes the similarities in image regions across users who created their password on the same image. Accordingly, out of 102 gestures made by 34 users (n=36, but we exclude the two images that were selected by only one participant), six users chose one same region, two users chose two same regions, and no user selected all three same regions. We would also like to note that when we consider the exact order of the gestures, there are no observed similarities in image regions across all subgroups and all participants.

Table 2: Summary of the similarities in image regions across users who created their password on the same image.

Image Subgroup	# of users that selected the same image	Common regions in password selections		
		1 out of 3	2 out of 3	3 out of 3
1	7	2	1	-
2	5	-	-	-
3	7	1	-	-
4	4	1	-	-
5	6	2	1	-
6	5	-	-	-
Total	**34**	**6**	**2**	**0**

5.5.2 Graphical password composition efficiency between the control and experimental group (RQ_2)

To investigate ***RQ_2***, we ran a two-way mixed analysis of variance (ANOVA) with the user group (control *vs.* experimental) and users' password selections (three consecutive selections) as the independent variables, and the time to make each password selection as the dependent variable. There were no significant outliers, as assessed by inspection of boxplots. The data were normally distributed, as assessed by Shapiro-Wilk's test of normality (p>.05). There was homogeneity of variances (p>.05), as assessed by Levene's test of homogeneity of variances. The analysis revealed significant differences between the three users' password selections on the time to compose the graphical password, $F(1, 66)=86.942$, $p<.01$, *partial* $\eta^2=.568$. The analysis revealed that there was no interaction between the user group and users' password selections on the time to compose the graphical password, $F(1, 66)=1.459$, $p=.231$, *partial* $\eta^2=.022$. **Figure 5 left** depicts the time to make each of the three graphical password selections.

We further examined simple main effects for each password selection. Data are mean ± standard error, unless otherwise stated. The analysis revealed that the time to create the last (third) selection between the two groups was statistically significant (*control:* 1109.81 ± 1222.91 msec *vs. experimental:* 650.83 ± 509.97 msec) with a mean difference of 458.97 seconds, $F(1, 66)=4.161$, $p=.043$, *partial* $\eta^2=.06$. With regards to the first and second selections, there were no significant differences between the control and experimental groups (p>.05).

5.5.3 Memorability differences between the control and experimental group (RQ_3)

To investigate $\boldsymbol{RQ_3}$, we measured login task completion time for each of the login sessions over the seven-day period, and memory time, which is the maximum amount of time (in hours) someone could effectively remember their password from the day of creation. Accordingly, we initially analyzed login task completion time using a mixed effects analysis (with the *lme4* package in R) [10] since this enabled us to handle all the variables of the study while accounting for repeated-measures of individuals (4 login sessions over a period of seven days) and for handling missing data of users, *e.g.*, a user that has not participated in some sessions across the seven days of the study can be used in the analysis without requiring removing the user from the sample [82]. In this respect, we performed a mixed effects analysis of the relationship between the time to successfully authenticate (by also including any failed attempts that eventually ended in a successful authentication) and the user group. As fixed effects, we entered the user group (control and experimental) into the model. As random effects, we used subjects in order to account for non-independence of measures. Visual inspection of residual plots revealed that linearity and homoscedasticity were not violated. *P*-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question [107]. The analysis revealed that the user group had no impact on the time needed to authenticate ($x^2(1)=.171$, $p=.679$). The mean login time for the users of the control group was 7.47 ± 4.78 seconds, while for the users of the experimental group was 7.63 ± 6.3 seconds. **Figure 5 right** depicts the mean login time across user group for each of the four sessions. We have further analyzed the users' login attempts, indicating that overall, the vast majority of login attempts was completed using the graphical password. Regarding the textual password login attempts, 5 out of 36 individuals from the experimental group also logged in once using textual passphrase with mean login time 5.21 ± 2.43 seconds, while 10 out of 32 individuals from the control group also logged in once using textual passphrase with mean login time 8.66 ± 3.05 seconds.

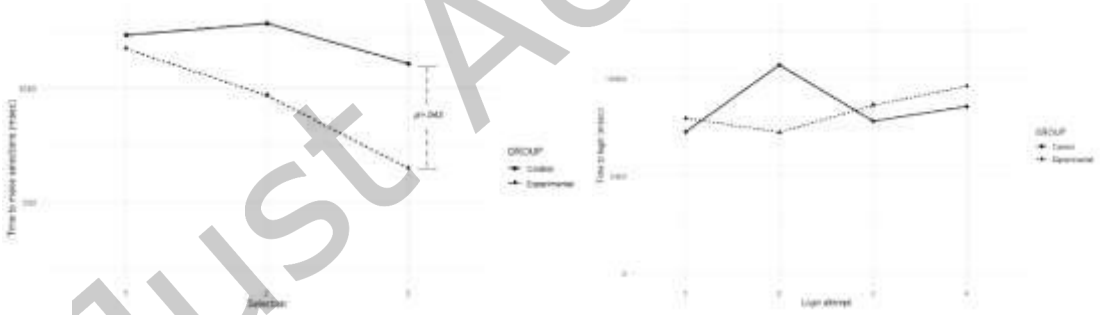


Figure 5. Time to make password selections (left); time to login across the seven-day period for each group (right).

The maximum memory time that someone could achieve was approximately 168 hours (7 days x 24 hours). To investigate memorability, we conducted an independent-samples t-test, with the user group (control *vs.* experimental) as the independent variable, and the memory time as the dependent variable. The analysis revealed that memory time between the two user groups was not statistically significant different, $t(66)=-.961$, $p=.340$. Memory time of the control group was 106.13 ± 76.8 hours, while memory time of the experimental group was 121.58 ± 55.2 hours.

5.5.4 Users' Perceptions towards security and experience of the DuoPass approach (RQ4)

To investigate ***RQ4***, we conducted a post-study survey to elicit the users' perceptions (experimental group) on the security, memorability, trust, usability and likeability towards the DuoPass system based on their interactions. For this purpose, we have designed a questionnaire by following state-of-the-art works and guidelines on eliciting perceived security, trust, memorability, usability and user experience [8, 17, 21]. Example statements of the survey were: *"Overall, how secure do you find the DuoPass password system?"*, *"How mentally demanding was the login task?"*, *"I trust in the ability of the DuoPass password system to protect my privacy"*, etc. Users rated the statements through a 5-point Likert scale, with the labels changing depending on the question (*e.g.*, 1: Strongly disagree – 5: Strong agree; 1: Very insecure – 5: Very secure). Perceived usability was measured through the SUS [17], which is an accredited and widely applied system usability instrument and widely used in password studies [21]. The survey also investigated the likeability towards the DuoPass personalized and flexible approach of DuoPass with users rating the statements through a 5-point Likert scale (1: Not at all – 5: Absolutely). **Figure 6** illustrates the responses of participants towards perceived security, memorability, trust and likeability.

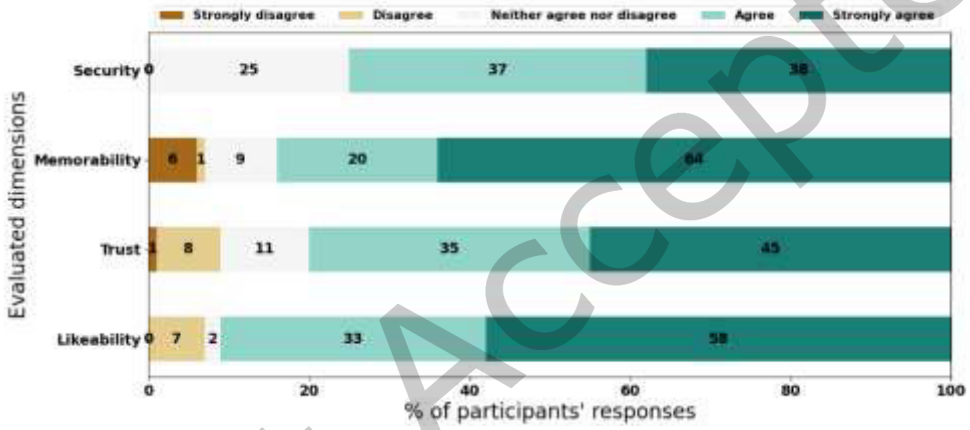


Figure 6. User responses with regards to perceived security, perceived memorability, perceived trust and likeability towards the personalization and flexible approach.

Results revealed that the majority of participants perceive the DuoPass system as secure (75%), with low mental demand (77%) in recalling the password and users could effectively recall their password (84%), while 80% of the participants trust the technology and its ability to keep their data private and secure. Furthermore, when participants were asked on whether they like the flexible and personalized approach for user authentication, the majority of participants (91%) extremely (21/36) or very much (12/36) liked the idea, with 3 users either moderately (1/36) and slightly (2/36) liking the idea. **Table 3** also summarizes the likeability scores per healthcare organization, indicating that patients across organizations had a consensus on liking the suggested approach and increase ecological validity.

Representative positive responses from participants include: *"I like it very much, it a great new approach", "easy yet hard to crack by (hackers)", "no one knows what I see in it", "easier to remember", "it's a much trickier way for criminals to find out what someone has entered", "accessible to everyone", "easy to use,*

especially through the use of images that capture the imagination of the user and therefore easier to remember and harder to find for people who want to harm. Each image is basically different and very personal", "very personal password", "I would like a lot that the images displayed were personal images that I could upload", "genius", "a sentence is better for me to remember", "point out the many possibilities with regard to places in the photo", "it was very easy to remember", "Very individual. Less prompts requirements usual password and feels very secure because of that", "pictures are good", "Easy to use and bespoke", "Works for people who learn using pictures. Hopefully easier to remember", "great not to have to remember yet another password", "it is very personalized", "not having to remember text", "It is nice and quick to be able to log with 3 clicks you remember from an almost infinite possibility of clicks and orders", "For people who have visual memory, using images instead of text it's probably easy to remember", "Better for dyslexic people", "I like the concept of using picture passwords, it can make it easier to remember".

Table 3: Likeability of DuoPass per healthcare organization.

	Healthcare Organization 1	Healthcare Organization 2	Healthcare Organization 3	Total
Extremely	10	7	4	21
Very much	5	3	4	12
Moderately	1	0	0	1
Slightly	1	0	1	2
Not at al	0	0	0	0

Negative responses from participants include: *"there are some places in the images that obviously pop out more and possibly people is more prone to use them in their password, making this system more insecure", "people will use the easiest* gestures *and this could be a safety concern", "if you don't use it often would it be still memorable?", "with a photo you have to be able to use the same photo everywhere, because otherwise I can't remember the login code", "My main concern is that choices of pass gestures and locations are not truly random and can be figured out using a user's publicly available data, such as for instance a social network profile", "Perhaps tapping on the heads is too obvious, people might create weak passwords", "security, if I pick 3 simple gestures it may be easier for a criminal to guess".*

Finally, we measured system usability based on SUS with participants scoring an overall SUS score of 74.77%. **Table 4** summarizes the SUS scores of patients across the participating healthcare organizations. Based on the literature, the average SUS score is 68% [89]. In case the score is under 68%, the system entails various usability issues that need improvement, while a score above 68%, indicates that the system entails good usability practices. Accordingly, the scores across the three healthcare organizations ranged between 72.14% for Healthcare Organization 1, 74.58% for Healthcare Organization 2 and 80% for Healthcare Organization 3, with an overall score of 74.77%. Such results are encouraging for further investigating and improving the system since the score suggests that the DuoPass system scores very well in usability, end-users like the system and they can easily complete the authentication-related tasks. Nevertheless, given that the score is below 80%, there are still aspects that require improvements, *e.g.*, during the studies, some patients had difficulties in entering their graphical password through the developed gesture input mechanism, hence, next steps entail improving the gesture input functionality. Additionally, we conducted a one-way ANOVA to determine if the SUS score was different for people belonging to different healthcare

organizations. There were no outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by Shapiro-Wilk test (p>.05); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances (p=.843). Data is presented as mean ± standard error. SUS score increased from 72.14 ± 4.23%, to 74.58 ± 5.59%, to 80 ± 5.10%, in that order, but the differences between the healthcare organizations was not statistically significant, $F(2, 36)=0.629$, p=.538.

Table 4: System Usability Scale scores across healthcare organizations.

	Healthcare Organization 1	Healthcare Organization 2	Healthcare Organization 3	Overall
SUS Score	72.14%	74.58%	80%	74.77%

5.5.5 Summary of Main Findings

The experimental evaluation study revealed interesting insights related to security, memorability and user experience with regards to the suggested DuoPass approach. **Table 5** provides a summary of the main findings. Users make arbitrary choices in knowledge-based user authentication, which decreases the security. In locimetric passwords this is a well-known and highly researched issue. The DuoPass approach aims to overcome this issue by recommending personalized images that are related to the users' prior experiences in the healthcare environment. As such, we expected to improve security and at least retain memorability and user acceptance. Both security and memorability analyses provide evidence that the DuoPass approach assisted end-users in making password choices based on their experiences, overcoming arbitrary choices, which are one of the main reasons for decreased security and memorability in locimetric approaches.

From a security perspective, we report DuoPass' superiority against the state-of-the-art approach given that the experimental user group scored significantly higher guessability compared to the control group. This can be accredited to the fact that users from the experimental group created graphical passwords on images that were related to their prior experiences within the hospital, and hence, created selections on regions based on their experiences, rather than generic regions, which may be susceptible to a brute force attack. Such a finding is in line with existing research, which revealed that depicting images related to the users' prior sociocultural experiences increases the security of user-selected graphical secrets [28]. Furthermore, task completion efficiency analyses revealed that during the last password selections (third gesture), there were significant differences in user selections, with users from the experimental group making a significantly faster selection compared to the users from the control group. This can be explained by the fact that users from the control group needed more time to reason about a secret story on an image that was rather not familiar to them, whereas in the experimental group, users created a story based on their familiarity with the image, and consequently, as they composed their password, they were faster in making their last selections.

From a memorability and user experience perspective, descriptive statistics reveal that users from the experimental group scored higher memory time compared to the control group, indicating good memorability aspects of the approach, however, this difference was not statistically significant. This finding was triangulated with end-users' qualitative feedback in which participants perceived the DuoPass secrets as highly memorable, users were able to memorize their secret for the whole period of the study,

the majority reported low mental demand (77%) in recalling their password and that they could effectively recall their password (84%).

Finally, DuoPass scores well in usability based on participant responses to the SUS (74.77%), however indicating that there is still room for improvement given that best SUS scores should be 80% and above. From feedback received during the studies, some patients had difficulties in entering their graphical password through the developed gesture input mechanism, hence, our efforts are focused on improving the interaction design of gestures to address cross-compatibility issues and heterogeneity of devices. When users were asked on likeability aspects of the approach, the significant majority of users (90%) extremely and very much like the flexible and personalized approach, and the majority would like to use DuoPass as an alternative password system (75%).

Table 5: Summary of main findings.

	Experimental Group	Control Group	Significance
RQ_1: Is there a significant improvement in security strength of the selected graphical passwords between the DuoPass condition (experimental group) and the state-of-the-art condition (control group)?			
Naïve Brute Force Attack	53 million	31 million	*Mean difference:* 22 million guesses (95% CI, -5.7 million to 1.28 million), $t(66)=-1.261$, $p=.021$
PoI-assisted Brute Force Attack	47 million	17 million	*Mean difference:* 30 million guesses (95% CI, -6 million to -346 thousand), $t(36.165)=-2.140$, $p=.039$
PoI Selections	0.45 ± 0.04	0.71 ± 0.04	*Mean difference:* 0.26 ± 0.05 (95% CI, .15 to .37), $t(65)=4.93$, $p<.001$
Representative user responses: "Very individual. Less prompts requirements usual password and feels very secure because of that", "The picture with gestures seems very robust, I think it would be very hard to hack"			
RQ_2: Is there a significant difference in graphical password entry efficiency between the DuoPass condition (experimental group) and the state-of-the-art condition (control group)?			
Password Composition (third selection)	650.83 ± 509.97 msec	1109.81 ± 1222.91 msec	*Mean difference:* 458.97 seconds, $F(1, 66)=4.161$, $p=.043$
Representative user responses: "3 clicks is faster and easier to remember than a 16 character password", "it's something I would like as an option, but I'd still need a text password. I think connecting dots in a user chosen pattern would work better rather than arbitrary shapes on a screen"			
RQ_3: Is there a significant improvement in memorability between the DuoPass condition (experimental group) and the state-of-the-art condition (control group)?			
Login time	7.63 ± 6.3 seconds	7.47 ± 4.78 seconds	$x^2(1)=.171$, $p=.679$
Memorability	121.58 ± 55.2 hours	106.13 ± 76.8 hours	$t(66)=-.961$, $p=.340$
Representative user responses: "I like the concept of using picture passwords, it can make it easier to remember", "easier to remember", "This is an easy to remember password sequence that visually minded users will likely find very appealing"			
RQ_4: Do end-users score positively with regards to perceived usability and likeability towards the DuoPass paradigm?			
Perceived Security	75% positive	-	-

	Experimental Group	Control Group	Significance
Perceived Memorability	84% positive	-	-
Perceived Trust	80% positive	-	-
Perceived Usability (SUS)	74.77% positive	-	-
Likeability	91% positive	-	-
Representative user responses: "I like it very much, it a great new approach", "I feel it could be a robust system" "easy yet hard to crack by (hackers)", "no one knows what I see in it", "it's a much trickier way for criminals to find out what someone has entered"			
RQ_5 Which authentication type (graphical vs. textual) do users prefer for authentication?			
Authentication type preference	Graphical: 30 users Textual: 6 users	-	p<.001
Representative user responses: "I like the idea of using picture passwords", "Some people would find choosing a picture password easier than remembering an only word password. It will also be more secure", "This is an easy to remember password sequence that visually minded users will likely find very appealing", "great not to have to remember yet another password"			

6 HUMAN GUESSING ATTACK STUDY

Bearing in mind that when using location-aware images in graphical passwords, the password selections are based on the end-users' existing experiences within the depicted sceneries. Hence, it is probable that the individuals who share common experiences with the end-users might be able to guess their selections. In order to shed light on this aspect, we have conducted a human attack study focusing on guessing vulnerabilities among people sharing common experiences. Each session of the study embraced pairs of participants that were closely related (*e.g.*, family members, friends, patients, medical staff, nurses, etc.) and who shared common experiences between them. In each session, both participants were first requested to create a graphical password independently, and then each participant was requested to guess the password selections of the other participant from the same pair.

6.1 Research Question

RQ. Does the suggested user-adaptable and personalized authentication paradigm, which utilizes location-aware images for graphical passwords, entail guessing vulnerabilities in terms of allowing attackers who share common experiences with the end-users to more easily identify regions of their selected secrets?

6.2 Image Set – Location-aware Image Semantics

We extended the location-aware image set from Subsection 5.2 to include images that were related to individuals' (*e.g.*, patients, medical staff, nurses, etc.) location-based experiences and memories within the hospital. **Figure 7** illustrates a subset of the images used in the human guessing attack study. We assigned each participant a specific location-aware image based on their role and their relationship with the other participant from the same pair. Furthermore, to control the image complexity and number of PoIs, we carefully selected images that had similar content complexity and number of PoIs following the approach described in Subsection 5.2. **Table 6** illustrates the means of image complexity and mean number of PoIs of the initial and the extended location-aware image sets.

Figure 7. Location-aware images depicting sceneries from the healthcare organization 1 (left) and 2 (right).

Table 6: Means of image complexity and number of PoIs for each location-aware image set.

	Initial		Extended	
	Mean	**St. Dev.**	**Mean**	**St. Dev.**
Complexity in bits	7.47	.14	7.31	.38
Number of PoI Regions	7.11	.73	7.29	.72

6.3 Data Metrics

With regards to calculating the graphical password strength, we adjusted the PoI-assisted brute-force attack model from Subsection 5.4 to start from segments covering the segments provided by each attacker, then checking the neighboring segments, then checking the PoI segments and their neighboring segments, and finally checking the rest of the segments.

6.4 Procedure and Participants

Participants were split into pairs, and they were first requested to create a graphical password independently, and then guess the password of each other from the same pair. The study was run remotely with the researcher supporting the participants. The study was split in two phases as follows:

Phase A – Password Creation. During the first phase, each pair of participants connected to a meeting via an online means of communication (*i.e.*, Microsoft Teams™) in a pre-scheduled time, and participants were asked independently to create a graphical password in order to access an online service. To avoid bias effects during the attack phase, each participant created a password on a different location-aware image that depicted places in which they share common experiences within the hospital.

Phase B – Human Guessing Attack. In this phase, we switched the image of the pairs and each participant was requested to guess the other participant's secrets as follows: *i)* by first indicating 3 areas (*x*, *y* segments on the grid) on the image for which they believe that the other participant made their selections around them; and then *ii)* by actually drawing 3 gestures for a total of 3 attempts to guess the actual password (*i.e.*, considering the ordering of gestures and type of gestures). At the end of the attack phase, participants submitted their feedback about the rationale behind their selections as attackers. This allowed us to elicit whether the attacker's rationale is related to the shared memories and experiences she possesses with the other participant from the same pair. Finally, both participants completed a questionnaire on demographics.

A total of 92 individuals, ranging in age from 28 to 62, were recruited from two healthcare organizations (44 from healthcare organization 1; 48 from healthcare organization 2). Since the purpose of this study was to understand how individuals decide on their selections when performing an attack on a password created by another individual with whom they share common experiences within places depicted on location-aware images at the hospital, we intentionally recruited pairs of participants that are close to each other (*e.g.*, family members (n=18), friends (n=18), patients (n=20), medical staff (n=18), and nurses (n=18)). To assure that participants were motivated to use secure passwords, we applied the user authentication task in the frame of an online service. Users were asked to perform specific tasks (*e.g.*, access a specific service and view information) that first required them to login. This way we did not explicitly ask the participants to login to keep the authentication task as a secondary task of interaction, and hence increase ecological validity. All the individuals participated voluntarily and provided their consent that their interactions would be recorded anonymously in the context of an experimental research study. Also, the participants could opt out of the study at any time they liked.

6.5 Analysis of Results

6.5.1 Euclidean distance of attackers' selections from the end-users' secret selections

To investigate the *RQ*, we conducted three analyses: *i)* we calculated the Euclidean distance of the attackers' guessing selections from the legitimate end-users' password secret selections; *ii)* based on the first analysis (Euclidean distance), we adjusted the brute-force attack performed in Subsection 5.5.1 in order to investigate whether users who share common experiences were able to run a more effective attack by starting to guess regions they suspected that the users selected their password; and *iii)* we performed a qualitative analysis based on the participants' feedback at the end of the human guessing attack study to better understand the approach followed by attackers on graphical passwords created on location-aware images. In the analyses that follow, data are mean ± standard error. There were no significant outliers in the data.

A. Disregarding the type of the gesture and the exact order. **Figure 8** depicts the Euclidean distance of each gesture of each participant by disregarding the type and the exact order of the attackers' gestures and the end-user's gestures. For the analysis, we adopted a threshold of 3 segments, by considering the allowed tolerance of the graphical password mechanism[4]. Accordingly, among

4 Microsoft™ Picture Password blog - bit.ly/2SajCDO

276 gestures (3 gestures x 92 participants), 49 gestures (17.7%) were in close proximity with the attacker's guessed selections. Furthermore, we conducted a one-way multivariate analysis of variance (MANOVA) to determine the effect of relationship between attackers and legitimate end-users on how far the attackers' password selections were from the legitimate end-users' password selections. Three measures were assessed: Euclidean distance of the legitimate-users and the attackers on the first gesture, second gesture, and third gesture of the graphical password. Participants belonged to one of the following categories: family member, friend, medical staff, patient, and nurse. Preliminary assumption checking revealed that data was normally distributed, as assessed by Shapiro-Wilk test (p>.05); there were no univariate or multivariate outliers, as assessed by boxplot and Mahalanobis distance (p>.001), respectively; there were linear relationships, as assessed by scatterplot; no multicollinearity (r=.117, p=.004 between gesture one and gesture two; r=.021, p=.007 between gesture one and gesture three; and r=.133, p=.010 between gesture two and gesture three); and there was homogeneity of variance-covariance matrices, as assessed by Box's M test (p=.007). The analysis revealed that the differences between groups on the combined dependent variables was statistically significant, $F(12, 225.180)=4.356$, p<.0005; *Wilks'* Λ=.576; *partial* η^2=.168. Follow-up univariate ANOVAs revealed that all three gestures were statistically significantly different between the participants from different relationship group (First gesture: $F(4, 87)=5.351$, p=.001; *partial* η^2=.197; Second gesture: $F(4, 87)=3.305$, p=.014; *partial* η^2=.132; Third gesture: $F(4, 87)=4.550$, p= .002; *partial* η^2=.173; Tukey-Kramer post-hoc tests showed that for the first gesture, participants from the family group had statistically significantly lower mean scores than participants from either the patient group (p=.002) or the nurse group (p=.050), while participants from the friend group had statistically significantly lower mean scores than participants from the patient group (p=.006). Regarding the second gesture, participants from the family group had statistically significantly lower mean scores than participants from the nurse group (p=.020). Regarding the third gesture, participants from the family group had statistically significantly lower mean scores than participants from the medical staff group (p=.036), the patient group (p=.025) and the nurse group (p=.001).

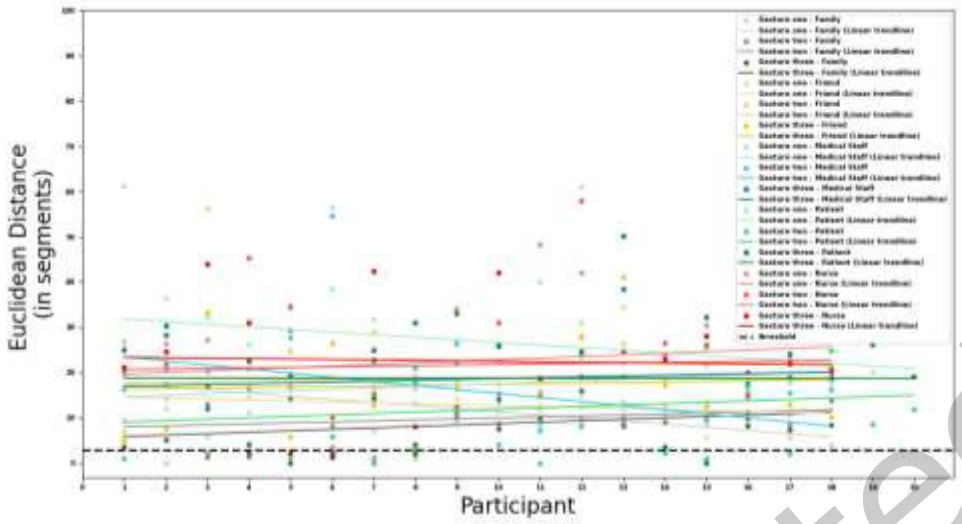


Figure 8. Euclidean distance between attackers' gestures and end-users' gestures, by disregarding the type and the exact order (*i.e.*, the attacker's first gesture was compared to the end-user's closest gesture, the attacker's second gesture was compared to the end-user's closest gesture, and the attacker's third gesture was compared to the end-user's closest gesture).

B. Disregarding the type of the gesture but considering the exact order. **Figure 9** depicts the Euclidean distance by disregarding the type of selections, but considering the exact order of the attackers' gestures and the end-users' gestures. Applying the same threshold of 3 segments, the analysis revealed that among 276 gestures (3 gestures x 92 participants) made by the participants, 19 gestures (6.8%) were in close proximity with the attacker's guessed selections. Furthermore, we conducted a one-way MANOVA to determine the effect of relationship between attackers and legitimate end-users on how far the attackers' password selections were from the legitimate end-users' password selections. Three measures were assessed: Euclidean distance of the legitimate-users and the attackers on the first gesture, second gesture, and third gesture of the graphical password. Participants belonged to one of the following categories: family member, friend, medical staff, patient, nurse. Preliminary assumption checking revealed that data was normally distributed, as assessed by Shapiro-Wilk test (p>.05); there were no univariate or multivariate outliers, as assessed by boxplot and Mahalanobis distance (p>.001), respectively; there were linear relationships, as assessed by scatterplot; no multicollinearity (r=-.089, p=.004 between gesture one and gesture two; r=.253, p=.008 between gesture one and gesture three; and r=.055, p=.009 between gesture two and gesture three); and there was homogeneity of variance-covariance matrices, as assessed by Box's M test (p<. 0005). The analysis revealed that the differences between groups on the combined dependent variables was statistically significant, $F(12, 225.180)=11.500$, p<.0005; *Wilks' Λ*=.282; *partial* η^2= .344. Follow-up univariate ANOVAs revealed that all three gestures were statistically significantly different between the participants from different relationship group (First gesture: $F(4, 87)=12.567$, p<.0005; *partial* η^2=.366; Second gesture: $F(4, 87)=3.930$, p=.006; *partial* η^2=.153; Third gesture: $F(4, 87)=13.832$, p<.0005; *partial* η^2=.389; Tukey-Kramer post-hoc tests showed that for the first gesture, participants from the family group had statistically significantly lower mean scores than participants from the medical staff group (p<.0005),

the patient group (p=.001) and the nurse group (p<.0005), while participants from the friend group had statistically significantly lower mean scores than participants from the medical staff group (p=.001), the patient group (p=.004) and the nurse group (p<.0005). Regarding the second gesture, participants from the family group had statistically significantly lower mean scores than participants from the nurse group (p=.002). Regarding the third gesture, participants from the family group had statistically significantly lower mean scores than participants from the medical staff group (p=.011), the patient group (p<.0005) and the nurse group (p<.005), while participants from the friend group had statistically significantly lower mean scores than participants from the patient group (p<.0005) and the nurse group (p<.0005).

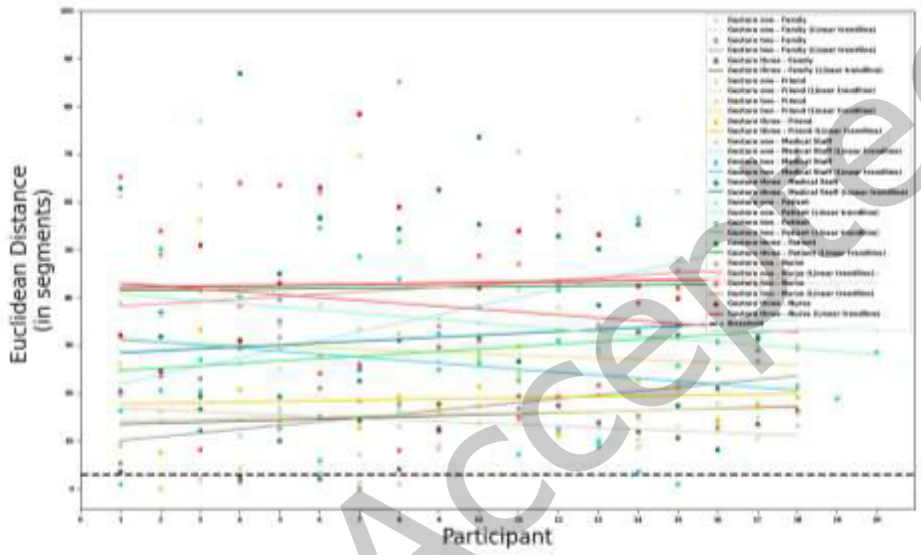


Figure 9. Euclidean distance between attackers' gestures and end-users' gestures, by disregarding the type but considering the exact order (*i.e.*, the attacker's first gesture was compared to the end-user's first gesture, the attacker's second gesture was compared to the end-user's second gesture, and the attacker's third gesture was compared to the end-user's third gesture).

C. Considering the type of the gesture and exact order. We compared the 3 attempts of each attacker with the end-user's stored password from the same pair of participants. From a total of 276 attacking guesses (3 attempts of each attacker x 92 participants), there was no successful attempt, yielding an online success guessing rate of 0%.

6.5.2 Security strength of the created graphical passwords based on experience-spot-driven brute-force attack

To investigate whether the suggested location-aware image approach holds against attacks when considering the experience-spots indicated by each participant that acted as an attacker, we conducted an offline attack comparing a PoI-assisted brute-force attack (the same attack that considers PoIs as described in Subsection 5.5.1) and a personalized PoI-assisted brute-force attack that was further enhanced to consider the experience-spots regions as indicated by the human attacker.

A. Disregarding order and type of gestures across all participants. Given that the implementation of PGA-like mechanisms takes into consideration the order and the type of gestures, which could impact the total guesses required to crack a graphical password (*e.g.*, circles are more complex than simple taps but less complex than lines[4]), it is interesting to first understand how each attack type (*PoI-assisted brute-force attack vs. Personalized PoI-assisted brute-force attack*) performs when we disregard the order and the type of the gestures and rather focus on the positions of the password selections. To do so, we simplify the gesture type as follows: For circles we disregard the radius and the directionality and keep only the center of the circle as a *x, y* segment, while for lines we consider only the *x, y* segment of the start of the line.

A one-way MANOVA was run to determine the effect of relationship between attackers and legitimate end-users on the number of guesses required to crack the passwords when using a PoI-assisted brute-force attack *vs.* a personalized PoI-assisted brute-force attack by considering also the experience-spots provided by the attackers. Two measures were assessed: Number of guesses required to crack the passwords when using a PoI-assisted brute-force attack and number of guesses when using a personalized PoI-assisted brute-force attack. Participants belonged to one of the following categories: family member, friend, medical staff, patient, nurse. Preliminary assumption checking revealed that data was normally distributed, as assessed by Shapiro-Wilk test (p>.05); there were no univariate or multivariate outliers, as assessed by boxplot and Mahalanobis distance (p>.001), respectively; there were linear relationships, as assessed by scatterplot; no multicollinearity (r=.394, p<.0005); and there was homogeneity of variance-covariance matrices, as assessed by Box's M test (p=.002). The analysis revealed that the differences between groups on the combined dependent variables was statistically significant, $F(8, 172)=4.546$, p<.0005; *Wilks'* Λ=.681; *partial* η^2=.175. Follow-up univariate ANOVAs revealed that in both types of attacks the number of guesses required to crack the passwords was statistically significantly different between the participants from different relationship group (PoI-assisted brute-force attack: $F(4, 87)=4.650$, p=.002; *partial* η^2=.176; Personalized PoI-assisted brute-force attack: $F(4, 87)=7.473$, p<.0005; *partial* η^2=.256; Tukey-Kramer post-hoc tests showed that for the PoI-assisted brute-force attack, participants from the family group had statistically significantly lower mean scores than participants from the nurse group (p=.001), while participants from the friend group had statistically significantly lower mean scores than participants from the nurse group (p=.024). Regarding the personalized PoI-assisted brute-force attack, participants from the family group had statistically significantly lower mean scores than participants from the nurse group (p<.0005), participants from the friend group had statistically significantly lower mean scores than participants from the nurse group (p=.002), participants from the medical staff group had statistically significantly lower mean scores than participants from the nurse group (p=.003), and participants from the patient group had statistically significantly lower mean scores than participants from the nurse group (p=.008). In the PoI-assisted brute-force attack, the mean number of guesses required to crack the passwords per group was as follows: *i)* family member: 43576.94 ± 7488.35; *ii)* friend: 71732.66 ± 17131.09; *iii)* medical staff: 99921.83 ± 18620.08; *iv)* patient: 112232.45 ± 22590.04; and *v)* nurse: 159116.55 ± 27663.24. In the personalized PoI-assisted brute-force attack, the mean number of guesses required to crack the passwords per group was as follows: *i)* family member: 28694.61 ± 4596.83; *ii)* friend: 52546.77 ± 19951.18; *iii)* medical staff: 55511.50 ± 11560.53; *iv)* patient: 62751.50 ± 11903.80; and *v)* nurse: 124987.88 ± 12485.52.

Figure 10 depicts the means of password strength among attack types by disregarding the order and the type of gestures across all participants.

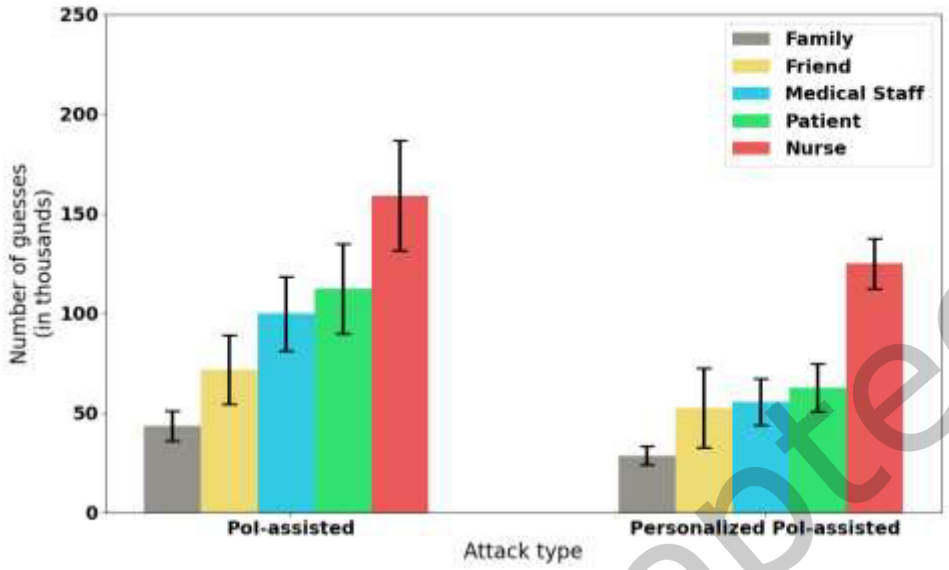


Figure 10. Means of password strength among attack types by disregarding the order and the type of gestures across all participants.

B. Disregarding order and type of gestures across participants with at least one gesture containing experience-spot. A one-way MANOVA was run to determine the effect of relationship between attackers and legitimate end-users on the number of guesses required to crack the passwords when using a PoI-assisted brute-force attack *vs.* a personalized PoI-assisted brute-force attack, by considering participants with at least one gesture containing experience-spot. Two measures were assessed: Number of guesses required to crack the passwords when using a PoI-assisted brute-force attack and number of guesses when using a personalized PoI-assisted brute-force attack. Participants belonged to one of the following categories: family member, friend, medical staff, patient, nurse. Preliminary assumption checking revealed that data was normally distributed, as assessed by Shapiro-Wilk test (p>.05); there were no univariate or multivariate outliers, as assessed by boxplot and Mahalanobis distance (p>.001), respectively; there were linear relationships, as assessed by scatterplot; no multicollinearity (r=.764, p<.0005); and there was homogeneity of variance-covariance matrices, as assessed by Box's M test (p=.001). The analysis revealed that the differences between groups on the combined dependent variables was statistically significant, $F(8, 96)=43.855$, p<.0005; *Wilks'* Λ=.046; *partial* η^2=.785. Follow-up univariate ANOVAs revealed that in both types of attacks the number of guesses required to crack the passwords was statistically significantly different between the participants from different relationship group (PoI-assisted brute-force attack: $F(4, 49)=66.535$, p=< .0005; *partial* η^2=.845; Personalized PoI-assisted brute-force attack: $F(4, 49)=81.915$, p<.0005; *partial* η^2=.870; Tukey-Kramer post-hoc tests showed that for the PoI-assisted brute-force attack, participants from the family group had statistically significantly lower mean scores than participants from the friend group (p=.032), the medical staff group (p<.0005), the patient group (p<.0005), and the nurse

group (p<.0005). Participants from the friend group had statistically significantly lower mean scores than participants from the medical staff group (p=.001), the patient group (p=.020), and the nurse group (p<.0005). Participants from the medical staff group had statistically significantly lower mean scores than participants from the nurse group (p<.0005), while participants from the patient group had statistically significantly lower mean scores than participants from the nurse group (p<.0005). Regarding the personalized PoI-assisted brute-force attack, participants from the family group had statistically significantly lower mean scores than participants from the medical staff group (p=.001), the patient group (p<.0005), and the nurse group (p<.0005). Participants from the friend group had statistically significantly lower mean scores than participants from the medical staff group (p=.002), the patient group (p<.0005), and the nurse group (p<.0005). Participants from the medical staff group had statistically significantly lower mean scores than participants from the patient group (p<.0005) and the nurse group (p<.0005), while participants from the patient group had statistically significantly lower mean scores than participants from the nurse group (p=.050). In the PoI-assisted brute-force attack, the mean number of guesses required to crack the passwords per group was as follows: *i)* family member: 25308.50 ± 1454.21; *ii)* friend: 32241.70 ± 1330.90; *iii)* medical staff: 41972.60 ± 2062.57; *iv)* patient: 39267.41 ± 1436.24; and *v)* nurse: 58851.83 ± 1498.21. In the personalized PoI-assisted brute-force attack, the mean number of guesses required to crack the passwords per group was as follows: *i)* family member: 7536.10 ± 351.76; *ii)* friend: 7815.60 ± 308.75; *iii)* medical staff: 12074.80 ± 394.72; *iv)* patient: 19233.83 ± 969.97; and *v)* nurse: 22047.91 ± 1000.61. **Figure 11** depicts the means of password strength among attack types by disregarding the order and the type of gestures across participants with at least one gesture containing experience-spot.

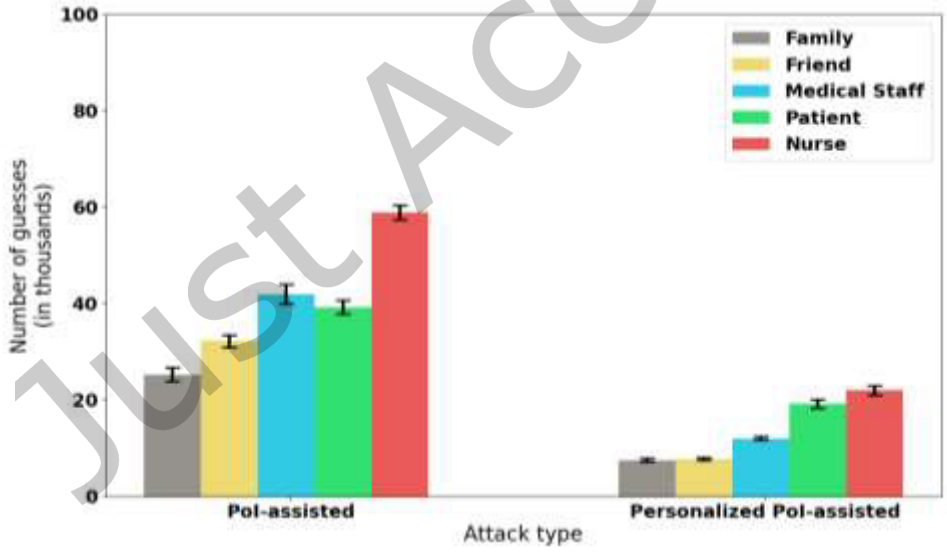


Figure 11. Means of password strength among attack types by disregarding the order and the type of gestures across participants with at least one gesture containing experience-spot.

C. Considering order and type of gestures across all participants. A one-way MANOVA was run to determine the effect of relationship between attackers and legitimate end-users on the number of guesses required to crack the passwords when using a PoI-assisted brute-force attack *vs.* a personalized PoI-assisted brute-force attack, by taking into account the order and type of gestures across all participants. Two measures were assessed: Number of guesses required to crack the passwords when using a PoI-assisted brute-force attack and number of guesses when using a personalized PoI-assisted brute-force attack. Participants belonged to one of the following categories: family member, friend, medical staff, patient, nurse. Data are expressed as mean ± standard error. Preliminary assumption checking revealed that data was normally distributed, as assessed by Shapiro-Wilk test (p>.05); there were no univariate or multivariate outliers, as assessed by boxplot and Mahalanobis distance (p>.001), respectively; there were linear relationships, as assessed by scatterplot; no multicollinearity (r=.183, p=.008); and there was homogeneity of variance-covariance matrices, as assessed by Box's M test (p=.012). The analysis revealed that the differences between groups on the combined dependent variables was not statistically significant, $F(8, 172)$=.020, p=0.99; *Wilks' Λ*= .998; partial η2 = .001. In the PoI-assisted brute-force attack, the mean number of guesses required to crack the passwords per group was as follows: *i)* family member: 1660935.38 ± 348113.52; *ii)* friend: 1687709.11 ± 447019.45; *iii)* medical staff: 1629472.16 ± 124750.15; *iv)* patient: 1624675.54 ± 849169.82; and *v)* nurse: 1622255.16 ± 66263.27. In the personalized PoI-assisted brute-force attack, the mean number of guesses required to crack the passwords per group was as follows: *i)* family member: 1745307.61 ± 253165.96; *ii)* friend: 1690173.83 ± 126111.69; *iii)* medical staff: 1789121.611 ± 217395.14; *iv)* patient: 1797917.49 ± 475354.52; and *v)* nurse: 1818772.27 ± 164674.17. **Figure 12** depicts the means of password strength among attack types by considering the order and the type of gestures across all participants.

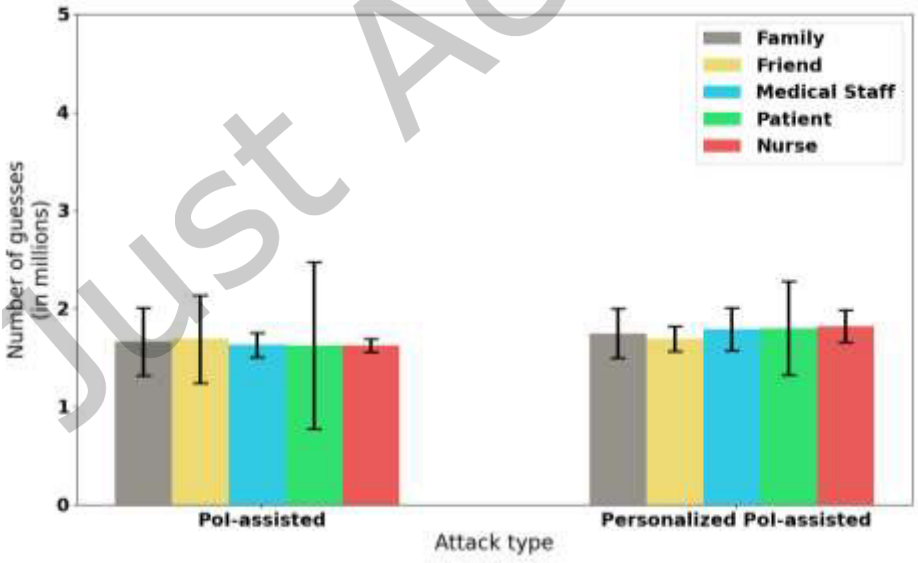


Figure 12. Means of password strength among attack types.

6.5.3 *Qualitative Analysis*

To further shed light and understand the approach followed by attackers on graphical passwords created on location-aware images, we used the data gathered from the feedback mechanism at the end of the study, as well as observations made by the researchers during the attack phase. In many cases, attackers used knowledge about the end-user under attack, related to their habits, preferences and facts about their personality: *"My colleague is a great storyteller and I believe he would try to create a story for the password, like arriving at the entrance of the hospital, then entering by the stairs, then reading information on the panel. So, I decided to make my attack having this story-telling process in mind." – P19; "I thought she will have used the three possible gestures (instead of just one or two of the options), and marked colorful items, because of her personality." – P28; "Most of the times he has his coffee in the front yard outside the emergency room during his shift break. I would be surprised if he hadn't selected this particular area." – P56; "My colleague likes flowers and plants. I think that some of her selections must be on the flowers." – P39; "She likes painting at her free time and I think she drew straight lines on objects that have bright colors." – P11.*

In other cases, it is evident that the scenery depicted on the location-aware images impacted the selections of the attackers. In particular, attackers used a more personalized approach by considering specific information related to their common shared experiences with the end-user under attack within the places depicted on the location-aware images: *"Usually we have lunch together at the hospital's cafeteria and we tend to be seated at the tables near the entrance. Hence, I made my selections around these tables." – P7; "Considering direction of movement; places he usually sits or similar." – P14; "The pictures are from daily routines, so I'm trying to guess where he is going on a daily basis or important aspects for him in the photo." – P29; "I work at the emergency department and the colleague that I am* requested *to guess his password is the ambulance driver. I think it is very possible that he made some of his password selections near or on the ambulance outside of the emergency department." – P8; "The photo shows two parking lots, but I know that she usually parks her car to the one next to the left entrance of the hospital because it is more convenient for her. I guess that some of her selections must be within this specific parking lot area." – P16; "Being a hospital receptionist and a friendly person that interacts daily with many patients, I think that she must have selected the people standing in front of the reception desk." – P33.*

In very few cases, attackers did not employ any sophisticated attack, but rather focused on the obvious PoIs of the images: *"Tried to guess likely features in the images, but type of gesture I used was just random." – P23; "I clicked on the most dominant views. I chose them because they caught my eye." – P34; "I believe he must have selected the chairs because these are the most visible points in the image." – P42.*

The above observations were concentrated in a coding schema relevant to the approach followed by the attackers as follows:

- Habits/Preferences/Characteristics of end-users (e.g., storytelling, coffee, flowers, painting);
- Common shared experiences (i.e., experiences within the depicted place/scenery);
- Random-guessing approach relying on areas of the image that attract peoples' attention (i.e., PoIs).

Table 7 summarizes the responses about the approach employed by the attackers based on the aforementioned coding schema.

Table 7: Summary of the approach followed by the attackers based on the coding schema extracted from data collected during the attack phase.

Attacking approach followed	Frequency
Habits/preferences/characteristics of end-users	30 out of 92
Common shared experiences	48 out of 92
Random-guessing approach	14 out of 92

6.6 Summary of Main Findings

The human guessing attack study revealed that the suggested user-adaptable and personalized authentication paradigm, which utilizes location-aware images for graphical passwords, increases guessing vulnerabilities in case someone knows the user, since analyses indicate that individuals that share common experiences may spot certain regions that the end-user used to create the graphical password gestures. In particular, the main findings of the analyses are as follows:

- human guessing vulnerabilities exist when we disregard the type of the gesture and the exact order, since in some cases participants from specific groups (*e.g.*, family member, friend) scored lower Euclidean distances than participants from other groups (*e.g.*, patient, nurse, and medical staff);
- human guessing vulnerabilities also exist when we disregard the type of the gesture but consider the exact order, since in some cases participants from specific groups (*e.g.*, family member, friend) scored lower Euclidean distances than participants from other groups (*e.g.*, patient, nurse, and medical staff);
- there were no human guessing vulnerabilities when we consider the type of the gesture and the exact order, since there was no successful attempt, yielding an online success guessing rate of 0%;
- with regards to the security strength when we disregard the order and type of gestures across all participants, we observed differences in the number of guesses required to crack the passwords using the PoI-assisted brute-force attack, since in some cases participants from specific groups (*e.g.*, family member, friend) scored lower number of guesses than participants from other groups (*e.g.*, nurse). Similarly, in the personalized PoI-assisted brute-force attack, participants from specific groups (*e.g.*, family member, friend, medical staff, patient) scored lower number of guesses than participants from other groups (*e.g.*, nurse);
- with regards to the security strength when we disregard the order and type of gestures across participants with at least one gesture containing experience-spot, we also observed differences in the number of guesses required to crack the passwords using the PoI-assisted brute-force attack, since in some cases participants from specific groups (*e.g.*, family member, friend, medical staff, patient) scored lower number of guesses than participants from other groups (*e.g.*, friend, medical staff, patient, nurse). Similarly, in the personalized PoI-assisted brute-force attack, participants from specific groups (*e.g.*, family member, friend, medical staff, patient) scored lower number of guesses than participants from other groups (*e.g.*, medical staff, patient, nurse);
- with regards to the security strength when we considering the order and type of gestures across all participants, there were no observed differences in the number of guesses required to crack the passwords using either the PoI-assisted brute-force attack or the personalized PoI-assisted brute-force attack.

Based on the aforementioned, we can conclude that some relationship groups are able to run a more effective attack by starting to guess regions they suspected that the users selected their password. Nonetheless, based on the brute-force attacks on the DuoPass graphical mechanism as a whole (*i.e.*, when

also considering the order and type of gesture[4]), this did not affect the security of the created graphical passwords on location-aware images.

7 DISCUSSION AND IMPLICATIONS

In this section we elaborate about the applicability of DuoPass in the broader healthcare domain and provide guidelines that can serve as a basis for implementing an adaptation and personalization system based on the suggested authentication paradigm, as well as the limitations of this research work.

We envision that DuoPass may be deployed as a standalone Web-based user authentication system within healthcare organizations, which will extend the existing textual-based password solutions that patients currently use to access their personal health records through the Web portal [38]. Given that DuoPass would rely on location-based experiences, habits, and memories created during patients' visits at the hospital, we anticipate that it would be more suitable for patients that visit the same hospital on a frequent or regular basis. At a first stage, an organization would need to identify mainstream spatial areas of the hospital, *i.e.*, areas that are visited by the majority of individuals (medical staff, patients, relatives, visitors, etc.). Next, the spatial relevance of each mainstream area should be identified in order to create a neighborhood/relationship map among the diverse mainstream spatial areas identified, *e.g.*, the mainstream spatial area "reception hall" is related to the hospital's "cafeteria", hence, a relationship rule would be created connecting the two areas. Finally, the system administrator would need to prepare and upload relevant images depicting sceneries for each of the identified mainstream of the hospital. These images would then be processed through an adaptation and recommendation engine that would recommend best-fit images to end-users aiming to improve memorability and security of passwords. For doing so, the recommendation engine would also receive as input the end-user's visitation record in order to extract the relevant experiences and visits the end-users had in specific mainstream spatial areas of the hospital. In this respect, DuoPass will leverage on the existing authentication infrastructure that exists in the healthcare organization for retrieving the user's visitation record.

The following scenarios are anticipated: *i) Enrolment Scenario:* during user enrolment, the system would retrieve (based on the username and a unique enrolment code) the user's visitation record within the hospital. Based on the semantic similarity of the user visits and the mainstream spatial areas of the hospital, the system would recommend three relevant images to choose from for creating their graphical password. Note that the three images would have the same level of complexity and PoIs to avoid scenarios in which the user would create predictable passwords; *ii) Login Scenario:* during login, the system would illustrate two options for authentication (graphical *vs.* textual), and accordingly the user would enter their secret credentials to login; and *iii) Reset Scenario:* password reset could be initiated either by the user (*e.g.*, in case they forget their password) or by the system based on the organization's applied policy. In this case, the same procedure would follow as in the enrolment scenario, considering however the previous image selections of the user, in order to avoid users selecting the same password.

DuoPass would consist of the following modules: *i)* the System Administration module; *ii)* the User Modeling module; *iii)* the Recommendation module; and *iv)* the Flexible User Authentication module. The System Administration module would allow administrators to upload and maintain images that depict sceneries of various locations of the hospital (*e.g.*, reception hall, main rooms of the hospital). The system's

image database would also be filled by end-users, who would be able to upload their own images taken within the hospital, once approved by the system administrator by following organizational policies and requirements. The User Modeling module would analyze the existing health record of the patients based on their activity and visits at the hospital (*e.g.*, patient may visit doctors of the ophthalmology department, orthopedic department, etc.). Based on the analysis, the module would infer the patient's frequent visits and important locations within the hospital, which would be then provided as input to the Recommendation module to recommend images depicting sceneries from the patient's most common visits. The Recommendation module would be further enhanced with image analysis technologies aiming to semantically automatically annotate the images with the depicted content, which may be used during password creation for recommendation and user guidance for the creation of more memorable and secure passwords. Finally, the Flexible User Authentication module would be responsible for authenticating users based on an easy-to-use and a flexible authentication paradigm that would be based on the recommended and/or user-adaptable graphical passwords.

Algorithm #1 in the Appendix presents our content-based recommendation algorithm that will recommend relevant images during password creation/reset based on the rules of mainstream spatial areas and user's visitation records and experiences within the hospital. The algorithm initially requires configuration by the organization in terms of identifying the mainstream spatial areas of the hospital and then creating the relationship map between them. Next, the set of candidate images is generated as follows: *i)* the system administrator uploads images that depict the mainstream spatial areas of the hospital, as well as approves relevant images within the hospital that were provided by the end-users; *ii)* a set of tags and sentences that describe the semantic content of the image is generated by explicit (*i.e.*, annotated by the system administrator) and implicit (*i.e.*, annotated by computer vision techniques for object and label detection[5,6,7]) methods; and *iii)* the set of tags and sentences is pre-processed and cleaned. The part of image recommendation involves the following steps: *i)* for each user, the frequent visits and locations within the hospital are inferred from their existing health records based on their activity and visits; *ii)* a set of tags and sentences that describe the semantic content of the users' visits and locations is generated; *iii)* the set of tags and sentences is then filtered to contain relevant information based on the relationship map between the mainstream areas; *iv)* the set of tags and sentences is pre-processed and cleaned; *v)* for each image in the set of candidate images, a semantic similarity score is calculated (*i.e.*, through Natural Language Processing (NLP) techniques (*e.g.*, BERT [34])) between the set of tags and sentences that describe the users' frequent visits and locations and the set of tags and sentences that describe the semantic content of the images; and *vi)* finally, the semantic similarity scores are sorted and the top *N* images are recommended to each end-user.

7.1 Limitations

Despite our efforts to keep the validity of the study, some design aspects of the experiments introduce limitations. We used specific personalized location-aware images in order to control the factors of the

[5] *Google Cloud Vision, https://cloud.google.com/vision*
[6] *Amazon Rekognition, https://aws.amazon.com/rekognition*
[7] *Tensorflow, https://www.tensorflow.org*

study (location-aware *vs.* non-location-aware images). Although users' choices may be affected by the content and complexity of the image [37, 105], we provided images of the most widely used image categories (*i.e.*, depicting sceneries and people [4, 37]) and of similar complexity [28, 59]. Although works exist on location-based authentication [2, 3, 85], expansion of our research will also consider a greater variety of location-aware image categories for triangulating the findings with diverse user communities and location-based experiences on different levels of abstractions (*i.e.*, individual, group, organizational, national, global) [28], and thus increase the validity of the study. Furthermore, the proposed personalization approach in DuoPass was compared against one baseline generic approach. Nevertheless, this was intentional in order to get comparable results, which probably would not be the case had we compared the suggested non-intrusive personalized approach against other intrusive approaches (*e.g.*, "presentation effect" [100], hiding salient areas [18], etc.). Also, in order to control the similarity of image factors in terms of complexity and PoIs, we intentionally did not compare the suggested approach against user-uploaded images which could have introduced images of varying complexity and PoIs.

Moreover, considering that DuoPass relies on location-based experiences, habits, and memories created during patients' visits at the hospital, it is probable that patients could share similar experiences with other patients or with other people that are close to them and know their habits (*e.g.*, enter the building through the same entrance, walk in the same corridor, visit the same hospital's cafeteria and order the same drink, visit the hospital with their accompanying caregiver, etc.). Although such scenarios could entail password guessing vulnerabilities in terms of allowing people sharing same experiences or are close to them to more easily identify regions of their selected secrets [26, 28], the human guessing attack study we conducted revealed that the security of such a personalized graphical password mechanism is not compromised when additional measures (*i.e.*, type and order of gestures) are considered. Furthermore, similar to most graphical password systems, DuoPass is also susceptible to shoulder-surfing attacks [99], as it was not designed to account for such threat scenarios. In the case that the username, the image, and the gestures are observed through shoulder-surfing, then an attacker has all the information needed to break in to the account, as is the case with most other graphical password systems [22]. Another challenge of DuoPass relates to generating and maintaining a diverse pool of location-aware set of images, in order to form a dictionary that contains adequate images that people can reflect upon based on their experiences.

Also, we stress that DuoPass graphical authentication mechanism primarily relies on visual elements, requiring end-users to perceive, process and recall visual information, and accordingly select certain regions on an image by using human motor functions, *i.e.*, by pointing on and selecting secret regions of the image through a computer mouse or finger input on a touch screen. Consequently, such graphical user authentication systems create accessibility issues for some user populations that might have visual and/or human motor difficulties. To address visual accessibility issues, Braille code-based images and haptics could be used in the graphical user authentication process. Such an approach would require utilizing and/or implementing certain hardware and software technology for storing the Braille code in the DuoPass system, and end-users to read and select secret regions of the Braille code image through haptic technology. However, many people with vision difficulties do not actually know or use Braille, and Braille also differs largely across countries (*e.g.*, British Braille, American Braille). Hence, it is more likely that people with vision difficulties would use speech recognition for passwords or biometric passwords.

Another limitation relates to getting useful patients' visitation records to form relevant image recommendations. We envision that DuoPass' User Modeling module could be extended by existing third-party services, such as, indoor positioning systems that track locations and activities of individuals while they are within the premises of an organization (*e.g.*, healthcare institution). Nonetheless, such an extension would require additional infrastructure and usage of third-party services that might increase the operational costs of the organization.

Finally, due to the inherent nature of memory, the suggested personalization approach of DuoPass might be practical for patients who: *i)* visit the same hospital on a frequent or regular basis; *ii)* do not suffer from memory impairment (*e.g.*, Alzheimer's Disease, Dementia); and *iii)* are familiar with specific information of healthcare locations and are able to memorize it (*e.g.*, patients that required emergency hospitalization due to an accident or were unconscious during their visit at the hospital might not be able to memorize the locations of the hospital). Nonetheless, in cases of one-off patients (*i.e.*, patients that visit hospitals/clinics once a year for routine check-ups or required emergency hospitalization and are not able to memorize healthcare locations), DuoPass will be configured to not recommend personalized location-aware images but instead will recommend state-of-the-art non-location-aware/generic images that have been previously approved by the system administrator in terms of complexity and policies. Expansion of our research will also consider the practicability of DuoPass with diverse patient communities, such as, elderly population with memory impairment.

8 CONCLUSIONS

This paper presents a novel knowledge-based user authentication paradigm, which aims to provide a secure, memorable and patient-centric authentication solution within current highly heterogeneous computational realms of healthcare environments. Results of a feasibility study, during which users interacted with the suggested authentication paradigm, revealed significant differences on users' password selections falling into Points-of-Interest regions of the images and subsequently on the security strength of the selected graphical passwords between the experimental and control groups. Furthermore, there was no interaction between the user group and users' password selections on the time to compose the graphical password, however, the experimental group required significantly less time to create the last (third) selection of their password compared to the control group. Simultaneously, both experimental and control groups performed similarly in terms of memorability and login efficiency. Moreover, responses from the post-study survey revealed that the suggested paradigm scored high in terms of users' likeability, perceived security, usability and trust. On the downside, the suggested paradigm introduces password guessing vulnerabilities in terms of allowing attackers, who share common experiences with the end-users, to identify regions of the end-users' selected secrets more easily. Nonetheless, the results of the human guessing attack revealed that the security of the suggested paradigm is not compromised when additional measures (*i.e.*, type and order of gestures) are considered.

We anticipate that the suggested approach will have a positive impact on both healthcare organizations and end-users. From the organization's perspective, the flexible approach will assist healthcare organizations to easily adjust their policies to the varying roles of their end-users (patients, doctors, nurses), in which current practice indicates that the "one-size-fits-all" approach is not adequate in the

highly dynamic and heterogeneous contexts of use in the healthcare domain. From the end-user's perspective, the suggested flexible and personalized paradigm and supported results open new directions for considering novel knowledge-based user authentication mechanisms to assist end-users to choose the "best-fit" authentication scheme depending on preference, unique characteristics and the context of interaction (*e.g.*, interaction in the office, in the emergency room, off the network, etc.).

From a procedural perspective, given that DuoPass is solely based on knowledge-based authentication approaches, it is less expensive compared to token-based and biometric-based solution, which entail increased implementation and maintenance costs, but at the same time results of this study reveal increased security and a positive user experience. In addition, through the flexible and adaptable character of DuoPass (*i.e.*, shift between graphical and textual passwords), it still supports the user interactions within the current state-of-the-art authentication approaches in the healthcare domain, which is solely based on traditional text-based approaches. In this respect, DuoPass also adapts easily to current multi-factor authentication approaches, *i.e.*, DuoPass can be used as the first step during authentication, and any additional layer may be added as a following step to increase security.

A side-effect of the approach relates to the password creation efficiency since users require more time to create their password (graphical and textual) than the traditional approach (only textual). Nonetheless, the majority of the participants commented that this has not negatively affected their likeability towards DuoPass, *e.g.*, a user commented that *"The creation phase happens only once so I'm ok with that"*. Future work will focus on improving the efficiency of the password creation phase with alternative visual and interaction designs. Furthermore, the open-ended nature of the suggested authentication paradigm raises new security threats and might affect users towards misuse strategies that need to be carefully addressed. In order to assure that users will not create semantically insecure (predictable) selections on images as a side effect of allowing them to create their own images for the graphical passwords, automated image tagging technologies will be used to prevent users' unsafe coping strategies. Furthermore, evidence suggests that individuals, in an attempt to reduce the memory load of remembering multiple passwords, tend to reuse the same or similar passwords across multiple accounts [13], which has a negative impact on the security. Hence, another future research prospect would be to investigate whether differences exist in individuals' perceptions about reduced memory load between individuals who utilize location-aware images in DuoPass and individuals who utilize non-location-aware images in other graphical user authentication schemes. Also, future work entails investigating whether differences exist in password reuse approach between individuals who utilize location-aware images in DuoPass and individuals who utilize non-location-aware images in other graphical user authentication schemes.

Bearing in mind that within nowadays information era, patients and medical staff interact in highly dynamic healthcare environments and contexts, and tend to use multiple devices to authenticate themselves, it is obvious that the current widely deployed "one-size-fits-all" text-based authentication paradigm might soon become obsolete. Hence, we believe that approaches like DuoPass provide an alternative solution to current state-of-the-art research and practice, and have the potential to be easily adopted with a rather inexpensive solution compared to other token-based (*e.g.*, smartcards) and biometric-based solutions (*e.g.*, fingerprint), which necessitate increased implementation and maintenance

costs. Although initial experiments are promising, further studies are required to evaluate DuoPass in the wild with the aim to get further insights on its validity, user acceptance and real-world user behavior.

ACKNOWLEDGEMENTS

We sincerely thank the participants of the healthcare organizations (Zuyderland Medical Center, Netherlands; Hospital Clinic Barcelona, Spain; Western General Hospital within NHS Lothian, Scotland, UK) for their time and efforts in conducting the user needs' verification and evaluation studies, and the valuable feedback received during the surveys and focus groups. This research has been partially supported by the EU Horizon 2020 Grant 826278 "Securing Medical Data in Smart Patient-Centric Healthcare Systems" (Serums), and the Research and Innovation Foundation (Project DiversePass: COMPLEMENTARY/0916/0182).

REFERENCES

[1] Abdellaoui, A., Khamlichi, Y.I., Chaoui, H. (2016). A robust authentication scheme for telecare medicine information system. Procedia Computer Science, 58, 584-589. DOI: https://doi.org/10.1016/j.procs.2016.09.091

[2] Al-Ameen, M.N., Wright, M. (2015). Multiple-password interference in the geopass user authentication scheme. In Proc. Workshop Usable Secur.(USEC), 1-6. DOI: http://dx.doi.org/10.14722/usec.2015.23004

[3] Al-Ameen, M.N., Wright, M. (2017). Exploring the potential of geopass: A geographic location-password scheme. Interacting with Computers, 29(4), 605-627. DOI: https://doi.org/10.1093/iwc/iww033

[4] Alt, F., Schneegass, S., Shirazi, A.S., Hassib, M., Bulling, A. (2015). Graphical passwords in the wild: Understanding how users choose pictures and passwords in image-based authentication schemes. In Proceedings of MobileHCI 2015, 316-322. DOI: http://dx.doi.org/10.1145/2785830.2785882

[5] Amin, R., Islam, S.H., Biswas, G.P., Khan, M.K., Obaidat, M.S. (2015). Design and analysis of an enhanced patient-server mutual authentication protocol for telecare medical information system. Journal of Medical Systems, 39(11), 1-20. DOI: https://doi.org/10.1007/s10916-015-0307-2

[6] Atkinson, R.C., Shiffrin, R.M. (1968). Human memory: a proposed system and its control processes. In: Spence, K.W., Spence, J.T. (eds.), The psychology of learning and motivation (Volume 2), Academic Press, 89-195. DOI: https://doi.org/10.1016/S0079-7421(08)60422-3

[7] Avancha, S., Baxi, A., Kotz, D. (2012). Privacy in mobile technology for personal healthcare. ACM Computing Surveys, 45(1), 1-54. DOI: https://doi.org/10.1145/2379776.2379779

[8] Ayalon, O., Toch, E. (2019). Evaluating users' perceptions about a system's privacy: differentiating social and institutional aspects. Symposium on Usable Privacy and Security (SOUPS 2019), USENIX Association, 41-59

[9] Baddeley, A. (1990). Human memory: theory and practice. Lawrence-Erlbaum, London

[10] Bates, D., Mächler, M., Bolker, B., Walker, S., (2014). Fitting linear mixed-effects models using lme4. DOI: https://doi.org/10.18637/jss.v067.i01

[11] Belk M., Fidas C., Germanakos P., Samaras G. (2017). The interplay between humans, technology and user authentication: a cognitive processing perspective. Computers in Human Behavior, 76, 184-200. DOI: https://doi.org/10.1016/j.chb.2017.06.042

[12] Belk, M., Fidas, C., Pitsillides, A. (2019). FlexPass: Symbiosis of seamless user authentication schemes in IoT, ACM SIGCHI Human Factors in Computing Systems (CHI 2019), ACM Press, LBW2318, 1–6. DOI: https://doi.org/10.1145/3290607.3312951

[13] Biddle, R., Chiasson, S., van Oorschot, P. (2012). Graphical passwords: Learning from the first twelve years. ACM Computing Surveys, 44(4), 41 pages. DOI: https://doi.org/10.1145/2333112.2333114

[14] Blasco, J., Chen, T., Tapiador, J., Peris-Lopez, P. (2016). A Survey of wearable biometric recognition systems. ACM Computing Surveys, 49, 3, article 43, 35 pages. DOI: https://doi.org/10.1145/2968215

[15] Bonneau, J., Herley, C., van Oorschot, P.C., Stajano, F. (2012). The quest to replace passwords: a framework for comparative evaluation of web authentication schemes. Security and Privacy, IEEE Symposium, 553-567. DOI: https://doi.org/10.1109/SP.2012.44

[16] Bowles, J., Mendoza-Santana, J., Webber, T. (2020). Interacting with next-generation smart patient-centric healthcare systems. Adaptive and Personalized Privacy and Security Workshop (APPS 2020), UMAP (Adjunct Publication), ACM Press, 191–192. DOI: https://doi.org/10.1145/3386392.3399561

[17] Brooke, J. (1986). SUS: a "quick and dirty" usability scale. In Jordan, P.W., Thomas, B.,; Weerdmeester, B.A., McClelland, A.L. (eds.). Usability Evaluation in Industry. London: Taylor and Francis

[18] Bulling, A., Alt, F., Schmidt, A. (2012). Increasing the security of gaze-based cued-recall graphical passwords using saliency masks. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 3011-3020. DOI: https://doi.org/10.1145/2207676.2208712

[19] Burr, W.E., Dodson, D.F., Polk, W.T. (2006). Electronic authentication guideline. National Institute of Standards and Technology, Technical report

[20] Cardaci, M., Di Gesù, V., Petrou, M., Tabacchi, M.E. (2009). A fuzzy approach to the evaluation of image complexity. Fuzzy Sets and Systems, 160(10), 1474-1484. DOI: https://doi.org/10.1016/j.fss.2008.11.017

[21] Chiasson, S., van Oorschot, P., Biddle, R. (2006). A usability study and critique of two password managers. Security Symposium, USENIX Association

[22] Chiasson, S., van Oorschot, P. and Biddle, R. (2007). Graphical password authentication using cued click points. In European Symposium on Research in Computer Security (pp. 359-374). Springer, Berlin, Heidelberg. DOI: https://doi.org/10.1007/978-3-540-74835-9_24

[23] Choudhary, T., Manikandan, M.S. (2016). Robust photoplethysmographic (PPG) based biometric authentication for wireless body area networks and m-health applications. National Conference on Communication (NCC), 1-6. DOI: https://doi.org/10.1109/NCC.2016.7561152

[24] Cohen, J. (2013). Statistical power analysis for the behavioral sciences. Routledge.

[25] Constantinides, A., Belk, M., Fidas, C., Pitsillides, A. (2020a). Design and development of a patient-centric user authentication system. ACM User Modeling Adaptation and Personalization (UMAP 2020 Adjunct Publication), ACM Press, 201-203. DOI: https://doi.org/10.1145/3386392.3399564

[26] Constantinides, A., Belk, M., Fidas, C., Pitsillides, A. (2021a). Understanding Insider Attacks in Personalized Picture Password Schemes. In IFIP Conference on Human-Computer Interaction, 722-731. DOI: https://doi.org/10.1007/978-3-030-85610-6_42

[27] Constantinides, A., Belk, M., Fidas, C., Samaras, G. (2018a). On cultural-centered graphical passwords: leveraging on users' cultural experiences for improving password memorability. In Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization (pp. 245-249). DOI: https://doi.org/10.1145/3209219.3209254

[28] Constantinides, A., Fidas, C., Belk, M., Pietron, A., Han, T., Pitsillides, A. (2021b). From hot-spots towards experience-spots: Leveraging on users' sociocultural experiences to enhance security in cued-recall graphical authentication, International Journal of Human-Computer Studies, 149 (2021), 102602. DOI: https://doi.org/10.1016/j.ijhcs.2021.102602

[29] Constantinides, A., Fidas, C., Belk, M., Samaras, G. (2018b). On sociocultural-centered graphical passwords: an initial framework. In Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (pp. 277-284). DOI: https://doi.org/10.1145/3236112.3236150

[30] Constantinides, A., Pietron, A.M., Belk, M., Fidas, C., Han, T., Pitsillides, A. (2020b). A cross-cultural perspective for personalizing picture passwords. In Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization (pp. 43-52). DOI: https://doi.org/10.1145/3340631.3394859

[31] Coventry, L., Branley, D. (2018). Cybersecurity in healthcare: a narrative review of trends, threats and ways forward. Maturitas, 113, 48-52. DOI: https://doi.org/10.1016/j.maturitas.2018.04.008

[32] Cranor, L.F. (2014). What's wrong with your pa$$w0rd? TED Talk, March 2014

[33] Das, A.K. (2015). A secure user anonymity-preserving three-factor remote user authentication scheme for the telecare medicine information systems. Journal of Medical Systems, 39 (3). DOI: https://doi.org/10.1007/s10916-015-0218-2

[34] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. DOI: https://doi.org/10.48550/arXiv.1810.04805

[35] Dhillon, P.K., Kalra, S. (2018). Multi-factor user authentication scheme for IoT-based healthcare services. Journal of Reliable Intelligent Environments, 4(3), 141-160. DOI: https://doi.org/10.1007/s40860-018-0062-5

[36] Dineen, L.C., Blakesley, B. C. (1973). Algorithm AS 62: Generator for the sampling distribution of the Mann-Whitney U statistic. Applied Statistics, 22, 269-273. DOI: https://doi.org/10.2307/2346934

[37] Dunphy, P., Yan, J. (2007). Do background images improve "draw a secret" graphical passwords?. In Proceedings of the 14th ACM conference on Computer and communications security, 36-47. DOI: https://doi.org/10.1145/1315245.1315252

[38] Eikey, E. V., Murphy, A.R., Reddy, M.C., Xu, H. (2015). Designing for privacy management in hospitals: Understanding the gap between user activities and IT staff's understandings. International Journal of Medical Informatics, 84(12), 1065-1075. DOI: https://doi.org/10.1016/j.ijmedinf.2015.09.006

[39] Farke, F., Lorenz, L., Schnitzler, T., Markert, P., Dürmuth, M. (2020). "You still use the password after all" – Exploring FIDO2 security keys in a small company. Symposium on Usable Privacy and Security, USENIX Association, 19-35

[40] Fatima, K., Nawaz, S., Mehrban, S., (2019). Biometric authentication in health care sector: A survey. Conference on Innovative Computing (ICIC 2019), 1-10. DOI: https://doi.org/10.1109/ICIC48496.2019.8966699

[41] Fernández-Alemán, J.L., Señor, I.C., Lozoya, P., Toval, A. (2013). Security and privacy in electronic health records: A systematic literature review. Journal of Biomedical Informatics, 46(3), 541-562. DOI: https://doi.org/10.1016/j.jbi.2012.12.003

[42] Fidas, C., Belk, M., Hadjidemetriou, G., Pitsillides, A. (2019). Influences of mixed reality and human cognition on picture passwords: An eye tracking study. IFIP TC13 Human-Computer Interaction (INTERACT 2019), Springer-Verlag, 304-313. DOI: https://doi.org/10.1007/978-3-030-29384-0_19

[43] Fidas, C., Belk, M., Portugal, D., Pitsillides, A. (2021). Privacy-preserving biometric-driven data for student identity management: Challenges and approaches. Adaptive and Personalized Privacy and Security, Adjunct Proceedings of ACM User Modeling Adaptation and Personalization (UMAP 2021), ACM Press. DOI: https://doi.org/10.1145/3450614.3464470

[44] Giri, D., Maitra, T., Amin, R., Srivastava, P.D. (2015). An efficient and Robust RSA-based remote user authentication for telecare medical information systems. Journal of Medical Systems, 39(1), 1-9. DOI: https://doi.org/10.1007/s10916-014-0145-7

[45] Goncalves, R., Leonova, E., Puttini, R., Nascimento, A. (2015). A privacy-ensuring scheme for health data outsourcing. International Conference on Cloud Technologies and Applications (CloudTech), IEEE, 1-7. DOI: https://doi.org/10.1109/CloudTech.2015.7336982

[46] Guennouni, S., Mansouri, A., Ahaitouf, A. (2019). Biometric systems and their applications. Eye Tracking and New Trends. IntechOpen, 1-12. DOI: https://doi.org/10.5772/intechopen.84845

[47] Halunen, K., Häikiö, J., Vallivaara, V. (2017). Evaluation of user authentication methods in the gadget-free world. Pervasive Mobile Computing, 40, 220-241. DOI: https://doi.org/10.1016/j.pmcj.2017.06.017

[48] He, D., Kumar, N., Chilamkurti, N., Lee, J.-H. (2014). Lightweight ecc based rfid authentication integrated with an id verifier transfer protocol. Journal of Medical Systems, 38(10), 1-6. DOI: https://doi.org/10.1007/s10916-014-0116-z

[49] Heckle, R., Lutters, W. (2011). Tensions of network security and collaborative work practice: Understanding a single sign-on deployment in a regional hospital. International Journal of Medical Informatics, 80(8), e49-e61. DOI: https://doi.org/10.1016/j.ijmedinf.2011.02.001

[50] Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. Journal of Educational Statistics, 6(2), pp.107-128. DOI: https://doi.org/10.3102/10769986006002107

[51] Herley, C., van Oorschot, P. (2012). A research agenda acknowledging the persistence of passwords. Security and Privacy, 10(1), 28-36. DOI: https://doi.org/10.1109/MSP.2011.150

[52] Ibrahim, A., Mahmood, B., Singhal, M. (2016). A secure framework for medical information exchange (MI-X) between healthcare providers. International Conference on Healthcare Informatics (ICHI), IEEE, 234-243. DOI: https://www.doi.org/10.1109/ICHI.2016.33

[53] Imura, S., Hosobe, H. (2016). Biometric authentication using the motion of a hand. Symposium on Spatial User Interaction (SUI 2016), ACM, 221. DOI: https://doi.org/10.1145/2983310.2989210

[54] Jain, A.K., Nandakumar, K., Ross, A. (2016). 50 Years of biometric research: accomplishments, challenges, and opportunities. Pattern Recognition Letters, 79(Supplement C), 80-105. DOI: https://doi.org/10.1016/j.patrec.2015.12.013

[55] Janjic, V., Bowles, J.K.F., Vermeulen, A. F., Silvina, A., Belk, M., Fidas, C., Pitsillides, A., Kumar, M., Rossborry, M., Vinov, M., Given-Wilson, T., Legay, A., Blackledge, E., Arredouani, R., Stylianou, G., Huang, W. (2019). The SERUMS tool-chain: ensuring security and privacy of medical data in smart patient-centric healthcare systems. IEEE Big Data 2019, IEEE Press, 2726-2735. DOI: https://www.doi.org/10.1109/BigData47090.2019.9005600

[56] Jayabalan, M., O'Daniel, T. (2019). A study on authentication factors in electronic health records. Applied Technology and Innovation, 3(1), 7-14

[57] Jiang, Q., Khan, M.K., Lu, X., Ma, J., He, D. (2016). A privacy preserving three-factor authentication protocol for e-Health clouds. Journal of Supercomputing, 72(10), 3826-3849. DOI: https://doi.org/10.1007/s11227-015-1610-x

[58] Johnson, J., Seixeiro, S., Pace, Z., van der Bogert, G., Gilmour, S., Siebens, L., Tubbs, K., Microsoft Corp (2014). Picture gesture authentication. U.S. Patent No. 8,650,636. Retrieved from https://google.com/patents/US8910253

[59] Katsini, C., Fidas, C., Raptis, G.E., Belk, M., Samaras, G. Avouris, N. (2018). Influences of human cognition and visual behavior on password strength during picture password composition. ACM Human Factors in Computing Systems (CHI 2018), ACM Press, 1-14. DOI: https://doi.org/10.1145/3173574.3173661

[60] Khan, M., Sakamura, K. (2015). Tamper-resistant security for cyber-physical systems with eTRON architecture. International Conference on Data Science and Data Intensive Systems, IEEE, 196-203. DOI: https://www.doi.org/10.1109/DSDIS.2015.98

[61] Kogetsu, A., Ogishima, S., Kato, K. (2018). Authentication of patients and participants in health information exchange and consent for medical research: A key step for privacy protection, respect for autonomy, and trustworthiness. Frontiers in Genetics, 9, 167. DOI: https://doi.org/10.3389/fgene.2018.00167

[62] Komanduri, S., Shay, R., Kelley, P., Mazurek, M., Bauer, L., Christin, N., Cranor, L., Egelman, S. (2011). Of passwords and people: Measuring the effect of password-composition policies. ACM Conference on Human Factors in Computing Systems (CHI 2011), ACM Press, 2595-2604. DOI: https://doi.org/10.1145/1978942.1979321

[63] Koppel, R., Smith, S., Blythe, J., Kothari, V. (2015). Workarounds to computer access in healthcare organizations: you want my password or a dead patient? Studies in Health Technology Informatics, 208, 215-220. DOI: https://www.doi.org/10.3233/978-1-61499-488-6-215

[64] Kothari, V., Koppel, R., Mare, S., Rudkin, S., Thimbleby, H. (2017). On developing authentication solutions for healthcare settings. Panel at Who are you? Adventures in Authentication Workshop (WAY 2017), Symposium on Usable Privacy and Security (SOUPS 2017), USENIX Association

[65] Kumar, T., Braeken, A., Jurcut, A.D., Liyanage, M., Ylianttila, M. (2020). AGE: authentication in gadget-free healthcare environments. Information Technoly Management, 21(2), 95-114. DOI: https://doi.org/10.1007/s10799-019-00306-z

[66] Kumar, V., Jangirala, S., Ahmad, M. (2018). An efficient mutual authentication framework for healthcare system in cloud computing. Journal of Medical Systems, 42, article 142. DOI: https://doi.org/10.1007/s10916-018-0987-5

[67] Lang, J., Czeskis, A., Balfanz, D., Schilder, M., Srinivas, S. (2016). Security keys: Practical cryptographic second factors for the modern web. International Conference on Financial Cryptography and Data Security, 422-440. DOI: https://doi.org/10.1007/978-3-662-54970-4_25

[68] Lee, K., Kaiser, B., Mayer, J., Narayanan, A. (2020). An empirical study of wireless carrier authentication for SIM swaps. Symposium on Usable Privacy and Security (SOUPS 2020), USENIX Association, 61-79

[69] Lee, T., Chang, I., Wang, C. (2013). Simple group password-based authenticated key agreements for the integrated EPR information system. Journal of Medical Systems, 37(2), 1-6. DOI: https://doi.org/10.1007/s10916-012-9916-1

[70] Leon, B., Boštjan, B. (2019). Rejecting the death of passwords: Advice for the future. Computer Science and Information Systems, 16(1), 313-332. DOI: https://doi.org/10.2298/CSIS180328016B

[71] Liu, C., Chung, Y. (2017). Secure user authentication scheme for wireless healthcare sensor networks. Computer Electronic Engineering, 59, 250-261. DOI: https://doi.org/10.1016/j.compeleceng.2016.01.002

[72] Mainali, P., Shepherd, C., Petitcolas, F. (2019). Privacy-enhancing context authentication from location-sensitive data. International Conference on Availability, Reliability and Security (ARES 2019), ACM Press, article 87, 1-10. DOI: https://doi.org/10.1145/3339252.3340334

[73] Mare, S., Baker, M., Gummeson, J. (2016). A study of authentication in daily life. Symposium on Usable Privacy and Security (SOUPS 2016), USENIX Association, 189-206

[74] Mason, J., Dave, R., Chatterjee, P., Graham-Allen, I., Esterline, A., Roy, K. (2020). An investigation of biometric authentication in the healthcare environment. Array 8, 100042. DOI: https://doi.org/10.1016/j.array.2020.100042

[75] Memon, Q., AlKassim, Z., Al Hassan, E., Omer, M., Alsiddig, M. (2017). Audio-visual biometric authentication for secured access into personal devices. Bioinformatics and Biomedical Science (ICBBS 2017), ACM Press, 85-89. DOI: https://doi.org/10.1145/3121138.3121165

[76] Murphy, A., Reddy, M., Xu, H. (2014). Privacy practices in collaborative environments: A study of emergency department staff. ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW 2014), ACM Press, 269-282. DOI: https://doi.org/10.1145/2531602.2531643

[77] Muzammal, S.M., Shah, M.A., Zhang, S.-J., Yang, H.-J. (2016). Conceivable security risks and authentication techniques for smart devices: A comparative evaluation of security practices. International Journal of Automation and Computing, 13, 350–363. DOI: https://doi.org/10.1007/s11633-016-1011-5

[78] Nascimento, M., Batista, L., Cavalcanti, N. (2015). A new approach to biometric recognition based on hand geometry. ACM Symposium on Applied Computing (SAC 2015), ACM Press, 59-65. DOI: https://doi.org/10.1145/2695664.2695801

[79] Okoh, E., Awad, A.I. (2015). Biometrics applications in e-health security: A preliminary survey. International Conference on Health Information Science, Springer, 92-103. DOI: https://doi.org/10.1007/978-3-319-19156-0_10

[80] Paivio, A. (2006). Mind and its evolution: A dual coding theoretical approach. Lawrence-Erlbaum, Mahwah, NJ

[81] Perazzi, F., Krähenbühl, P., Pritch, Y., Hornung, A. (2012). Saliency filters: Contrast based filtering for salient region detection. International Conference on Computer Vision and Pattern Recognition, IEEE, 733-740. DOI: https://www.doi.org/10.1109/CVPR.2012.6247743

[82] Pinheiro, J., Bates, D. (2006). Mixed-effects models in S and S-PLUS. Springer Science & Business Media

[83] Raptis, G.E., Katsini, C., Cen, A.J.L., Arachchilage, N.A.G. and Nacke, L.E. (2021). Better, Funner, Stronger: A Gameful Approach to Nudge People into Making Less Predictable Graphical Password Choices. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-17). DOI: https://doi.org/10.1145/3411764.3445658

[84] Reese, K., Smith, T., Dutson, J., Armknecht, J., Cameron, J., Seamons, K. (2019). A usability study of five two-factor authentication methods. Symposium on Usable Privacy and Security (SOUPS 2019), USENIX Association

[85] Renaud, K.V. (2009). Guidelines for designing graphical authentication mechanism interfaces. International Journal of Information and Computer Security, 3(1), 60-85. DOI: https://doi.org/10.1504/IJICS.2009.026621

[86] Sadovnik, A. and Chen, T. (2013). A visual dictionary attack on Picture Passwords. In 2013 IEEE International Conference on Image Processing (pp. 4447-4451). IEEE. DOI: https://www.doi.org/10.1109/ICIP.2013.6738916

[87] Samuel, R., Markert, P., Aviv, A., Neamtiu, I. (2020). Knock, knock. Who's there? On the security of LG's knock codes. Symposium on Usable Privacy and Security (SOUPS 2020). USENIX Association, 37-59

[88] Saravanos, A., Zheng, D., Zervoudakis, S., Delfino, D., Hynes-Keller, L. (2021). An Exploration of Hot-Spots in Locimetric Passwords. Available at https://arxiv.org/abs/2102.13241v1

[89] Sauro, J. (2011). Measuring usability with the System Usability Scale (SUS). Available at https://measuringu.com/sus

[90] Schwartze, J., Haarbrandt, B., Fortmeier, D., Haux, R., Seidel, C. (2013). Authentication systems for securing clinical documentation workflows. A systematic literature review. Methods in Information Medicine, 53(1), 3-13. DOI: https://www.doi.org/10.3414/ME12-01-0078

[91] Shay, R., Bauer, L., Christin, N., Cranor, L., Forget, A., Komanduri, S., Mazurek, M., Melicher, W., Segreti, S., Ur, B. (2015). A spoonful of sugar? The impact of guidance and feedback on password-creation behavior. ACM Conference on Human Factors in Computing Systems (CHI 2015), ACM Press, 2903-2912. DOI: https://doi.org/10.1145/2702123.2702586

[92] Siddiqui, Z., Abdullah, A.H., Khan, M.K., Alghamdi, A.S. (2014). Smart environment as a service: Three factor cloud based user authentication for telecare medical information system. Journal of Medical Systems, 38(1), 1-14. DOI: https://doi.org/10.1007/s10916-013-9997-5

[93] Somasundaram, R., Thirugnanam, M. (2021). Review of security challenges in healthcare internet of things. Journal of Wireless Networks, 27(8), 5503-5509. DOI: https://doi.org/10.1007/s11276-020-02340-0

[94] Spanakis, E.G., Spanakis, M., Karantanas, A., Marias, K. (2016). Secure access to patient's health records using SpeechXRays a mutli-channel biometrics platform for user authentication. International Conference of the Engineering in Medicine and Biology Society (EMBC 2016), IEEE, 2541-2544. DOI: https://www.doi.org/10.1109/EMBC.2016.7591248

[95] Squire, L (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. Journal of Cognitive Neuroscience, 4(3), 232-243. DOI: https://doi.org/10.1162/jocn.1992.4.3.232

[96] Sternberg, R.J. (2003). Cognitive theory. Thomson Wadsworth, Belmont, CA

[97] Stobert, E., Biddle, R. (2013). Memory retrieval and graphical passwords. Symposium on Usable Privacy and Security (SOUPS 2013), USENIX, 1-14. DOI: https://doi.org/10.1145/2501604.2501619

[98] Tan, J., Bauer, L., Christin, N., Cranor, L.F. (2020). Practical recommendations for stronger, more usable passwords combining minimum-strength, minimum-length, and blocklist requirements. ACM Conference on Computer and Communications Security (CCS 2020), ACM Press, 1407-1426.

DOI: https://doi.org/10.1145/3372297.3417882

[99] Tari, F., Ozok, A.A. and Holden, S.H. (2006). A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In Proceedings of the second symposium on Usable privacy and security (pp. 56-66). DOI: https://doi.org/10.1145/1143120.1143128

[100] Thorpe, J., Al-Badawi, M., MacRae, B., Salehi-Abari, A. (2014). The presentation effect on graphical passwords. In proceedings of the SIGCHI conference on human factors in computing systems, 2947-2950. DOI: https://doi.org/10.1145/2556288.2557212

[101] Tran, Q.N., Turnbull, B.P., Hu, J. (2021). Biometrics and privacy-preservation: How do they evolve? IEEE Journal of the Computer Society, 2, 179-191. DOI: https://www.doi.org/10.1109/OJCS.2021.3068385

[102] Tulving, E. (2002). Episodic memory: From mind to brain. Annual Review of Psychology, 53, 1-25

[103] Unar, J.A., Seng, W.C., Abbasi, A. (2014). A review of biometric technology along with trends and prospects. Pattern Recognition 47(8), 2673-2688. DOI: https://doi.org/10.1016/j.patcog.2014.01.016

[104] Wang, J., Katabi, D. (2013). Dude, where's my card?: RFID positioning that works with multipath and non-line of sight. ACM Conference on SIGCOMM (SIGCOMM 2013), ACM Press, 51-62. DOI: https://doi.org/10.1145/2486001.2486029

[105] Wiedenbeck, S., Waters, J., Birget, J.C., Brodskiy, A., Memon, N. (2005). Authentication using graphical passwords: Effects of tolerance and image choice. In Proceedings of the 2005 symposium on Usable privacy and security, 1-12. DOI: https://doi.org/10.1145/1073001.1073002

[106] Williams, H. L., Conway, M. A., Cohen, G. (2008). Autobiographical memory. In Cohen, G., Conway, M.A. (Eds.), Memory in the Real World (3rd ed.), 21-90, Hove, UK: Psychology Press

[107] Winter, B., Grawunder, S. (2012). The phonetic profile of Korean formal and informal speech registers. Journal of Phonetics, 40(6), pp. 808-815. DOI: https://doi.org/10.1016/j.wocn.2012.08.006

[108] Wu, F., Li, X., Sangaiah, A.K., Xu, L., Kumari, S., Wu, L., Shen, J. (2018). A lightweight and robust two-factor authentication scheme for personalized healthcare systems using wireless medical sensor networks. Future Generation Computing Systems, 82, 727-737. DOI: https://doi.org/10.1016/j.future.2017.08.042

[109] Yildirim, M., Mackie, I. (2019). Encouraging users to improve password security and memorability. Journal of Information Security, 18, 741-759. DOI: https://doi.org/10.1007/s10207-019-00429-y

[110] Zeb, K., Saleem, K., Al Muhtadi, J., Thuemmler, C. (2016). U-prove based security framework for mobile device authentication in eHealth networks. International Conference on e-Health Networking, Applications and Services (Healthcom 2016), IEEE, 1-6. DOI: https://doi.org/10.1109/HealthCom.2016.7749518

[111] Zhang, L., Zhang, Y., Tang, S., Luo, H. (2017). Privacy protection for e-health systems by means of dynamic authentication and three-factor key agreement. IEEE Transactions on Indoor Electronics, 65(3), 2795-2805. DOI: https://www.doi.org/10.1109/TIE.2017.2739683

[112] Zhang, X., Qin, Z., Lyu, Y. (2018). Biometric authentication via finger photoplethysmogram. Computer Science and Artificial Intelligence (CSAI 2018), ACM Press, 263-267. DOI: https://doi.org/10.1145/3297156.3297174

[113] Zhao, Z., Ahn, G.J., Hu, H. (2015). Picture gesture authentication: Empirical analysis, automated attacks, and scheme evaluation. ACM Transactions on Information and System Security (TISSEC), 17(4), 1-37. DOI: https://doi.org/10.1145/2701423

[114] Zhao, Z., Ahn, G.J., Seo, J., Hu, H. (2013). On the security of picture gesture authentication. USENIX Security Symposium (USENIX Security 13), 383-398

APPENDIX

Table A1: Research methodology outline.

Timeline	
Phase A ***(6 months)***	**Literature Review and Needs Verification** ***(Section 2 of the paper)*** *Verify and triangulate the state-of-the-art user authentication literature in the healthcare domain with diverse stakeholders of three European healthcare organizations* **Literature Review** - *Sources:* ACM Digital Library, IEEE Xplore - *Keywords:* authentication; password; biometric; locimetric; drawmetric; healthcare; health - *Number of papers reviewed based on inclusion criteria:* 40 - *Publication date:* 01/01/2015-01/06/2021 **Triangulation of Literature with Healthcare Organizations** - Zuyderland Medical Center, Netherlands; ~100K annual patients and users - Hospital Clinic of Barcelona, Spain; ~25K annual patients and users - Western General Hospital, Scotland; ~20K annual patients and users **Procedure** - Semi-structured interviews with key stakeholders (n=9) - *Stakeholder Profiles:* Chief Information Security Officers, Enterprise Architects, IT Department Managers, Security Experts, Doctors, Project Managers
Phase B ***(12 months)***	**DuoPass Design and Development** ***(Section 4 of the paper)*** *Design and development of the DuoPass authentication system based on the "Single-Secret Two Reflections" paradigm, following a user-centered design approach* **Design Considerations** - Security Factors - Usability and User Experience Factors - Adaptation and Personalization Factors - Security and Usability Key Performance Indicators **Key Performance Indicators adopted for the evaluation study** - Password guessability - Password creation efficiency - Memory time - Login time - Users' perceived security, usability, trust and likeability
Phase C ***(11 months)***	**User Evaluation with Participants of Healthcare Organizations** ***(Section 5 and 6 of the paper)*** *Record users' interactions with the suggested DuoPass approach or a state-of-the-art authentication approach, aiming to evaluate its security, memorability and user experience* **Sampling and Procedure** - Between-subjects' feasibility study (n=68); human guessing attack study (n=92) - Experimental group used the DuoPass authentication system, which included a graphical password system with *location-aware images* based on the suggested authentication paradigm - Control group used a state-of-the-art authentication system, including *non-location aware images* based on current state-of-the-art authentication approaches

Algorithm for image recommendation during password creation/reset.

Algorithm #1: Image recommendation during password creation/reset

Input: A set of user models ($um = um_1, um_2, ..., um_m$) that describe the users' frequent visits and important locations at the hospital, filtered to contain relevant information based on the relationship map between the mainstream areas, and a set of candidate images $ci = (ci_1, ci_2, ..., ci_k)$ that depict the mainstream spatial areas of the hospital, provided by the system administrator and the end-users.

Output: The top N images that are recommended to the end-user based on the semantic similarity scores.

```
procedure Mainstream_Map()
    ma = identify_mainstream_areas()
    mm = create_relationship_map(ma)
    return mm
end procedure
procedure Candidate_Images()
    ci_set = upload_images()
    for i := 1 to k do begin
        eit_i = explicit_image_tags() # Annotated by system administrator
        iit_i = implicit_image_tags() # Annotated by computer vision techniques
        ci_i = eit_i ∪ iit_i
        ci_i = clean_text(ci_i)
        append_to_set(ci_set, ci_i)
    end for
    return ci_set
end procedure
procedure Recommend_Images(mm, ci)
    for i := 1 to m do begin
        semantic_ranking = {}
        fv_i = frequent_visits()
        il_i = important_locations()
        fmm_i = filter_mainstream_map(mm, fv_i, il_i)
        um_i = fv_i ∪ il_i ∪ fmm_i
        um_i = clean_text(um_i)
        for j := 1 to k do begin
            ss_ij = semantic_similarity(um_i, ci_j) # through NLP techniques
            semantic_ranking[i][j] = ss_ij
        end for
        sort_by_value(semantic_ranking)
        recommend_top_N(um_i, semantic_ranking)
    end for
end procedure
mm = Mainstream_Map()
ci = Candidate_Images()
Recommend_Images(mm, ci)
```