
Evaluation of a rhythm based user authentication system for mobile devices

Project report submitted in part-
fulfilment for BEng Computer Science
in the Department of Computer
Science, University of York

March 17th 2009

Christopher Northwood

Supervisor: Alistair Edwards

Word Count: 18,747, as calculated by Microsoft Word

Page count: 56

Excluding appendices

1 ABSTRACT

This study evaluates *Tapas*, a system for authenticating to mobile devices using a tap sequence on a touch screen designed by Marriner (2007) and implemented on the Windows Mobile platform by Jolley (2008).

It undertakes 3 experiments: a diary experiment with participants using the system extensively; a usability experiment which asked users to register, and later recall a tap sequence; and a security experiment which investigates whether users overhearing tap sequences can successfully log in to the system.

Results were varied. Memorability of tap sequences appears to be at least as good as a traditional PIN/password system, yet many users found the system frustrating in other aspects. The security of the system is also poor, as all tap sequences were reproduced simply by overhearing them being entered.

In conclusion, further work must be undertaken with the *Tapas* system to address these issues, specifically with regards to security, but the concept of authenticating using a sequence of taps is sound.

2 TABLE OF CONTENTS

1	Abstract	3
2	Table Of Contents	4
3	Table Of Figures	6
4	Introduction	7
5	Project Planning	8
6	Literature Review	9
6.1	Security & Usability	9
6.2	Alternatives To Passwords	10
6.3	Introducing <i>Tapas</i>	12
7	Experiment Design	17
7.1	Planning.....	17
7.2	Ethical Considerations	19
7.3	Long-Term Experiment.....	19
7.4	Usability Experiment	21
7.5	Security Experiment	23
8	Preparation	25
8.1	Long-Term Experiment.....	25
8.2	Usability Experiment	26
9	Results & Analysis	27
9.1	Long-term Experiment	27
9.2	Security Experiment	33
9.3	Usability Experiment	38
10	Conclusion.....	51
10.1	How secure is <i>Tapas</i> ?	51
10.2	How memorable are tap sequences?	52
10.3	How does <i>Tapas</i> compare to the traditional PIN or password based authentication mechanisms?.....	52
10.4	Potential Future Work.....	52
11	Bibliography	54
Appendix A	Disclaimer	57
Appendix B	Extended Use Experiment Participant Instructions	58
B 1	Software Disclaimer	58
B 2	Part One.....	58
B 3	Part Two	59
B 4	Part Three.....	62

Appendix C	Usability Experiment Participant Instructions	64
C 1	Part One	64
C 2	Part Two	65
Appendix D	Usability Control Experiment Participant Instructions	67
D 1	Part One	67
D 2	Part Two	68
Appendix E	Security Experiment Participant Instructions.....	70
E 1	Setting Up The Experiment	70
E 2	The Experiment.....	70

3 TABLE OF FIGURES

Figure 1 - Tapas Unlock Screen	25
Figure 2 - Experiment Control Screen	26
Figure 3 - Results from password questionnaire	29
Figure 4 - Results from Tapas Questionnaire.....	30
Figure 5 - Results from Security Experiment	35
Figure 6 - How Taps Converted into the rhythmic contour	37
Figure 7 - Age of Participants	38
Figure 8 - Users with Touchscreen Experience	39
Figure 9 - Users who use a PIN on their device	39
Figure 10 - Users Who play or Had Previously Played an instrument	40
Figure 11 - Number of Attempts for a Successful Tap Sequence Capture	41
Figure 12 - Capture Frustration.....	41
Figure 13 - Capture Attempts vs. Frustration	42
Figure 14 - Perceived Complexity of Tap Sequence.....	42
Figure 15 - Number of Taps In Sequence	42
Figure 16 - Attempts vs. Complexity	44
Figure 17 - Attempts vs. Experience	44
Figure 18 - Post Capture Confidence	44
Figure 19 - Confidence Prior to Recall.....	45
Figure 20 - Changes in Confidence.....	45
Figure 21 - Recall Attempts Before Successful Login	46
Figure 22 - Mental Demand of remembering sequence	46
Figure 23 - Effort to Recall and enter Tap Sequence	46
Figure 24 - Frustration of Recalling Tap Sequence	47
Figure 25 - Login Attempts By Musical Ability	48
Figure 26 - Confidence prior to login vs. Attempts required to log in.....	48
Figure 27 - Preference.....	48
Figure 28 - Perceptions on Security	49

4 INTRODUCTION

Tapas was proposed by Edwards (2005) as a system to authenticate users on mobile devices by tapping a rhythm onto a touch screen mobile device, with the aim of increasing the usability of the security systems on such devices.

An initial investigation was done by Marriner (2007) into the problem, resulting in an algorithm for comparing match patterns and a small-scale feasibility study, proving positive. This work was then taken by Jolley (2008) and turned into an implementation for the Windows Mobile platform. More information on these works can be found in section 6.3.

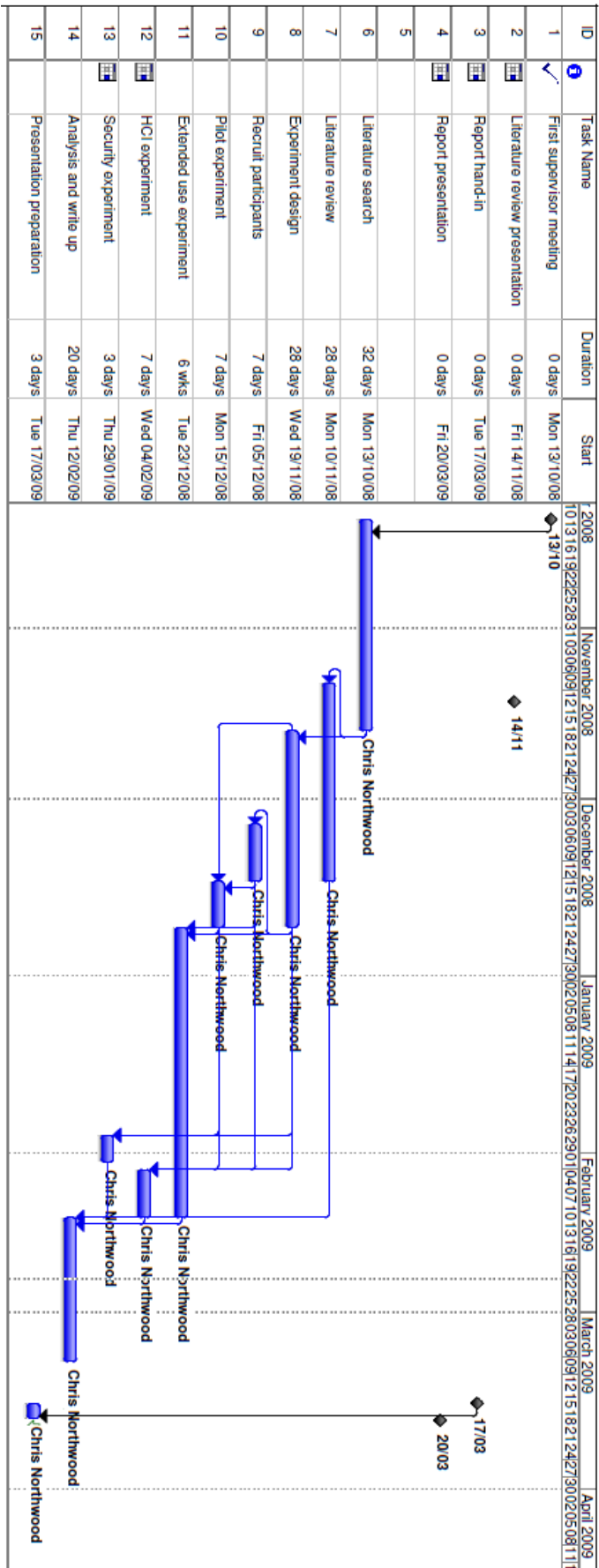
This study uses the implementation by Jolley (2008) to perform a wider evaluation with larger scale experiments in order to answer key questions.

In the project definition, “*a need for larger-scale experiments to establish the security and practicality of the approach*” is given (Edwards, 2008), with a project outline of the general type of experiment expected to be performed. Using the scope given in this project definition, by the end of the study I aim to have answered the following broad questions:

1. How secure is *Tapas*?
2. How memorable are tap sequences?
3. How does *Tapas* compare to the traditional PIN or password based authentication mechanisms?

Techniques for performing experiments in human-computer interaction are well developed and documented and can be used to answer these questions. More detail on the specific approach taken is given in section 7.

5 PROJECT PLANNING



6 LITERATURE REVIEW

6.1 SECURITY & USABILITY

“An underlying goal has been to provide [...] security at minimal inconvenience to the users of the system” (Morris & Thompson, 1979).

This quote sums up the essence of usability concerns in computer security concisely. This requirement has been known for some time – Kerckhoff wrote his 6th principle in 1883 which dealt with the usability of military cryptographic ciphers (Petitcolas, 2008). Some suggest that this principle is *“more easily forgotten than understood”* (Gutmann & Grigg, 2005), with some secure system designers going out of their way to purposely annoy the user (Espiner, 2008).

Traditionally, passwords have been the method used for a shared secret between the user and a device that allows for user authentication; however this method has many known problems, such as users forgetting passwords, or writing them on pieces of paper hence compromising the security of such a system (Sasse, Brostoff & Weirich, 2001).

Sasse, Brostoff & Weirich (2001) suggests that the problem with computer security is the users – the so called “weakest link”, and that with this point of view, the typical method of increasing security by using increasingly sophisticated and complex technological means does not work, as these do not address the core problem. The example of Whitten & Tygar (1999) is given, where users were given 90 minutes to sign and encrypt an e-mail message using PGP5 – but most users could not accomplish this task, despite PGP5 having a good user interface using typical HCI analysis methods.

Sasse, Brostoff & Weirich (2001) performed their study based chiefly on passwords, but also included a voicemail system that required a 6 digit PIN. Although only one study on the PIN was conducted, it showed recall problems with the PIN to be worse than with the passwords. Both passwords and PINs are common logon mechanisms for PDAs and smart phones.

Sasse, Brostoff & Weirich (2001) came to the conclusion that although passwords are not fundamentally flawed, common implementations are, and several methods were suggested to alleviate these problems, such as single sign-on systems to reduce the memory load, but this has little relevance to mobile devices. A common thread in these conclusions is reducing memory load, and this study aims to investigate the memory load of a sequence of taps.

With regards to users’ views to security on mobile devices, Harrington & Mayhew (2001) suggest that many users are reluctant to use PINs, and that most mobile phone implementations of PIN security are poor, in that they only prompt for the PIN on start up, whereas most phones are kept on constantly (Clarke, Furnell, Rodwell, & Reynolds, 2002).

This particular research is possibly of limited use, as the intended audience of *Tapas* is chiefly business people, who are likely to store more sensitive (both commercially and legally) data on their mobile devices than a typical consumer. Additionally, this business person is also likely to have policies mandated by their employer, meaning they have little choice in whether to use the security or not. However, the participants in this study are likely to include students, typical consumers, so it is important to bear this research in mind when analyzing their views on the *Tapas* system.

Studies have also been done on business users of PDAs (Leyden, 2003), which found that two-thirds of business users protect their PDA, but this still means one-third do not, despite having a large amount of sensitive data on their mobile device. This news article is now 5 years old, and a study only a year later (The Register, 2004) shows 91% of corporate respondents citing security as their top priority. With data loss moving increasingly into the public eye both for large corporations (Oates, 2008; BBC, 2008) and for smaller organisations (Sanchez, 2008) in addition to pressure from the Government (Information Commissioner's Office, 2008), these statistics may be very much out of date.

6.2 ALTERNATIVES TO PASSWORDS

Elftmann (2006) conducted a thorough investigation into password alternatives, and identifies the problem of passwords as being a hangover from a command-line driven age which was not updated when GUIs became common. It is fairly easy to see how the same argument could be applied as computing transitioned to the mobile world, where devices do not necessarily have keypads and input is through a touch screen.

Passwords and PINs are classed as a type of security scheme called secret, or knowledge, based, however approaches such as biometric-based schemes, image-based schemes and token-based schemes can also be considered.

Biometric security generally consists of identifying yourself by providing a specific characteristic to the system requiring identification. However, biometrics suffers from multiple problems, specifically that of reading the biometric characteristic being measured. This is traditionally solved by using a peripheral device intended for reading a specific characteristic, such as a fingerprint reader or an iris scanner, but having to carry an additional peripheral for a mobile device around with you is unwelcoming, and, with the exception of some notebook computers, these readers are not integrated into the mobile device, making them not highly usable.

Biometrics also is plagued by trust issues; Eschenburg, Lylykangas, Krämer, Surakka, Troitzsch, Vuorinen & Bente (2005) show that just over 40% of people will accept using biometrics to log in to their PC or check their e-mail, and presumably by extension, log in to their mobile device.

Keystroke analysis is a form of biometrics which looks at how a user entered a password in order to increase the security of the password – even if you discover someone's password,

you also have to enter it in a sufficiently similar way to the owner in order for it to be accepted. This is traditionally done on a hard keyboard, but Saevanee & Bhatarakosol (2008) have investigated identifying keystroke dynamics on a touch screen. Combined with a relatively high user acceptance rate for keystroke dynamics, this may prove to be a promising area for increasing password security, however nothing is done to deal with the memory problem, as it just adds an additional layer of security, instead of replacing the existing one.

A token-based system, similar to keystroke analysis, is traditionally used to complement password based authentication, rather than replace them. In the most popular token-based implementations, such as RSA's SecurID (RSA Security, n.d.), the user is supplied with a physical token which shows a PIN on the screen generated by a cryptographic function dependent on time, which is then given, along with a password, to a login screen. Such a system is typically used for authentication into remote systems, such as a VPN, rather than local systems.

Most research for token-based systems with mobile devices appear to be focussed on using the device itself as a token (a so-called soft token), rather than using a token for logging in to the device itself. Although token-based authentication does appear to provide additional protection, it is important to bear in mind that two different aspects of security are already being covered: "something you know" and "something you have", by possession of the device. If an attacker already has the PDA, there's a good chance they have the token (especially if the two are kept together, which is likely), negating the additional security added.

Although token-based systems do add additional security, they also add additional inconvenience (having to check the token and enter an additional PIN), and it is not clear whether this trade-off is justified without context.

There are also a number of schemes available that are a slot in replacement for passwords, including some specifically developed for mobile devices. They generally work on the same principle of recall, but use images or graphical based systems in order to do it, as human brains are typically better at remembering images than strong passwords (Elftmann, 2006).

Elftmann (2006) breaks down these graphical systems into three classes: *pure recall*, *cued recall* and *recognition*.

Pure recall requires the user to directly draw their pattern onto the screen without any hints to what their pattern may be. The Draw-A-Secret system was an early implementation of this, which processed a pattern drawn onto a grid to match against a stored pattern. The T-Mobile G1 smart phone, based on the Android platform, contains a system that requires the user to recall a pattern for connecting dots which falls into this category.

Cued recall works by showing users an image and asking them to click on specific locations or on specific objects in the image in a specific order to unlock the device. A system called

Passlogix was an early implementation of this, which defined ranges around objects in an image, so clicking anywhere on the image would work. An improved variant called PassPoints was later implemented, which works by remembering the location in the image that was clicked on, rather than objects. This also means any image (such as one with personal meaning to the user) can be used and the password space is larger, as anywhere in the image can be used, not just defined areas.

The final class, recognition, works by showing the user some images during registration, and then asking the user to pick one of the images they selected out of a group containing that image and other random ones. Again, this takes advantage of the brain being a lot better at remembering images than text. Two major implementations include Déjà Vu (Dhamija & Perrig, 2000), which works with random, abstract images, and Passfaces (Passfaces, 2009), which uses random faces for recognition.

Pure recall is the most difficult memory task, with cued recall being less so and recognition as the easiest of the three. There are many studies that come to the conclusion that graphical passwords (even those in the pure recall class) are believed to be more memorable than textual passwords, but they tend to take longer to register and log in to the system than textual passwords (Elftmann, 2006; Dhamija & Perrig, 2000).

Research into the security of graphical passwords compared to textual passwords is limited to small scale studies, as there are very few major deployments of graphical passwords. Some commentators (Suo, Zhu & Owen, 2005) believe them to be resistant to many of the attacks passwords suffer from, studies on Passfaces (Davis, Monroe & Reiter, 2004) have found that many users choose faces highly correlated to their own race or gender, hence greatly reducing the security of the system.

It would appear that alternative password schemes, specifically graphical passwords, are a promising area showing improvements over the traditional password mechanism, however research and large scale studies on the same scale as those done on traditional passwords are still needed to conclusively show that switching from passwords is possible. Additionally, there is still no conclusive research showing which password alternative is the best solution, as even within the field of graphical passwords there are a multitude of different systems.

6.3 INTRODUCING *TAPAS*

Tapas was proposed by Edwards (2005) as a method to increase usability of security systems on mobile devices. An implementation of *Tapas* (called *Tap-Pass* at the time) was proposed as a system to unlock a PDA by tapping a secret, personal rhythm on the screen. It was hypothesised that remembering and entering this series of taps is more usable than that of a PIN or password, as well as fulfilling specific security criteria:

- *The thief is aware of the tapping mechanism, has observed the owner using it and attempts to imitate it. The hope is that rhythms will be sufficiently difficult to forge. This will be the subject of this project.*

- *The thief is aware of the mechanism, but has not observed the owner's rhythm. He or she therefore tries to guess the rhythm - and will not succeed.*
- *The thief is unaware of the mechanism. Attempts to enter text, shake the device and so on are fruitless. (Edwards, 2005).*

It is easy to see that this shares some principles with the pure recall graphical scheme discussed above – the user is asked to enter a tap sequence without any cues. The assumption is that tap sequences, like images, are more memorable than passwords.

With this in mind, Marriner (2007) undertook an investigation into the concept by developing some prototypes and performing a series of small scale experiments in order to check the feasibility of the concept.

The first experiment undertaken by Marriner (2007) was one to capture a series of taps in order to perform analysis on the tap sequences used by participants. As a result of this analysis, Marriner (2007) found that the mean tap sequence length was approximately 13 taps. For clarity, I will use taps to mean a full tap, that is placing the stylus on the screen and then lifting it up again, and event to refer to a single up or down event. Marriner (2007) is not clear whether he means tap or event in his original report, however analysis of the raw data gives a mean of 13 events and 7 taps for his small scale sample, so it is likely he meant events.

At this point in the experiment Marriner (2007) had not developed a matching algorithm, so participants were not asked to enter their patterns twice to confirm; therefore it is unknown if the patterns captured were accurately captured, or accurate representations of what the user wanted to capture. Furthermore, Marriner (2007) detected multiple occasions where taps were detected with length 0 or 1, which were discarded. It is unknown what caused these anomalies and whether they had any effect on the recorded data.

In this experiment, Marriner (2007) also attempted to address the problem of security, however the usefulness of the results he gathered are questionable as the matching algorithm used is different to that in the final product.

Finally, users were asked to re-enter their taps and the matching was considered in the same way as in the security experiment. 46% of the re-entered patterns were a different length to the originally captured pattern - this is a significant amount, as the final matching algorithm used by Marriner (2007) immediately discards patterns of different lengths.

Using the analysis from this, Marriner (2007) recommends an algorithm based on a technique proposed by Peters, Anthony & Schwartz (2005) to identify song titles based on tapping a rhythm on a keyboard.

This technique considers relative frequency of taps in a sequence by converting the time between events into a character representing if that particular tap was tapped at a different

rate to the previous tap. Given a tap time n and the two tap times immediately before it: n_{-1} and n_{-2} :

1. If $n - n_{-1} > n_{-1} - n_{-2}$, that is, the time between this tap and the one before it is greater than the time between the two previous taps, then this gap is represented as 'u', that is, the values go *up*.
2. If $n - n_{-1} < n_{-1} - n_{-2}$, that is, the time between this tap and the one before it is less than the time between the two previous taps, then this gap is represented as 'd', that is, the values go *down*.
3. If $n - n_{-1} = n_{-1} - n_{-2}$, that is, the time between this tap and the one before it is the same as the time between the two previous taps, then this gap is represented as 's', that is, the values stay the *same*.

The actual algorithm is slightly more complex than this, by first converting the timings to a rhythmic contour, and by also introducing a small amount of allowed fluctuation between the two values in the third case. Marriner (2007) experimentally determined a value of 1.5 ms to allow for two tap distances to be considered the same.

Using this algorithm, Marriner (2007) undertook further tests with the system, asking users to enter a series of taps and then to repeat it. This gives more useful results than the previous experiment, showing that it takes on average 3.6 times for a user to successfully reproduce two tap sequences. This is a very high amount, although further research is required to see if this would decrease as users become more used to the pattern. From the outset though, it appears as if enrolment may be a chore and cause negative impressions in the user's mind of *Tapas* of being unreliable from first use. Assuming that the 3.6 times to successfully reproduce a tap assuming no memory issues is a constant and does not decrease, and assuming a 6 second time to enter a tap sequence (3 seconds for the actual sequence + the tap timeout), this leads to a seemingly high 21.6s wait on average to unlock a PDA. The biggest problem identified was the detection of different number of taps by the touch screen than entered.

Experiment 3 by Marriner (2007) gives different results, of those who could successfully log in, a mean of 1.6 attempts were needed, which is more acceptable, however no reason for this disparity is given. One fifth of people attempting to log in could not, and gave up after 10 events. This amount is unacceptably high.

The security of the algorithm can also be considered. Given that a tap event can be encoded in 3 different ways, and each tap generates 2 events (an up and a down), then for a tap sequence of length n , the number of event pairs is $2n - 1$, so the possible permutations of the generated string is 3^{2n-1} , if the length of the taps is known. Given our average rhythmic contour string is of length 13, this gives us 531,441 possible permutations, which appears to compare favourably to a 4 digit PIN which only has 10,000 permutations. It remains to be seen whether users generate sequences with sufficient entropy to maximise this theoretical space.

Further to this, Jolley (2008) took the concept and prototypes developed by Marriner (2007) to implement the *Tapas* into a security subsystem for a Windows Mobile 5 device. It is this implementation that the experiments carried out by this study focus on.

Jolley (2008) lists several requirements and features for his implementation:

For the Set up Screen:

- *The user must be able to enter a Tapas pattern.*
- *This data must be saved onto the device.*
- *The user must be able to set a lock out time.*
- *The user must be able to set a backup password.*
- *The user must be able to change the Tapas pattern once an original pattern has been entered.*

For the Lock Screen:

- *The user must be able to enter a Tapas pattern.*
- *The user must be able to enter a backup password.*

To fulfil these requirements, Jolley (2008) investigated two methods of implementing *Tapas* onto a Windows Mobile device, firstly as a Local Authentication Plug-in (LAP) for the Local Authentication Subsystem (LASS), an integral part of Windows Mobile OS, or secondly as a persistent background application implementing its own security lockout system.

Given the downsides of implementing a persistent application (hit against memory usage, conflicts with the LASS, ease of disabling *Tapas*), and that the LASS is the official supported mechanism for this type of implementation, Jolley (2008) implemented *Tapas* as a LAP for the LASS.

Jolley (2008) also made the decision to re-implement the system for pattern capture and matching developed by Marriner (2007), mainly as LAPs must be written as a DLL and Visual Basic does not provide facilities for this kind of binding. Jolley (2008) also lists a number of usability concerns with the Marriner (2007) prototypes.

Jolley (2008) also conducted some small-scale usability tests of his particular software with 11 users; however despite some small bugs identified, the test results were positive. However, there are some inconsistencies in the analysis of the results, for example, Jolley (2008) claims “it was positive that all the users found the time out on the device to be working”, however the results show that the timeout lock did not work in 2 cases and no explanation is given for the discrepancy.

Additionally, the results of the tests are at odds with the small usability studies Marriner (2007) performed, which showed significant issues with being able to reproduce taps and the ability of the system to detect taps that were not evident in this experiment. Statistics such as how many attempts it took a user to successfully log in using *Tapas* would have been

useful. It is possible that the small scale of the experiments in Jolley (2008) meant the problem only occurred once, or it is possible that the device Jolley (2008) used, an O₂ XDA Mini S, uses a different touch screen technology to the LOOX 720 used by Marriner (2007). Unfortunately neither devices list in their datasheets any specific information about this touch screen so I would not be able to investigate these possibilities any further.

A further review of the specific software developed by Jolley (2008) was taken as part of the experiment preparation and is discussed in section 8.1.

7 EXPERIMENT DESIGN

7.1 PLANNING

As discussed in the Introduction, this study aims to address three key questions drawn from the scope of the project as defined by Edwards (2008):

1. How secure is *Tapas*?
2. How memorable are tap sequences?
3. How does *Tapas* compare to the traditional PIN or password based authentication mechanisms?

Experiments should be designed in order to answer these questions. Cairns & Cox (2008) cover the different form of human-computer interaction experiments that can be undertaken, specifically covering questionnaires, in-depth interviews and focus groups. Diary studies are also a common tool used in HCI so were also included for consideration into the designs of the experiments.

When designing the experiments it is also important to consider the small scale experiments done by Marriner (2007) and Jolley (2008) in order not to simply rehash their work. A final consideration is the scope of the project as defined by Edwards (2008) which dictates that the experiments should be on a larger scale than those done previously.

The experiments performed by Marriner (2007) and Jolley (2008) were limited in depth, and focussed mainly on quantitative data using questionnaires. The scale was also limited in breadth, dealing with a limited number of participants. This study should expand on both the breadth and depth of these previous experiments. A diary study would collect a great deal of qualitative data greatly increasing the depth, however a concern exists on whether or not a sufficiently broad group would take part – Windows Mobile has a limited market share at 13.8% (Canalys, 2008). A workaround for this would be to provide the devices for people to use, however resource limitations mean this is not possible.

An alternative is to perform a questionnaire based study, using a single device and inviting participants back regularly at differing intervals in order to see if they can recall and enter their tap sequence. This would allow many more participants to be involved, but is a more contrived situation, missing out on the real world usage a diary study would allow.

The focus group and interview scenarios discussed in Cairns & Cox (2008) could also be considered, however suffer their own problems. With the limited resources I had at my disposal (only a single device), a focus group would offer limited benefit over interviews or questionnaires as users could only partake one at a time, and participants may be biased by the other participants in the focus group. Interviews are also problematic with the great deal of qualitative data obtained which is difficult and time-consuming to analyse.

It would appear that designing a single experiment which is both sufficient in depth and breadth is complex, and a better solution may be to design two experiments, one which investigates the subject in depth, and a second experiment which is on a large enough scale to have a large breadth.

Revisiting the earlier diary study idea, this now appears to be a good solution, as a second experiment can be designed which addresses the breadth element that the diary study may be lacking in.

The diary study should aim to emulate as realistic a simulation as possible so the instructions will be kept simple and ambiguous in order to gather as much data as possible. The limited number of participants will make being able to analyse a large amount of qualitative data a possible task within the project.

The complementary experiment must therefore widen the participation, yet bring a manageable amount of results. Questions that bring quantitative answers are desirable here, although a limited amount of qualitative data can be corrected. Questionnaires about hands-on experience with the device, as opposed to interviews and focus groups, seem a good solution; however the nature of the questionnaire and the experiments needs to be considered. In order to investigate the memorability of taps, the format suggested by Edwards (2008) makes the most sense, with users being asked to register a tap sequence and then recall it later.

The time between registration and recall also needs to be considered. It should be long enough to be a real test of memory, yet short enough to be realistic. As a result of my literature review, there appears to be limited information about the memorability of this sort of sequences, so the time chosen is fairly arbitrary.

With the in-depth and questionnaire based studies, a study sufficiently broad and deep enough to deal with the question of memorability appears to be devised, however there remains the points of comparison to passwords and security to remain.

The questionnaire could be designed such as to measure users perceived security; however a real test of security would need either a more involved experiment, or another separate one. The diary experiment may also touch on security; however this would be a fairly limited amount depending on how many attempts are made to unlock the device. A more definite study is required in order to ensure there is sufficient data to be analysed.

A third experiment, with quantitative data to be collected experimentally, appears to be sensible. In order to keep the diary study a real world experience, attacks on devices should be simulated outside of the diary study. The data generated from this experiment should be quantitative in order to perform a detailed analysis on the system, and qualitative data gathering on perceived security can be incorporated into the other experiments. For the same reasons as discussed above, an experiment with the data captured by questionnaire appears to be the most appropriate.

The final concern is how to compare the system to the traditional PIN/password system; however this is simply accomplished by following good experimental design and using a control group with PINs, rather than taps.

Considering the data to be collected, and the relative strengths of each different method considered above, there appears to be three experiments that need to be designed to address the different aspects of the problem:

- A diary study in the real world in order to gather in depth data on usability and memorability;
- A questionnaire and registration/recall experiment in order to gather a broad range of experiences with *Tapas*;
- A further questionnaire and security based experiment in order to address the concerns regarding the security of *Tapas*.

7.2 ETHICAL CONSIDERATIONS

Blandford, Adams, Attfield, Buchanan & Gow (2008) suggests three important elements of ethical consideration:

- Vulnerable participants;
- Informed consent;
- Privacy, confidentiality and maintaining trust.

The participants were all able bodied and in good health, with the majority being students, with some professionals taking part – meaning that they would not feel pressured into taking part in the experiment and getting good results.

All participants were given a high level overview of the experiment in the advertisement for participants, and a more in-depth description when they responded to the advertisement and given the participant pack, as well as the full pack of instructions, so they were fully informed at all points about what the experiment entailed.

All participants were also fully informed that their data will be held confidentially for analysis, and only data that has been made anonymous will be made available in the report. The paper questionnaires and diaries were securely stored in my house with the raw computerised data and analysis in a password protected spreadsheet. Only the analysed and summarised anonymous data will appear in this final report or its' appendices.

These three facts address the three elements of ethics listed above, in addition to satisfying the ethical expectations of computer science professionals (British Computing Society, 2006).

7.3 LONG-TERM EXPERIMENT

7.3.1 OVERVIEW

For the long-term experiment, the aim was to investigate the usability of the *Tapas* system in everyday use, which would not be covered by the laboratory experiments. This allowed

Tapas to be used a lot more heavily than in the laboratory experiments, giving a lot of useful data about the individual's use of *Tapas*.

The long-term experiment was also useful for gathering usability data on *Tapas* in a variety of different environments. Jolley (2008) suggested the use case of logging into a device on a bus, and it is obvious how other environments could be encountered.

Due to resource limitations, participants were required to have a Windows Mobile device or smart-phone, this limits the sample size that were available to partake in this experiment; however this allowed analysis to focus more on qualitative data. As these participants already had experience with a Windows Mobile device, they were their own control group, and meaning that the experiences of using the mobile device both with and without *Tapas* could be related.

A possible consequence of the requirement for people to provide their own devices is that this limits the range of people involved and the sample may have been biased.

7.3.2 PARTICIPANT INSTRUCTIONS

For this experiment, participants were first asked to complete a background questionnaire and enable PIN/password protection for their Windows Mobile device if they did not already use it.

The questions regarding background were required in order to gauge the users' experiences with the software and place the diaries into a context. The control period at the beginning of the experiment can be used in order to address the area of comparisons between *Tapas* and PIN/password.

After 3 weeks, the participants were then asked to complete a questionnaire about their experiences with the PIN/password system, and were then given *Tapas* to install on their mobile devices for a further period of 3 weeks.

During these 3 weeks, they were asked to keep a diary recording how they found the system on each day – frustrating or transparent to use, as well as examples of particular situations where they used the system and found it particularly frustrating or had to use the backup password system.

These instructions were kept deliberately vague and broad in order to gain as much information as possible. However, consideration was made in order to ensure the instructions are not so vague so the user does not miss out crucial information. As Jolley (2008) suggested further work using different environments, specific mention was made to recording this type of information.

At the end of the three weeks, another questionnaire was completed, with the users asked to summarise their experiences using the PIN/password based system.

Both this questionnaire and the PIN/password one contained the same questions in order for direct comparisons to be made between the two. These questions were chosen in order to gather the range of perceived experience from the different aspects of the system – the effort of logging in, the mental demand of recall, and the overall frustration of the system. Additional space was also left for freeform qualitative text. The *Tapas* questionnaire also contained an additional question. Following the concerns of Marriner (2007) that the system was not accurately detecting all taps, a question was asked as to what the participants felt the reason any unsuccessful logins were in order to discover if this issue did re-occur.

The full details of the instructions and questionnaires are given in Appendix B.

7.3.3 PILOT EXPERIMENT

Due to the length of time this experiment would take, a full pilot experiment was not feasible, so a different approach was taken, mainly focussing on bug finding and the stability of *Tapas*.

Tapas was installed on a PDA for testing and used daily for 2 weeks during the preparation phase, where bugs were identified and fixed, and the experiment designed based on experiences of using this device. Due to the timing (*Tapas* being tested whilst the experiment was being designed), the actual questionnaires and diary experiment were not piloted, although as diary studies are well understood and the questionnaire similar to those in the other experiments which were piloted, this should present a minimal risk to the project.

7.4 USABILITY EXPERIMENT

7.4.1 OVERVIEW

The usability experiment is the complement to the long-term experiment. Where the long-term experiment aimed to investigate the use of *Tapas* in depth over a long period of time with a limited number of participants, the usability experiment aimed to cover a broader group of users, but with a less in-depth study of the technology.

This experiment was conducted in a laboratory setting and a variety of users chosen to represent a good range of experience with Windows Mobile devices and security, and technology in general.

With the experiment, people were invited to enter a series of taps, with a smaller control group asked to enter a PIN, and then invited back a week later to re-enter their taps or PIN. Quantitative data about the number of users who could recall their taps or PIN, the amount of attempts it took them, and the overall time it took them to log in to the system will be recorded and subsequently used for analysis.

7.4.2 PARTICIPANT INSTRUCTIONS

First, participants were asked to fill in a very brief questionnaire asking for background information about past experience with touch screen devices and PINs/passwords.

The questions asked in the questionnaire are designed so as they can be used to discover interesting correlations with other results. One question that can be addressed is whether or not touch screen experience gives any bearing to the usability of the system, and other useful background information about the participants experience with security may also be interesting in order to strengthen the claims of Clarke et al. (2002) with regards to the current state of mobile security.

Additionally, Jolley (2008) brought up an interesting question of discovering if there was a relationship between musical ability and the ability to remember taps. A question asked people for their musical background was added in order to address this and use this for a correlation.

After filling in this questionnaire, users were then presented with the modified *Tapas* program, as described in section 8.2, which was prepared with their participant ID. They were then asked to enrol on the system by entering their series of taps twice. It was noted how many times they attempted to enrol, characteristics of their tap sequence, and they were asked how confident they felt that they would be able to recall their taps in a week. The characteristics of the tap sequence can then be analysed to see if it bears any correlation to the ability to register and recall the tap sequence, and confidence once again can be used to analyse perceptions of the system.

Participants that were able to successfully enrol were asked to return in a week for the second stage of the experiment.

A control experiment was also designed to be used in the ratio of 1:9 of control to *Tapas*, which instead of using the taps were asked to use the backup password feature of *Tapas* to register, remember and recall a password, rather than a tap sequence. All questions were the same, with appropriate language changed to PIN/password rather than *Tapas*, in order for direct correlations to be drawn.

After the participants returned for the second part, they were first asked to record how confident they felt. With this data, a measure of change in confidence over time can be generated. It also would make sense that a high confidence immediately before recall shows that the user believes they have remembered the sequence, and then correlating this with the attempts made will show how accurate this confidence is.

Users were then presented with the modified *Tapas* program loaded with their user ID and were then asked to attempt to re-enter their taps. The overall time taken to log in, and the number of attempts required to log in was then noted.

Users that were unable to log in were asked why they believed they were unable to log in – either because they did not recall their taps, or because they believed that they could recall them but the system “did not accept” them.

After the experiment had concluded, users were given another very short questionnaire about their experiences with *Tapas* and how it compared against a PIN/password system.

This questionnaire deals with the same criteria as the long-term experiment to address different aspects of the system – the effort of logging in, the mental demand of recall, and the overall frustration of the system, and the penultimate question addresses the comparison question, asking the participants which system they would prefer and asking for qualitative justification for this decision. In order to address the security, or at least the perceived value of it, a final question was added to address this area.

The full instructions are given in Appendix C and for the control experiment in Appendix D.

7.4.3 PILOT EXPERIMENT

A small pilot experiment was performed with 2 users, and the pilot ran smoothly. One participant mentioned the Likert scales were not clear, so minor changes were made to improve the clarity of the scales.

7.5 SECURITY EXPERIMENT

7.5.1 OVERVIEW

Whereas the long-term and usability experiments focus mainly on the usability of the *Tapas* system, this focused mainly on the security of the system. This experiment had a user, Alice, and someone who is attempting to gain access to the PDA, Mallory (Rivest, Shamir, & Adleman, 1978). Using an increasing amount of knowledge and different techniques, Mallory attempted to log in to the PDA using the *Tapas* system.

The first method to be used was that of dumb guessing – would using a randomly different series of taps result in a combination that matches closely enough the user's real password, or using a well known tune or rhythm (such as the *Match of the Day* tune) result in a match?

The second was that where Mallory overheard the taps being entered, firstly in a typical outdoor environment where the more realistic issue of background noise was encountered, and secondly in a less realistic laboratory environment. Mallory will then try to replicate the overheard taps and the success rate will be measured.

The third experiment was where Mallory had both a visual and auditory clue to the tap sequence – that is, they could see Alice enter the taps.

A final experiment would attempt to replicate that of a social engineering attack, where Alice attempted to describe the series of taps to Mallory, and Mallory would then attempt to replicate the taps as described.

7.5.2 PARTICIPANT INSTRUCTIONS

In this experiment, participants will be put into pairs, with one person in the pair representing the 'Alice' user, and the other representing the 'Mallory' attacker.

'Alice' was first asked to enrol into the *Tapas* system away from 'Mallory'. 'Alice' was given guidance as to the complexity of the series of taps to use, so different pairs could simulate different levels of tap complexity. 'Alice' would then log into the system and hand over the

PDA to 'Mallory' (a less violent way of simulating theft) who would attempt to log into the PDA using the information gleaned from 'Alice'. The level of information 'Mallory' had gradually increased until they are able to log into the system and the point that they will be able to log in will be recorded.

With the usability aspects of *Tapas* well covered by the long-term study and the usability study, this experiment was designed just to focus on security, with the only data collected related to how many attempts a break in required. The iterations are design with real-world scenarios in mind which are loosely given in the descriptions, e.g., "lose your PDA on a train", becoming decreasingly realistic as they progress.

Direct comparison between this experiment and a control one involving passwords would not be possible, as a completely different experiment would need to be designed to evaluate password security of which the results may not be directly comparable – overhearing someone enter a password is not a fair comparison. With the wide range of literature relating to password security and the comparisons available from the previous 2 experiments, a direct control is not necessary.

The iterations are as described above, and the full instructions are given in Appendix E.

7.5.3 PILOT EXPERIMENT

A pilot experiment was carried out using a pair of users given the guidance to use a moderately complex sequence – the attacker had no previous experience with *Tapas*, whereas the user was involved in the previous pilot experiment.

The pilot experiment ran smoothly, however the attacker was able to break in surprisingly easily. Considering the problems discussed in 9.2.2, I suspect that it was only through luck and previous experience that the user was able to successfully enrol.

8 PREPARATION

8.1 LONG-TERM EXPERIMENT

To prepare for this experiment, the *Tapas* software implemented by Jolley (2008) will be required to be loaded on to the PDA or smart-phone of the end user.

Testing was performed as part of the pilot study for the long-term experiment, and various bugs were found in the Jolley (2008) implementation, mainly relating to the wrapper code rather than the core *Tapas* algorithm itself. The LAP as compiled by Jolley (2008) would only run on Windows Mobile 5 for Smartphone devices, whereas most modern devices will be running the more modern Windows Mobile 6 OS. Binaries for Windows Mobile 5 for Smartphone would run on a Windows Mobile 6 Professional device, but would lock up a Windows Mobile 6 Classic device on boot requiring a reinstall of the OS.

The cause of this lockup was isolated to the `MakeEmergencyPhoneCall()` procedure, which required the LAP to be linked against a 'phone.dll' library, which Windows Mobile 6 Classic devices do not have. Following the advice of Limosani (2008), this function was changed to check if the 'phone.dll' library is available, and conditionally make a phone call only if this library is available (otherwise do nothing).

A problem was identified due to a change of Windows Mobile's security API between versions 5 and 6, which led to the backup password feature not working. A backup password feature had to be implemented within *Tapas* itself, rather than relying on the external API.

The implementation by Jolley (2008) also had some debugging code left in, such as the 'Emergency Call' feature being diverted to his personal mobile phone, rather than the emergency services, and writing to a file "Debug.txt" in the root of the drive. This debugging code was made to be conditional depending on whether the code was compiled in debug mode or not.

There was also an issue with the implementation of the algorithm where the first tap in a pattern would be disregarded. Jolley (2008) explicitly states that this should not be the case, so I have corrected the comparison loop.

Other design issues were also identified – for example the software killed all other applications if it was not running as the foreground window to be displayed at the front,



FIGURE 1 - TAPAS UNLOCK SCREEN

which sometimes led to the 'Today screen' being killed, rendering the device unusable. This was changed by creating the dialog with a flag stating it should be above all else.

8.2 USABILITY EXPERIMENT

For this experiment, the *Tapas* software will have to be modified slightly in order to deal with the different participants. As the Jolley (2008) implementation of *Tapas* stores the taps in a registry key, a simple modification asking for a participant ID (as in Figure 2) when logging in and then loading the appropriate registry key is all that is required.

In the Jolley (2008) implementation of the *Tapas* software, the code was structured such that the UI and LAP API code were in the same file. The code was restructured so that the UI was separate from the LAP call-backs, and used this to create a simple program that launched the tap acquisition and recall dialogs with the taps being saved and loaded according to an entered participant ID.

For this experiment, a HP iPaq 114 running Windows Mobile 6 Classic was chosen. The device is an entry-level PDA featuring the bare minimum required for the project - a touch screen, and the Windows Mobile OS.

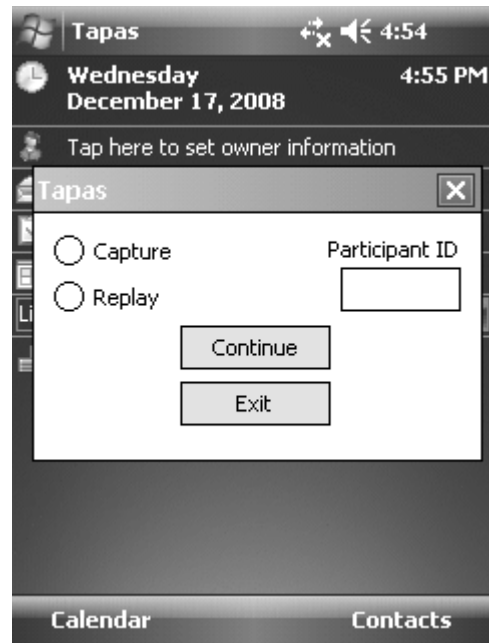


FIGURE 2 - EXPERIMENT CONTROL SCREEN

9 RESULTS & ANALYSIS

9.1 LONG-TERM EXPERIMENT

9.1.1 OVERVIEW

For this experiment, 7 participants with their own Windows Mobile devices were recruited and asked to keep diaries and fill in before and after questionnaires about their experience.

From this initial sample of 7, 2 stopped responding to e-mail after the experiment started. It is unknown if this is due to issues with *Tapas* or a lack of willingness to take part. A further 2 were then forced to drop out due to problems with the *Tapas* software. In one case, the software corrupted the phone's OS (Windows Mobile 5) causing it to crash on boot. I was unable to replicate the bug in the emulator or connect a debugger to the device so was unable to track down this problem. In the other case, the participant complained of the software disabling the physical buttons on his device (including the power button), meaning the only way he could wake up the device was to reset it. Once again, I was unable to track down the problem (there is no code in the *Tapas* LAP that handles physical button presses), and this participant was forced to drop out of the experiment.

Of the remaining 3 participants, one complained of time commitments that kept him from filling in an accurate logbook and also dropped out of the experiment. He was able to give some partial data from a fortnight of using *Tapas* on his thoughts and on bugs in the software, which will be discussed further below.

Having only the results of 2 users to analyse was less than expected, however the dichotomy of their experiences was very interesting, and this will also be discussed in further below.

9.1.2 PROBLEMS

The experiment was not without issues. Above, I identified two issues which meant two participants could no longer take part. I am at a loss as to the cause of the physical button disabling bug identified by one participant, and a further investigation into the bug by an expert on the Windows Mobile platform would be needed.

The second issue regarding the phone locking following installation is also unresolved. The device, when booted, displaying a window title of "Password", which is generated by the default LAP, which suggests that the installer did not work on this particular device and left the device half-corrupted. The inability to reproduce this bug on any other device, and lack of access to the particular device to use the remote debugging tools also left this bug unresolved.

A further problem was identified during the experiment from one of the diary study participants. In order to comply with the Microsoft requirement that locked phones should be able to call the emergency services, *Tapas* was implemented with a button on the touch screen to dial them. Unfortunately, this button is very easy to activate. On day 3, it was activated by a user absent-mindedly not paying attention to where he was tapping. However,

a more serious set of circumstances became apparent later on in the experiment though. When a Windows Mobile phone receives an SMS message, it will wake up the device and activate the touch screen for 1 minute, making it relatively easy to accidentally catch the button with keys or other items in the same pocket the phone is kept in. This happened on day 5, and then 3 times on day 10. The participant contacted me on day 10 to report the flaw and I immediately provided a version with the Emergency Call button disabled to provide a short term fix to the issue.

Two minor bugs were identified during the experiment, but were not reported to me until the end so no investigation was done into potential bug fixes. The first bug was that a user reported whilst the device was locked, any alarms set appeared to be silenced. The second bug only occurred once and was that the *Tapas* unlock window appeared on top of an incoming call screen, meaning the user had to “frantically fight the lock screen to unlock the device so [he] could receive the call”.

9.1.3 PARTICIPANT DATA

Both the participants who fully completed the study were male students studying Computer Science in their late teens or early 20s. Both had past experience with touch screen devices and currently use them as their main phone. In one case, this main phone was the device used in the experiment; however the other user had upgraded to an Apple iPhone for his main phone and used the older Windows Mobile device as a PDA only.

Both users currently use a security system on their phone, one using the default PIN system of an iPhone and the other using a graphical password system called “Throttle Lock” (ThrottleLauncher, 2008).

A third user, another male Computer Science student in his early 20s, also provided some partial results which are discussed below.

Despite the small sample size and similarity in users, the users’ usage patterns are significantly different (one being constantly used, the other being used as infrequently as once a day) in order to obtain a variety of results.

9.1.4 PASSWORDS

Although both users had previous experience with passwords, the first 3 weeks of the experiment involved them continuing to use the PIN/password system and rating their experience at the end of it. As these first 3 weeks were over the Christmas period, their phone/PDA usage patterns over this period would be atypical and not necessarily representative, continuing the experiment as planned using the 3 week control period made the most sense.

It is important to bear in mind the small sample size and that both users had previously used passwords extensively when analysing the results from the PIN/password use.

Both users gave roughly the same answers for the questionnaire regarding passwords at the end of the period, with the exception of the question regarding the effort of the login system.

Figure 3 shows the results given in the questionnaire, where 1 is the least demanding, least successful, least effort, least frustration and most insecure, and 7 is the most demanding, most successful, most effort, most frustration and most secure.

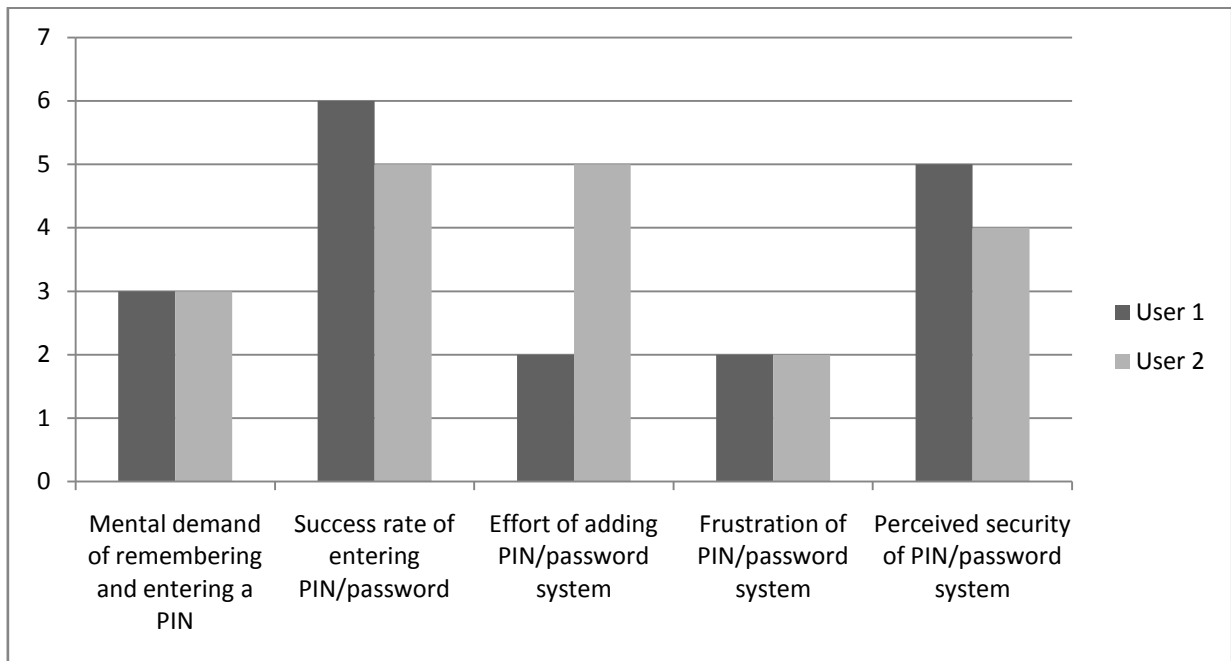


FIGURE 3 - RESULTS FROM PASSWORD QUESTIONNAIRE

The obvious conclusions to be drawn from these results are that remembering passwords does require some mental effort, but not an unduly demanding amount, and that the success rate when used frequently is high. Generally, people are used to the password system and do not find it frustrating, and have faith in the security, yet are aware of the limitations.

The difference in effort is interesting, with one user obviously feeling a considerable addition in effort level (presumably over the case of there being no security) yet the other felt the opposite. A conjectured cause for this is that users perceive security differently. If it is felt it is necessary, the effort added may be justified and the additional effort not felt as much. There is no data to support this conjecture, however.

Neither user felt the need to add any additional comments about the password system. This is possibly because password systems are ubiquitous as a security mechanism, meaning acceptance and understanding is high. The above results certainly support this conjecture.

9.1.5 TAPAS

After the end of the second 3 week block where Tapas was used, users were asked to record answers to specific questions about their experience in a questionnaire. As is expected from the dichotomy of experiences recorded in the diaries, the results of the *Tapas* questionnaire vary more between the participants than the password questionnaire.

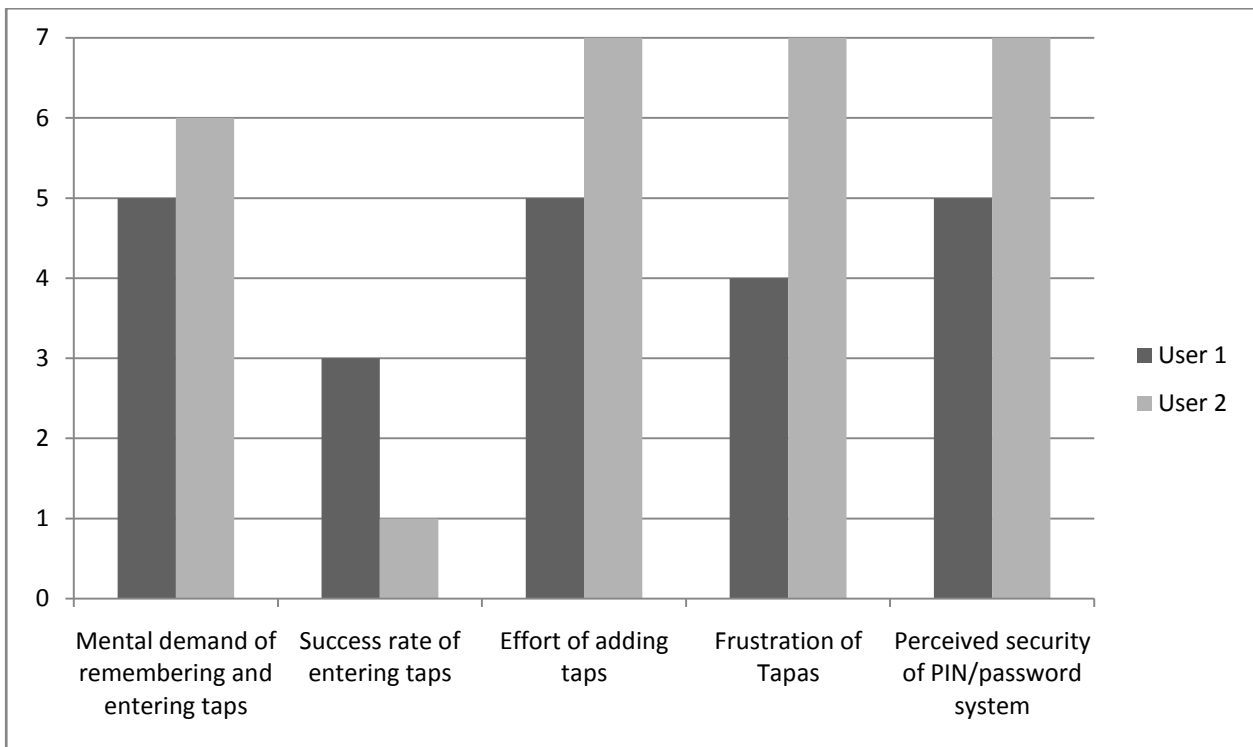


FIGURE 4 - RESULTS FROM TAPAS QUESTIONNAIRE

Figure 4 shows the results of the *Tapas* questionnaire, where 1 is the least demanding, least successful, least effort, least frustration and most insecure, and 7 is the most demanding, most successful, most effort, most frustration and most secure.

It is obvious by comparing this chart to the one for the password experiment above that in every way, with the exception of security, the participants rate the *Tapas* as worse than a password/PIN based system – more frustrating, more mentally demanding, less successful and requiring more effort. These factors are reflected in the question asking participants which system they prefer, PIN based or *Tapas*. In both cases, the users selected the PIN/password system. The question asking for the rationale behind the decision reflects the facts above, with one stating that PIN/password systems are “easier to use, easier to remember and have a higher success rate”, with the other simply stating that it’s “just easier”.

When answering the question for the reason for a low success rate, user 2 put it down to the system not recognising the taps, rather than any memory issues in recalling the taps.

With regards to the perceived security, user 1 felt it was the same as the password system, whereas user 2 rated it much more secure. User 2, however, was the one who had difficulty with taps being recognised. It is therefore reasonable to assume that user 2 thought the system was “too secure” – that is, it was so secure it kept him out.

The other user who did not fully partake in the experiment but did provide me with feedback held a different view. He felt that the system required a lower mental demand than the PIN/password system, but still experienced the same frustration with the system consistently failing to recognise taps. His final conclusion on the system was that, although

he plans to switch back to a PIN now, he would use *Tapas* if it were polished further to solve the bugs he identified.

It is unfortunate that the third user was unable to fully partake, as it would be interesting to see if he suffered from the same touch screen problem as user 2 or not.

9.1.6 ANALYSIS OF DIARIES

Both diaries reported a dichotomy of experiences, even from the very start. One user expressed issues installing the software and required help from a friend to complete the process, however the other did not. The user that had the issues was the one who was less out of practice with using the Windows Mobile platform; therefore it is likely the issues were as a result of the Windows Mobile platform rather than the software itself, especially considering that *Tapas* was packaged as a standard Windows Mobile .CAB installer, with the caveat that some security certificates had to be installed first. A production version of *Tapas* would not have this issue as it could be signed with a commercial Authenticode certificate instead of a self-signed certificate.

The other two users who dropped out after installing *Tapas* reported no difficulty with the installation procedure itself, and both of these users used the Windows Mobile device as their primary phone, which appears to support my theory above.

Once the initial installation was complete, both users reported no major problems with enrolling their initial taps. The only issue reported was that of the tap timeout. *Tapas* assumes the sequence is complete if there has been 3 seconds of no activity, meaning that no tap sequence can have a gap bigger than 3 seconds in it, and also that you must wait 3 seconds after entering your tap sequence before the device unlocks. As a result of this, one user made the comment that he “had to tap a bit quicker than expected” with his tap sequence, however the other felt there was a “lag” between him entering his taps and the device unlocking.

One point of note with the configuration is that one user needed to continuously revisit the configuration screen to set the time out time until he found a satisfactory lock-out time that balanced security and annoyance. It was not until day 13 he seemed to find a satisfactory time out time – 10 minutes.

Where the participants used *Tapas* on their main device, it was used in a considerably wider range of environments than the participant who used it only on his PDA. User 3 noted in his feedback two particular environments where he felt *Tapas* excelled where a PIN/password system would not.

These environments were whilst being a passenger in a car, where you may have difficulty holding the screen steady, and in a brightly lit environment where the screen may be washed out from the light. As the precise location on the touch screen is unimportant for taps, this meant that *Tapas* was successful where a traditional approach may not have been.

User 1 was limited to use in fairly uninteresting circumstances, likely a result of the PDA not being his main portable device. This may have contributed to the relative success of his usage of *Tapas*.

User 2 also only had one non-standard environment of interest, which is that of being intoxicated. This happened on days 6 and 15. On both occasions the user noted that it was “impossible” to use whilst drunk, however the user successfully managed to use the backup password feature and disable *Tapas* for the duration of the evening suggesting that the impairment was focussed on the ability to accurately recreate the rhythm.

User 2 did see some benefit of not being able to log in to the system, noting this “might actually be a good thing a bit like Google goggles”, presumably helping to avoid sending embarrassing SMS messages, in a similar way to the Google Mail Goggles system aims to avoid users sending e-mails they may later regret (Perlow, 2008).

However, when considering the day-to-day use of the system, user 1 and user 2’s diaries differ wildly. The partial feedback from user 3 also suggests his experience with *Tapas* was similar to that of user 2.

Where user 2 made note of occasions where he could successfully log in first time, user 1’s occasions of note were where he could not log in first time. The frustration user 2 felt with the system is clearly recognisable from the comments made in the diary, as well as his final questionnaire (discussed above). He described the system as annoying multiple times, even on relatively good days. His success rate improved towards the end of the third week (day 18 onwards), yet first time successes were still rare.

On day 7, user 2 makes the diagnosis of the problem that “it doesn’t seem to pick up all the taps”. This diagnosis is consistent with the conclusion I reached whilst investigating issues in the security experiment in section 9.2.2.

User 1’s use of *Tapas* is fairly uninteresting, in that it worked pretty much every time. Of the 23 days of diary entries he made, 15 of them did not report any issues claiming that it worked first time every time, and the majority of the remaining entries are brief notes such as “Second time lucky” or “Failed to log in twice, grr”. As log in systems are supposed to limit inconvenience for legitimate users, this shows that even small failures which cause minor inconvenience to the user greatly increase the frustration level.

User 3’s experiences seem to lie in the middle of these two extremes. He mentions that *Tapas* “doesn’t always recognise the tap pattern immediately” citing one occasion where it took him 5 attempts to log in. More detailed analysis is unfortunately not possible.

User 3 also brings up the issue of security with his partial results, mentioning that his phone has a screen hardener so taps are “clearly heard”, mentioning on one occasion that his phone was unlocked by a friend who had simply overheard the taps.

User 1 also had a similar situation, however his friends had not overheard him entering his taps first, they simply guessed the sequence to unlock the device and then changed the stored tap sequence as a practical joke.

I will address this issue in further depth in the security experiment in section 9.2.3.

The general conclusion to be drawn from the diary study is that the usability of *Tapas* appears to vary wildly depending on external influences – mainly the sensitivity and accuracy of the touch screen on the specific device.

9.2 SECURITY EXPERIMENT

9.2.1 OVERVIEW

Ten users took part in this experiment in pairs, meaning 5 sets of results were gathered.

These users were placed into pairs with one user creating a tap sequence, and the other then attempting to guess it and break into the system. In 3 of the attempts, the user was able to break into the system with no knowledge of the tap sequence, and in the remaining cases was able to break in immediately after overhearing the user enter their taps.

9.2.2 PROBLEMS

Despite what appeared to be the initial success of my pilot experiment, a first attempt at running the experiment ran into problems. Four participants (2 pairs) took part, and both experienced similar problems.

The users, both of who were not familiar with *Tapas* (unlike in my pilot experiment) were asked to enrol and given advice of using a “simple” and “complex” tap sequence respectively. The user with the simple tap sequence had difficulty enrolling (the two tap sequences reportedly did not match), but was eventually able to do so successfully, and the user with a “complex” tap sequence was unable to do so, and with each attempt simplified his tap sequence until he was able to successfully enrol. Additionally, he was tapping harder and harder in order to make the screen recognise his taps, until I was forced to intervene and ask him to stop tapping with such force as I was concerned that the device may be damaged.

When both users were eventually able to enrol, they were then unable to successfully repeat their tap sequences until after a number of attempts. Due to the very short period of time between enrolling and entering the taps (a matter of minutes), it appears unlikely forgetting the tap sequence would have been the cause; however one participant stated that they had forgotten which tap sequence they had eventually used, as he varied the tap sequences progressively making them simpler until he found one that worked.

It appeared that this experiment suffered from the same issue that Marriner (2007) discovered, where the system appeared to not be detecting the taps. One participant suggested it would be useful if the screen showed feedback when a tap was detected. Marriner (2007) had this feature in the prototypes, but Jolley (2008) did not implement them into his version, reasoning that the flash “was not very welcoming”. However, usability

issues resulting from a lack of feedback when entering information using a touch screen is a well understood problem (Schedlbauer, 2007).

Running the software in an emulator as part of Visual Studio and emulating the stylus using mouse clicks did not appear to have the same problems and all taps were registered. This suggests that the Marriner (2007) is correct in the diagnosis that the touch screen itself is the problem with tap detection, rather than anything the software is processing.

In both cases, the attacker was able to access the system and successfully recall the tap sequence during the “quiet overhearing” stage, however I suspect that due to the difficulty users had getting their systems to accept their taps on recall, it was only through chance that they were able to match their taps and log in. With the attackers, I suspect something similar was occurring, therefore over a sufficient period of time the attacker would have broken in with only very vague knowledge about the tap sequence, such as the length.

Due to the problems with the experiment, it was initially abandoned with an initial investigation done into the failures and modifications to the experiment to work round them. A modified version of the software which drew a line on the screen when a tap was detected was developed which confirmed the suspicion that the touch screen was not detecting all the taps as lines were not consistently being drawn.

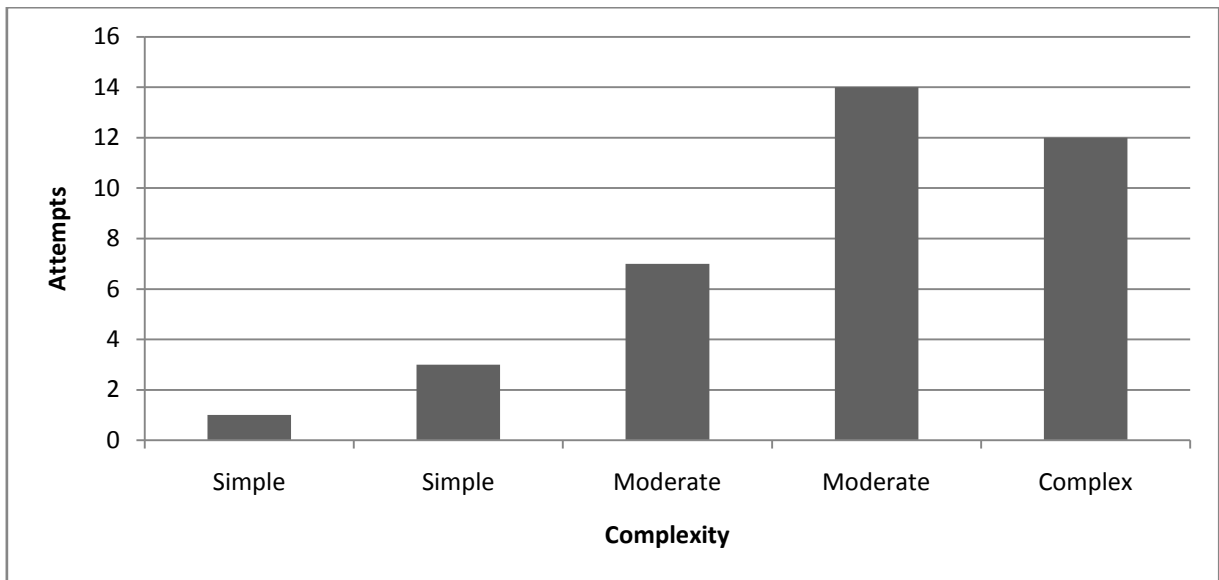
At first, plans were drawn up to emulate the tap sequence using a touch pad on a laptop and the Windows Mobile emulator (tapping the touchpad with a single finger emulates a mouse click and is similar in mechanism to touch screens that do not use a stylus). However, following the conclusion of the diary study and initial analysis of results, it would appear the same symptoms were exhibited on one device, but not the other (discussed in detail in section 9.1 above). The participant who did not have problems with his PDA detecting taps graciously agreed to let me borrow the device for this experiment, and the usability experiment.

The new device, a T-Mobile MDA, was manufactured by HTC, the same company that produces the O₂ XDA as used by Jolley (2008). It seems likely that HTC use a more sensitive type of touch screen than the other manufacturers that appear in this experiment and as originally used by Marriner (2007).

Testing was undertaken with the new hardware and the debug software showed that the touch screen was picking up taps much more consistently. The experiment was restarted as originally planned.

9.2.3 RESULTS

Five experiments were performed, 2 pairs were given the hint to use a “simple” complexity sequence, a further 2 were given the hint to use a “moderate” complexity, and the remaining pair to use a “complex” sequence, however there is no current quantitative definition of complexity so the users were asked to use their own judgement.



Complexity	Dumb Attack	Busy Overhearing Attack
Simple	First attempt	N/A
Simple	Third attempt	N/A
Moderate	Seventh attempt	N/A
Moderate	Unable to break	Fourth attempt
Complex	Unable to break	Second attempt

FIGURE 5 - RESULTS FROM SECURITY EXPERIMENT

Figure 5 shows the results I gathered in the security experiment, in both a graphical and text-based form. Three of the 5 sequences were broken by the attacker simply guessing the sequence – in one case even immediately, with the remaining 2 sequences broken after they were overheard by an attacker. In the chart, this is shown by the overall number of attempts. Attackers were given 10 attempts at each round before the extra information (overhearing the taps) was supplied. Five rounds were designed as part of the experiment however in all cases the device was successfully unlocked during the first 2.

Although there is not enough data here to form any strong correlations, there does appear to be a trend just by looking at the above chart. In the experiment, the simpler sequences were broken in the first round, whereas the sequences broken in the second round were more complex. Further investigation based on this initial investigation may be beneficial.

From this experiment I aimed to investigate the security of *Tapas* and whether or not complex tap sequences are more secure than simpler ones.

From the results above, it is obvious that *Tapas* is not very secure, in that in all cases it was broken either by guessing the tap sequence or by simply overhearing the taps in an environment with a realistic amount of background noise (in this case, either a computing lab during a busy practical, or in a newspaper office).

For the second question, from the above data there does seem to be a correlation between complexity and ease to break, although there is not enough data to determine whether this is statistically significant or not.

I have done a more theoretical review of the security of *Tapas* in section 9.2.4 which may explain the abysmal results demonstrated above.

9.2.4 FURTHER COMMENTS

It should be noted that even participants who were given the hint to use a “simple” sequence gave a tap sequence more complex than the simplest sequences given in the usability experiment, therefore there is a potential bias as users knew the experiment was related to the ability to break in to the system prior to registering the tap sequence so may have made their taps more secure. Conversely, in the usability experiment the participants may have made their taps simpler than a normal user would have in order to help ensure they would remember them.

A future study could possibly be modified so users do not know their taps will be broken into in advance, in order to better simulate a real life attack.

As discussed in section 6.3, *Tapas* calculates strings consisting of the character *s*, *d* and *u* to represent the rhythmic contours for rhythm comparison, with a theoretical maximal space of 531,441 permutations.

In order to discover if users would indeed create tap sequences with sufficiently unique timings, or instead stick to familiar patterns, a copy of all the rhythmic contour strings captured in both this and the usability experiment was created and analysed.

The first analysis was to determine how many unique patterns were captured. Out of 48 different strings, 26 (54%) were unique, with “ssssssss” being the most common rhythmic contour string, appearing 7 times (15%). This pattern translates into roughly 5 equidistant taps. The other most common patterns were similar to this differing only in the number of taps, but still consisting entirely of the ‘s’ character.

Further analysis was done on this raw data to discover the variance in rhythmic contours strings – that is, looking at the frequency of the particular characters.

As Figure 6 shows, the vast majority of taps (464 of the 510 rhythmic string characters) were converted into the ‘s’ character, meaning the tap was similar in length to the one before it. 22 (46%) of the rhythmic contour strings consisted solely of the character ‘s’.

In order to maximise our security, taps generated would have to be random with no bias towards a particular character (i.e., each character should occur one third of the time). The chi-square test can be applied between our observed data and the data expected if it was truly random and maximised the space, the result is $p < 0.0001$, which is a significant variation.

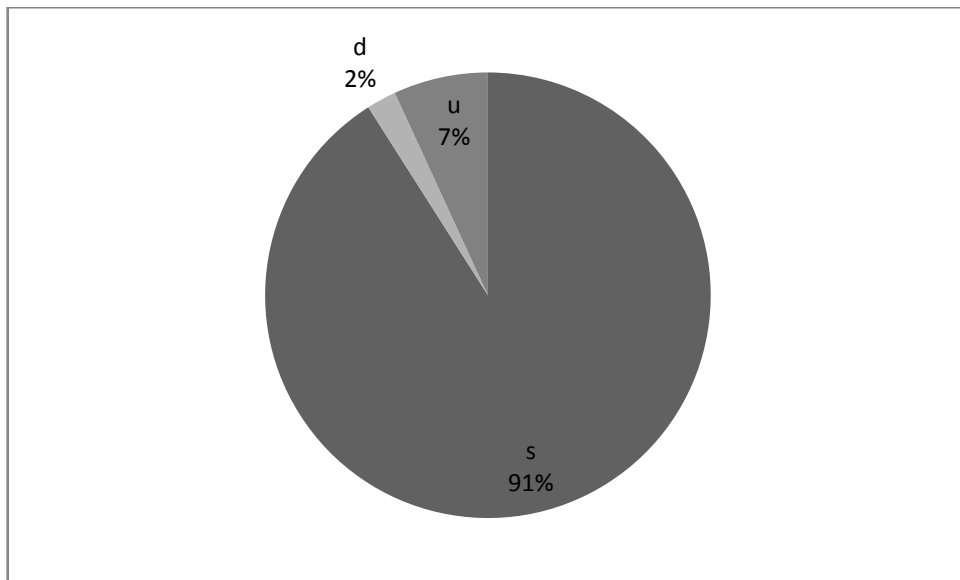


FIGURE 6 - HOW TAPS CONVERTED INTO THE RHYTHMIC CONTOUR

Using the default match percentage of 80%, if an attacker were to attempt to log in using a series of equidistant taps knowing the correct number of taps (that is, a string of the correct length, just consisting of only 's'), 39 of the 48 strings (81%) would match.

Considering there were no recorded string lengths longer than 14 taps, it would be trivial to try each sequence length in turn, and this attack would be very successful.

This is a significant weakness in the system and further research would be required to potentially identify a new or improved algorithm that addresses this concern. One potential source is an investigation of the '1.5' constant used in the algorithm, where it is used to discern if the time between two taps is close enough to be an 's' or not, and its use within the implementation, where it is casted to an integer for comparison.

Another possibility is that, unprompted, *Tapas* users simply do not come up with a sufficiently distinctive tap sequence, and that further guidance needs to be given into creating a strong tap sequence, similar to how guidance is given to users creating a strong password (Microsoft, 2006).

Such guidance may include varying time between taps more, as well as holding the stylus down on the screen for longer between taps. The second piece of advice would also be useful to confuse attackers who overhear the sequence, as lifting the stylus off the touch screen makes very little sound and the potential attacker would not have the entire sequence.

However, giving this advice may damage the memorability of the sequence, so any advice given must be carefully considered so as not to fall into the same trap as advice given about strong passwords generally falls into, which is not considering the memorability of the stronger password (Yan, Blackwell, Anderson, & Grant, 2000).

This suggested advice needs to be investigated further in order to gauge its accuracy. If it does turn out to be accurate, then a quantitative measure of the strength of a tap sequence can be obtained by measuring the Hamming distance between the rhythmic contour string and a string of the correct length consisting solely of the 's' character.

The general consensus of both this analysis and the results of the experiment given above shows that *Tapas* in its current form is ineffectual in providing a sufficiently secure platform for authentication.

9.3 USABILITY EXPERIMENT

9.3.1 OVERVIEW

In this experiment people were asked to enter a tap sequence into the device and then invited to recall the sequence in a week. A small control group was also used who instead remembered a password.

42 participants were recruited into the main group and 4 into the control group. Of these, 29 (69%) returned after one week (in some cases this was 8 days, rather than 7 however the extra day did not seem to have a measurable effect on my data), and 2 (50%) of the control group returned.

Of the 42, 40 (95%) were able to successfully enrol their taps in the system. Of the returning 29, 24 were able to correctly recall and enter their tap sequence (83%).

More in depth analysis of my results appear below.

9.3.2 PROBLEMS

In hindsight the design of my experiment was flawed with regards to the control group, compounded by being unable to locate 50% of my control participants after 1 week. A more sensible design would have been to ask users to create both a password and tap sequence for the device and then ask users to recall both, similar to Dhamija & Perrig (2000). I had

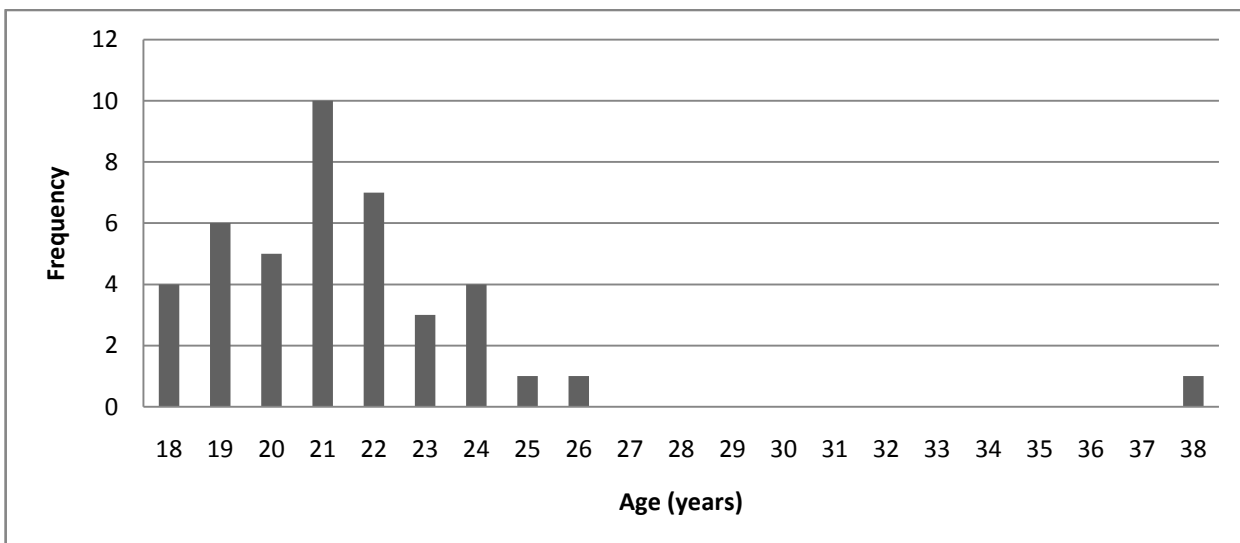


FIGURE 7 - AGE OF PARTICIPANTS

originally aimed to collect 90 *Tapas* participants and 10 control participants; however this target proved to be unrealistically high.

In order to work around this issue, the data from Dhamija & Perrig (2000) with regards to password and PIN recall after one week can be used in this experiment.

9.3.3 PRE-CAPTURE

The pre-capture section of the questionnaire aimed to gauge the previous experience users have had with touch screen devices, as well as their age and course.

The majority of the people who took part were Computer Science students in their late teens or early 20s, with one outlier at age 38 as shown in Figure 7. The remaining participants were also in the same age range, but instead studied arts subjects, such as History, English Literature and Politics.

Figure 8 shows the experience users said they had with touch screens, with 1 being “no experience” and 7 being “used daily for an extended period of time”. The mode here is to have a large amount of experience with touch screen devices, yet the other not so strong positive values appear to be dwarfed by the less experience values. This suggests that if users do use touch screen devices, for example if they own one, they use it frequently, whereas users who do not own such a device do not get a chance to use the technology.

This is supported by comments made by participants when answering the question,

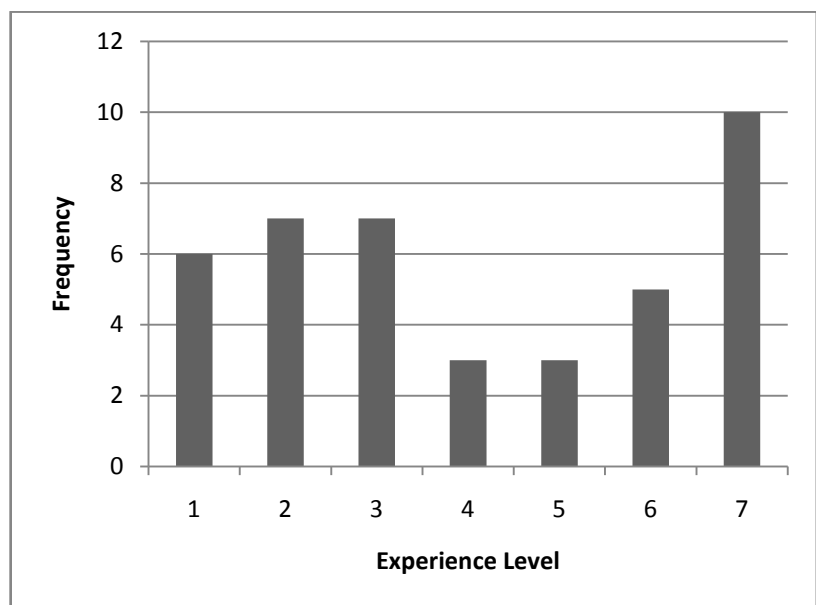


FIGURE 8 - USERS WITH TOUCHSCREEN EXPERIENCE

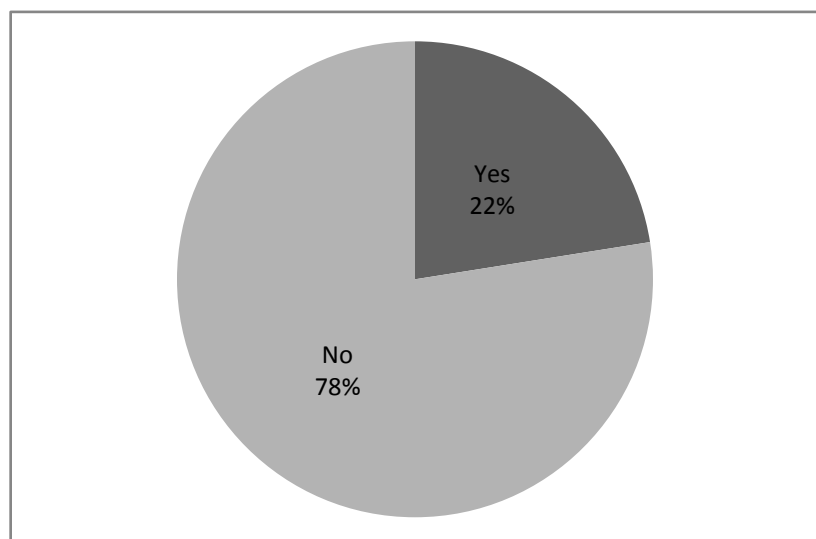


FIGURE 9 - USERS WHO USE A PIN ON THEIR DEVICE

such as “my friend has an iPod Touch”.

The next questions asked the users about their current security habits, firstly focussing on whether or not they used a PIN or password system to protect their phone, as shown in Figure 9. In 78% of cases, the users used no security on their phone, as suggested by Clarke et al. (2002) which gives strength to the argument that current mobile device security as used is not sufficient. When asked for the reason why they do not use it, they typically fell into two camps: they do not want the inconvenience, or they do not perceive a benefit to the additional security. It is important to remember that all participants in this experiment were students. Although it may seem that the type of data students store on their phones or PDAs are uninteresting to professional data thieves, the device may contain information such as private text messages or e-mails they would prefer not to release to their friends which they may not have considered. Some users did see a benefit despite not using security currently, with 3 making comments similar to “I should” in their answers.

Other answers of note included ones where the participant was unaware whether her phone contained the functionality, or in another case where the participant had simply never considered the idea. One participant also made the comment that he believed any security could easily be bypassed by professional data thieves, but then expressed that this was “bad reasoning”.

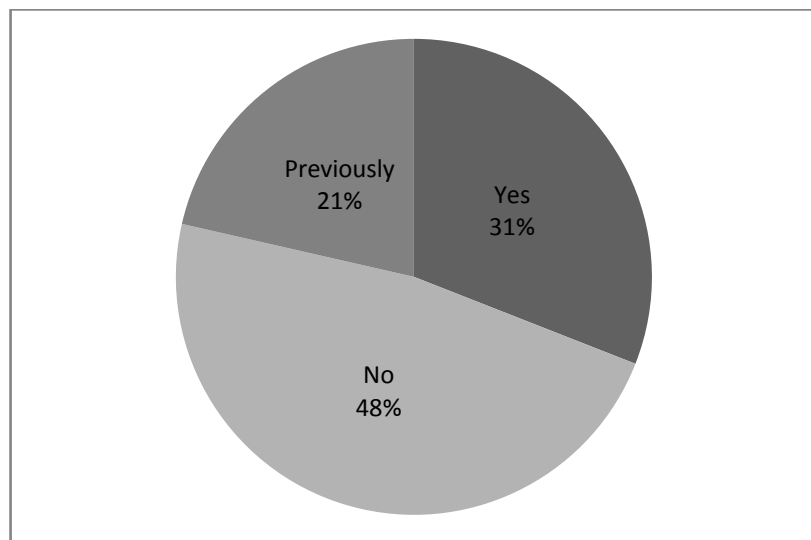


FIGURE 10 - USERS WHO PLAY OR HAD PREVIOUSLY PLAYED AN INSTRUMENT

Of the 9 users who did use security, 7 of those are only prompted when their phone is turned on. As most users kept their phones on constantly, this renders this form of protection useless, meaning that only 4.5% of the participants surveyed effectively secured their mobile phone.

The final question asked related to participants’ musical experience, as Jolley (2008) expressed a possible relationship between the ability to remember patterns, and the whether or not a person is musical.

As Figure 10 shows, there are two approximately equal groups of musical and non-musical participants in the experiment. Whether or not a musical background makes a difference will be analysed later.

9.3.4 POST-CAPTURE

Users were asked to create a tap sequence in the device using the enrolment procedure from the main *Tapas* software but modified into the software created in section 8.2. For this, users would have to enter the same (or a sufficiently matching) tap sequence twice.

One problem noticed by many users was that the 3 second lag after of the last tap appeared to be too long. After entering the first tap sequence, many users appeared to be confused about what was going to happen next for the brief moment until the tap screen appeared. Many users also appeared to be unsure on how to cancel the dialog box, just doing a tap pattern immediately without first clearing it by tapping the “OK” button. Although the dialog box is a standard UI element from the Windows Mobile API and this suggests an issue with the affordability of the Windows Mobile system rather than *Tapas*, *Tapas* could work around this. Instead of showing a dialog box after the first tap sequence, the text on the canvas could simply change to inform the user to enter the second set of taps.

Users were given a limit of 10 attempts to register their tap sequences, although some gave up out of frustration earlier (typically after the 3rd failure) and had to be prompted to continue the experiment. Of the 42 participants, 2 were unable to successfully register their taps, and a further 3 took more than 5 attempts to successfully register their taps. Promisingly, 13 (31%) managed to enrol first time. Many users took their first attempt in order to familiarise themselves with the system and touch screen, and some were confused by the instructions to “enter their tap sequence twice” and did so without waiting for the dialog box after the final tap to occur, therefore their first

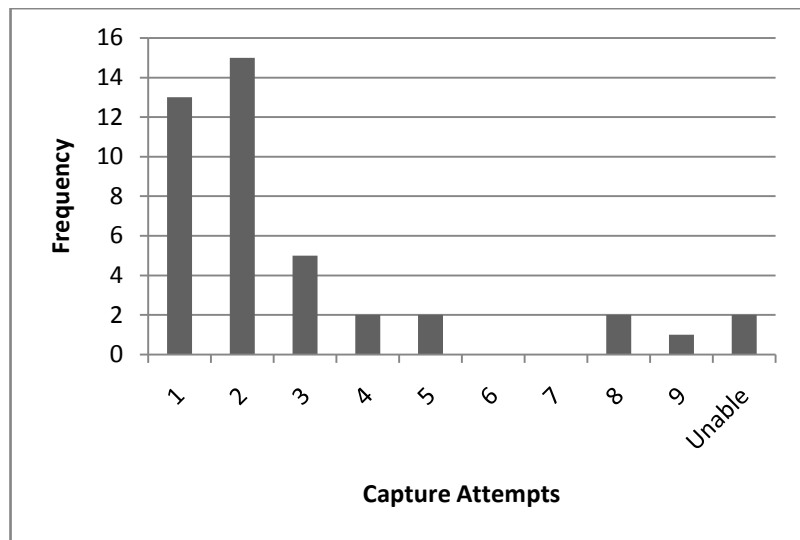


FIGURE 11 - NUMBER OF ATTEMPTS FOR A SUCCESSFUL TAP SEQUENCE CAPTURE

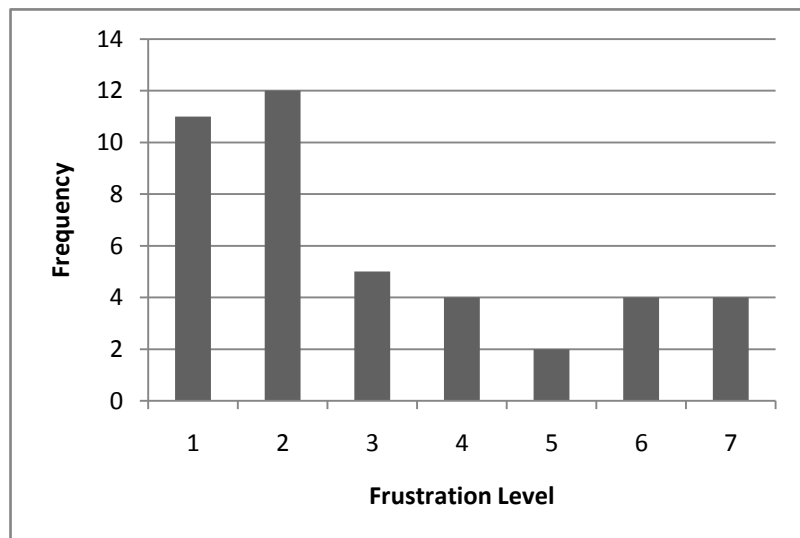


FIGURE 12 - CAPTURE FRUSTRATION

recorded pattern was actually the same pattern twice. If we look at the amount of successful registrations on the second attempt we find 15 successes (the mode value), meaning that 28 (67%) of enrolments are successful after the second time. Figure 11 shows the overall number of attempts required to successfully enrol.

As users made more attempts to enrol in the system, they often changed their patterns, making them simpler.

The 4.7% failure rate for enrolment compares favourably to the figure of 5% given by Dhamija & Perrig (2000) for a person who could not successfully log in using the recently created password. However, the two students who were unable to register were Computer Science students who must use passwords as part of their course. Similarly, all users must have at some point created and used a password (for it is a requirement of attending the University to regularly check e-mail) which suggests that 4.7% failure rate is actually unacceptably high. This is supported by the control study where all users were

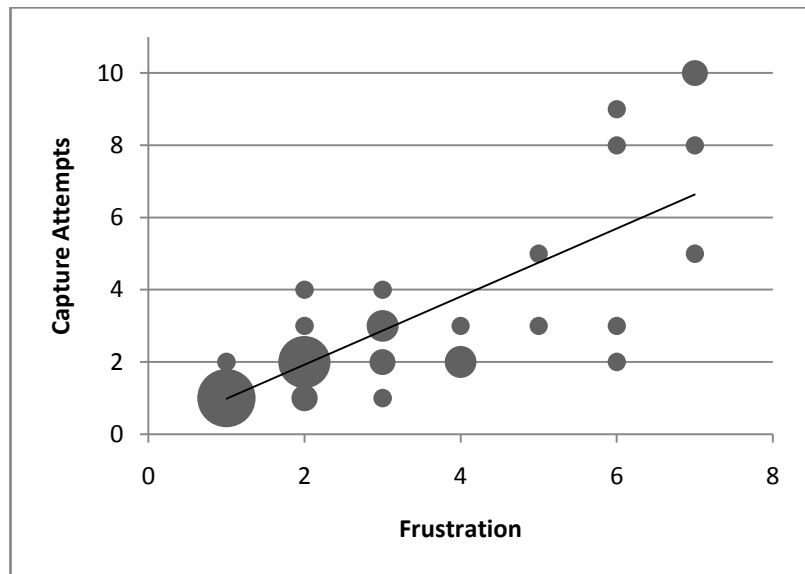


FIGURE 13 - CAPTURE ATTEMPTS VS. FRUSTRATION

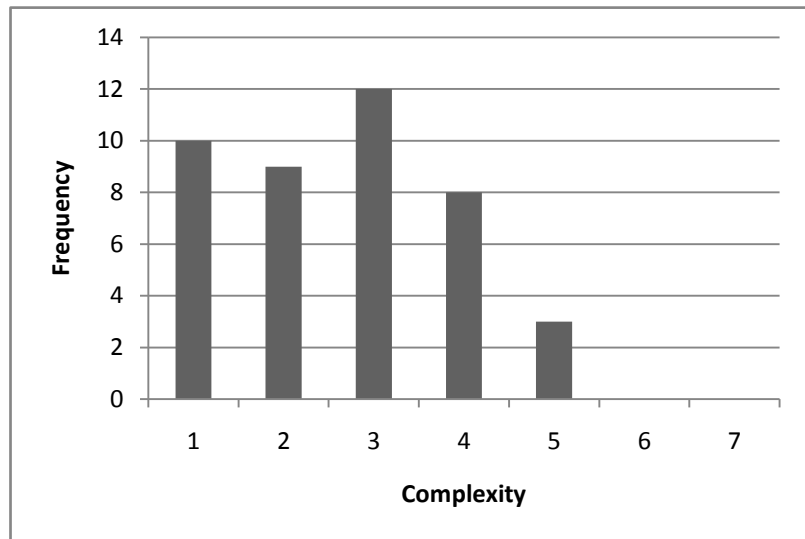


FIGURE 14 - PERCEIVED COMPLEXITY OF TAP SEQUENCE

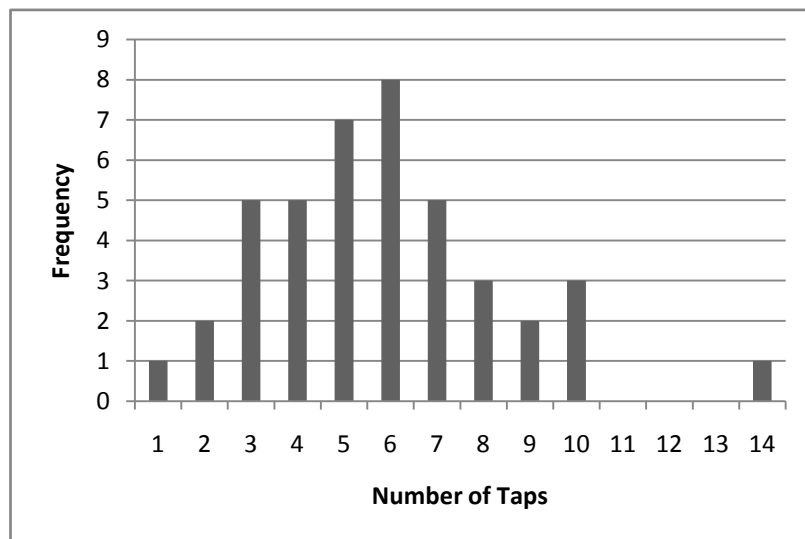


FIGURE 15 - NUMBER OF TAPS IN SEQUENCE

able to successfully register a password. Indeed, to an external observer, the fact without context that 1 in 20 people cannot successfully register with the system would appear to be a major problem.

Following the capture of the tap patterns, users were asked how frustrating they found the system. The frequency of answers is shown in Figure 12, where 1 represents the least frustration and 7 the most. We can see that the shape of Figure 12 compared to Figure 11 are similar, which suggests there may be a correlation between the two. Figure 13 plots the number of capture attempts against the frustration the user felt in a bubble graph and includes a linear trend line. For the purpose of analysis, failure to enrol is recorded as 10 attempts on this chart.

We can see that this plot and trend line appears to support this argument. The Pearson product-moment correlation coefficient (Cairns & Cox, 2008) for this graph is 0.80 which also shows that there is a strong correlation between the two.

Users were also asked to note information about their taps: how complex they felt their taps were (Figure 14 – with 1 being simple and 7 being complex), and secondly how many taps their sequence consisted of (Figure 15).

Figure 14 shows a bias towards simpler tap sequences, with no users choosing the extremely complex option. This may be due to the fact participants were aware that they would have to recall the tap sequence so deliberately chose ones that were easy to remember, but which possibly would not be secure.

Some users took this to their extreme, choosing very short tap sequences, as shown in Figure 15. The extreme case of a single tap was registered by a user who was frustrated by the system after only 2 attempts and decided to protest with a single tap.

The original software by Marriner (2007) did enforce a minimum tap length, but this was not implemented into the Jolley (2008) version of the software. It makes sense that shorter tap sequences would be easier to break into; however there has been no empirical evidence to support this. Indeed, with tap sequences consisting of less than 3 taps, there is no room for variance with the default 80% match threshold, as each character in the rhythmic contour represents 25%, so a single character wrong would drop you below the limit. This would make entering the sequence harder as it would have to be perfect every time. It stands to reason it is easier to consistently reproduce a 2 tap sequence than a longer sequence, however.

Figure 15 shows that sequences of length 6 are the most favoured. This is lower than the figures gained by the pilot studies of Marriner (2007), where a mean of 7 was determined. Once again, I believe this is due to users choosing passwords that are easy to remember as they knew the task beforehand and desired to explore well. A standard deviation of 2.6 with a mean of 5.8 however shows that there is still a reasonable spread of tap lengths.

An interesting question that we can consider here is whether or not more complex taps take more attempts to enter. As we can see from Figure 16, there appears to be no strong correlation, and this is supported by the Pearson product-moment correlation coefficient of -0.05 suggesting there is no correlation.

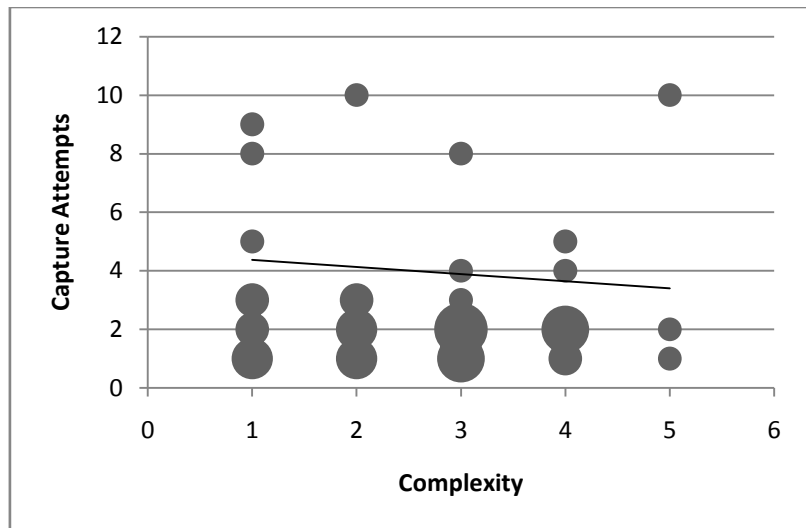


FIGURE 16 - ATTEMPTS VS. COMPLEXITY

Another question we can ask is to consider whether the number of attempts is dependent on the user's previous experience with touch screens. That is, do users who have more experience with touch screens have fewer problems with the system?

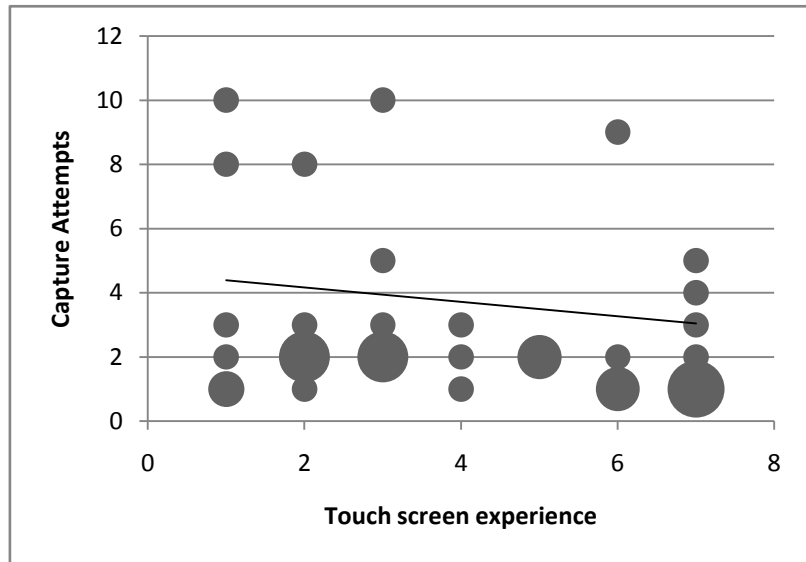


FIGURE 17 - ATTEMPTS VS. EXPERIENCE

Figure 17 shows that there is no strong correlation between the two, although most of the extreme capture attempt values (including the two cases where the users were unable to successfully register their taps) were by those users who recorded an experience level of 3 or less. However, with a Pearson product-moment correlation coefficient of -0.21, there appears to be limited statistical evidence to support this.

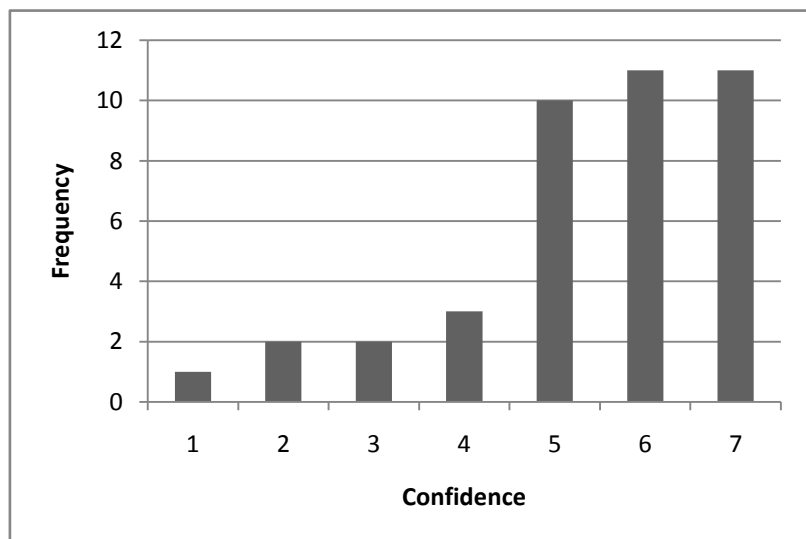


FIGURE 18 - POST CAPTURE CONFIDENCE

The final data gathered in the post-capture section of the questionnaire asked the 40 participants who were able to successfully enrol how confident they felt on whether or not they would be able to recall their taps in a week. Figure 18 shows the frequency of answers to the question, with 1 being the least confident and 7 being very confident. With a mean of 5.40 and a standard deviation of 1.55, this suggests that most people were confident that they would be able to remember their values.

9.3.5 RECALL

The recall part of the experiment was conducted approximately a week after the registration part. In some cases, it was 8 days afterwards due to differing commitments. 29 of the 40 who were able to

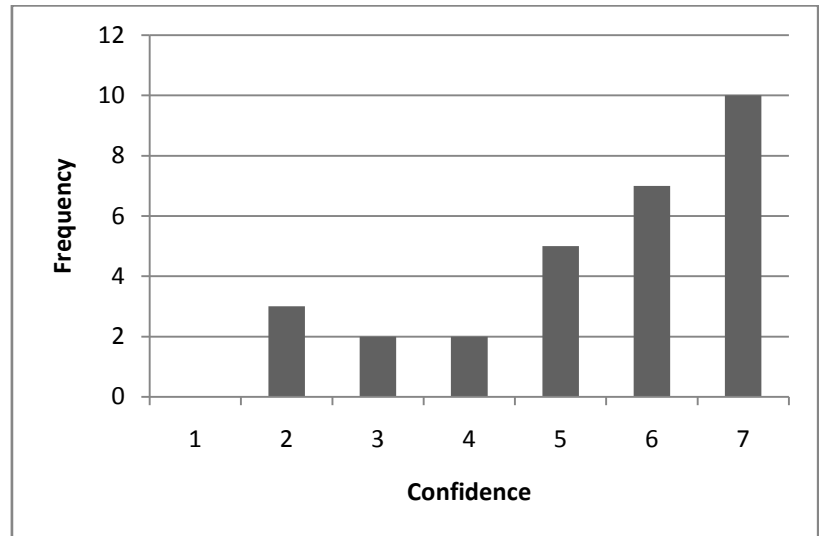


FIGURE 19 - CONFIDENCE PRIOR TO RECALL

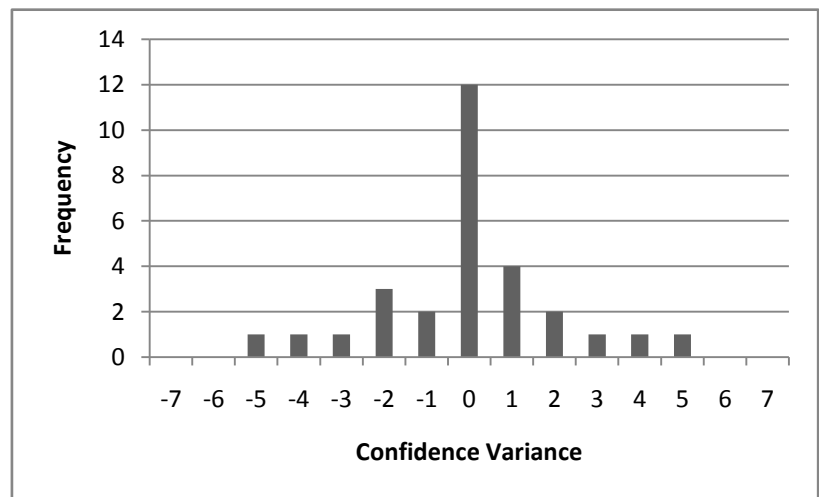


FIGURE 20 - CHANGES IN CONFIDENCE

successfully register were able to take part in the second part of the experiment.

Before users were asked to recall their tap sequence, they were asked as to how confident they felt that they would be able to accurately recall their tap sequence.

Figure 19 shows the frequency of how confident users felt before entering their taps. With a mean of 5.41 and a standard deviation of 1.68 this shows that the general confidence is approximately the same as immediately after the capture. This is supported by Figure 20 which shows how users' confidence changed between the first and second time they were asked.

In most cases, there was no change in confidence, with 9 people saying they felt more confident, and 8 people saying they felt less so. A standard deviation of 2.12 supports this showing that most people’s change in confidence, if any, was minor.

Users were then asked to log in to the system using the taps they created the previous week. As Figure 21 shows, most users succeeded with this, with 52% being able to immediately log in.

Although this initially looks promising, it is important to consider the other factors and data. 5 users were not able to successfully replicate their tap sequence and were therefore unable to log in – this represents 17% of participants.

This at first appears favourably compared to the results gathered by Dhamija & Perrig (2000), which found between 30-35% of users were unable to log on using their PIN or password after one week. However, applying the chi-square test (Cairns & Cox, 2008) to the data with the null hypothesis of *Tapas* being as memorable as passwords

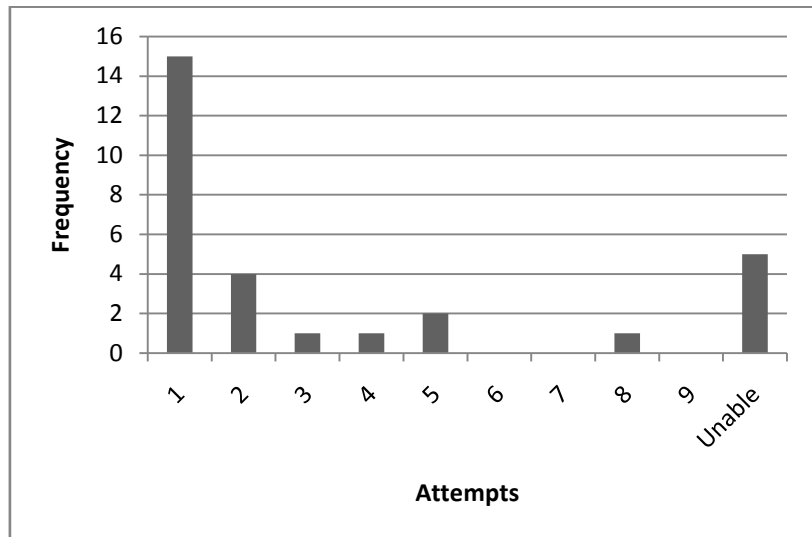


FIGURE 21 - RECALL ATTEMPTS BEFORE SUCCESSFUL LOGIN

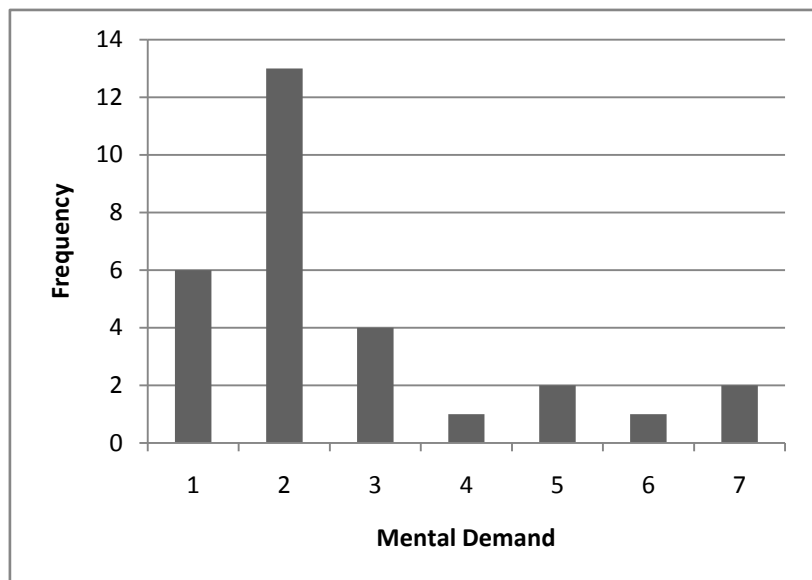


FIGURE 22 - MENTAL DEMAND OF REMEMBERING SEQUENCE

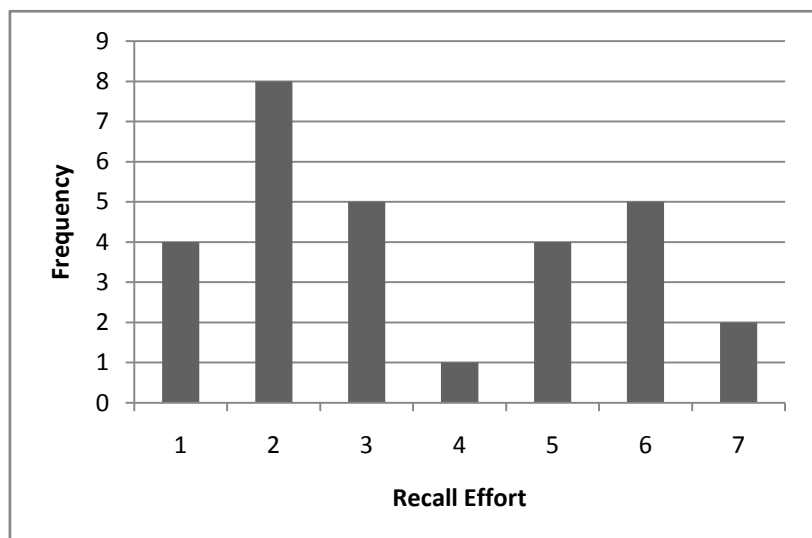


FIGURE 23 - EFFORT TO RECALL AND ENTER TAP SEQUENCE

using the data collected in Dhamija & Perrig (2000), we find the value $p = 0.0794$, which is not small enough to be considered a significant statistical variance.

For the 5 people who were unable to recall their taps, they were asked for the reason why. Two of the users said that this was due to *Tapas* being unable to recognise their taps. As one of these users was the one who choose a tap length of a single tap, this could indeed due to be a bug in the system when dealing with single taps. The remaining 3 simply put the failure down to the forgetting the tap sequence.

After confirming their taps (or otherwise), users were then asked on their opinions on the system. Firstly, how much mental demand they felt having to remember the tap sequence added, secondly how much effort they felt having to log in to the tap system required, and finally how frustrating they found the system. For each question, they were asked to rate their experiences on a scale of 1 (representing the least mental demand, effort or frustration) to 7 (representing the most mental demand, effort or frustration).

Figure 22 shows the mental demand users experienced whilst remembering their sequence. With a mode of 2, a mean of 2.7 and a standard deviation of 1.7 this shows that generally all users did not think they had an unduly difficult task to remember their tap sequence.

Figure 23 demonstrates the answer to the question of how much effort the participants felt they shows a slightly different picture with a far from normal distribution with a mode of 2, but a mean of 3.6. Although more participants chose below 4 as their option than those that chose above 4, this still means a considerable proportion of the participants felt that recalling and entering the tap sequence required quite a lot of effort.

Figure 24 also shows a fairly even spread, with a mode and mean of 3 and 3.3, although twice as many people choose below 4 as those who choose above 4, this suggests that although most people experience a low but non-trivial level of frustration, but there are still a considerable amount of people who do find the system frustrating.

One extreme example of note is of a participant who could remember the film which had the song he based his pattern off, but not the specific song and spent 4 minutes thinking hard to recall it. When he eventually did, he managed to successfully log in first

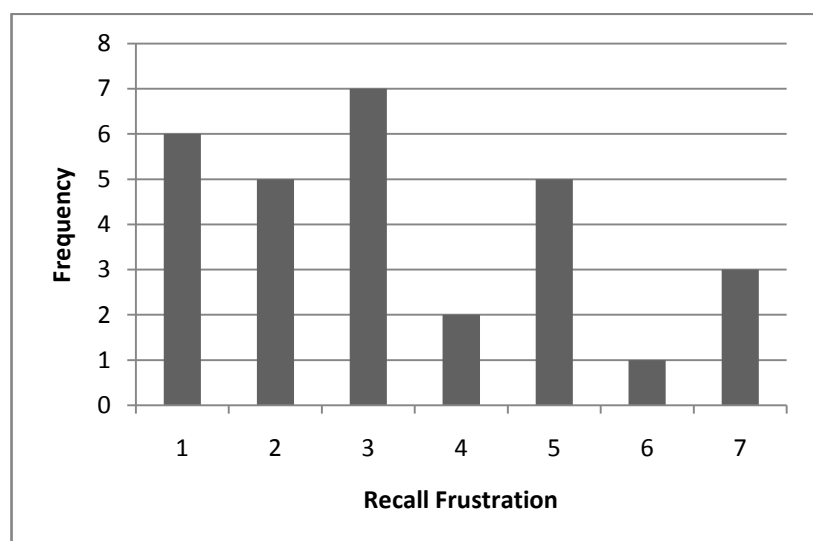


FIGURE 24 - FRUSTRATION OF RECALLING TAP SEQUENCE

time.

Now we have this data, we can start to answer the questions that it raises. Figure 25 shows a breakdown of login attempts, but this time analysed by the musical ability the user claimed to have. This shows some interesting data, specifically when it comes to the participants that were unable to log in. Every user who claimed to play a musical instrument managed to log in eventually, and it would appear a disproportionate amount of non-musical users were unable to log in (31%, compared to the standard 17%). If we apply the chi-square test to see if this variance is statistically significant, we get $p = 0.0002$ which does show that this variance is indeed statistically significant.

We can also ask whether or not users' confidence is misplaced, and whether or not their confidence made a difference as to whether they could reproduce the tap sequence or not. Figure 26 shows the results of this analysis that even a cursory glance showing that the number of login attempts appears to bear little relationship to the confidence the users felt, which is further

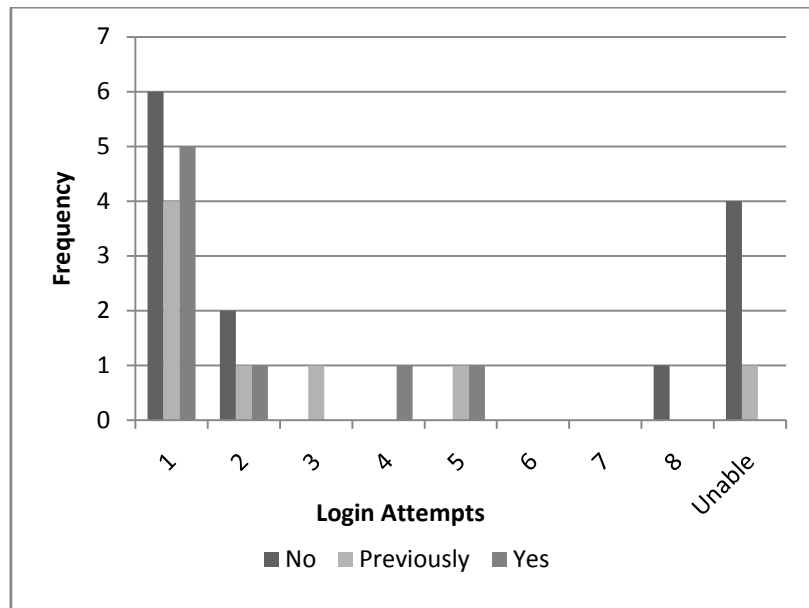


FIGURE 25 - LOGIN ATTEMPTS BY MUSICAL ABILITY

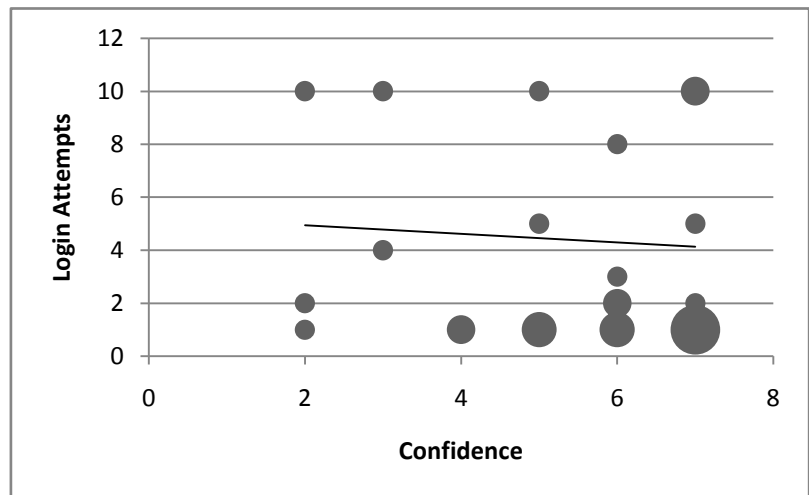


FIGURE 26 - CONFIDENCE PRIOR TO LOGIN VS. ATTEMPTS REQUIRED TO LOG IN

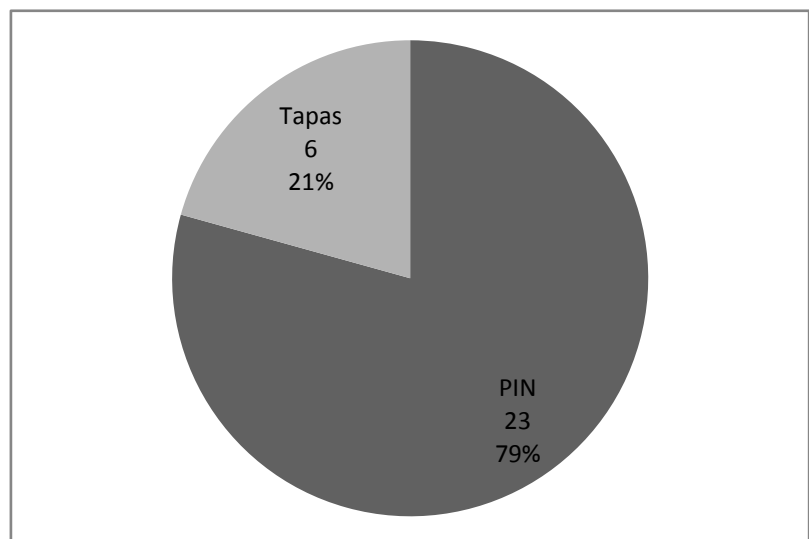


FIGURE 27 - PREFERENCE

demonstrated by the Pearson product-moment correlation coefficient of -0.15.

The penultimate set of questions asked users for their perception of *Tapas* and specifically whether would they prefer this or a traditional PIN/password system. A more in-depth security analysis is performed in section 9.2.

As Figure 27 shows, for most people *Tapas* was an unpopular choice compared to a PIN and password system. Participants were given space to further expand their reasons for their choice which raised a varied and interesting viewpoint. Many participants stated that they preferred the PIN/password system just because of the familiarity of it, with other citing security as another issue with specific concern to the ability of taps to be overheard. Another theme brought up was in regards to the memorability of taps, however opinions varied wildly. Some people felt that taps were easier to remember than passwords, even those who altogether preferred a traditional security system, however many made comments with the opposite effect, stating their belief that tap sequences were less memorable than a PIN or password. Many

individuals also made insightful comments contrasting the limited use *Tapas* to the ubiquity of passwords, claiming that as *Tapas* could only be used on touch screen devices, it is more likely to be forgotten than a password that is used everywhere. Another point raised is one not previously considered, which is if *Tapas* is used on multiple devices,

would different sequences be used on the different devices? Good password practice (Microsoft, 2006) recommends different passwords for different systems, but this is generally considered to be less memorable (Sasse, Brostoff, & Weirich, 2001).

Two participants also made comments related to the time to enter their tap sequences to unlock the system, feeling that it took longer to enter a tap sequence than a password. Although there is currently no empirical evidence to support this, it is a possible area of investigation for future research.

The final question asked was on perceived security of the system, rating from 1 (insecure) to 7 (secure) how secure participants felt the system was, as shown in Figure 28. When looking at this data it is important to bear in mind that 6 of the people involved with this experiment

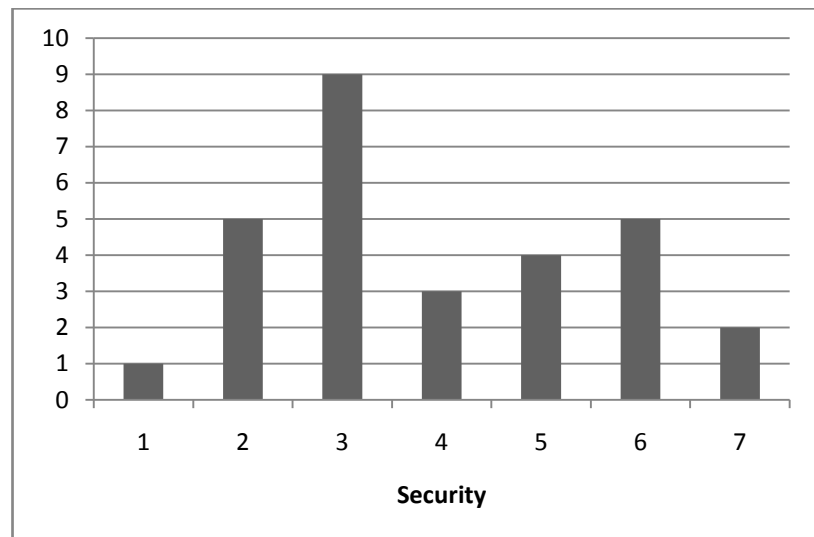


FIGURE 28 - PERCEPTIONS ON SECURITY

were also involved in the security experiment earlier which may bias their results. With a mode of 3 and a mean of 3.9 this shows that the participants perceived the system as generally insecure, but not trivially so.

This is reflected in the comments for the previous question, where users cited insecurity as a reason for not preferring *Tapas* over a PIN/password system, which the specific concern that people would be able to overhear their taps.

9.3.6 FURTHER COMMENTS

Using *Tapas* was not immediately clear to some people who did not have much experience with touch screen devices, however most understood the concept after one attempt. Additional verbal instructions were given to the users who needed help enrolling with the device.

Another common misconception was that the system was position based rather than purely based on the timings of the rhythm.

Although the first problem solves itself (hands on learning is often effective), the second issue is more of an issue, one that may possibly be resolved by better user education in the introduction to the system.

Another issue that arose was the lag at the end of the tap sequence. This was brought up as an issue in this and the other experiment as appearing to be “too long”. However, as the longest tap interval recorded was 2.5s it would not immediately seem that 3 seconds is too long. A more thorough investigation of this timeout needs to be performed for future improved versions of *Tapas*. A sensible solution may be to allow this timeout to be configurable.

The memorability of tap sequences, at least in infrequent usage appears to be at least as good as a PIN/password based system, however as we can see the system is far from perfect, specifically with regards to perceived security. As most users still prefer a PIN/password based system to *Tapas*, there appears to be a high barrier to user acceptance, especially when it comes to moving users out of their “comfort zone” of passwords/PINs.

10 CONCLUSION

When considering the results of this study it is necessary to consider the two aspects of *Tapas* - first, *Tapas* the abstract concept (that is, logging in using taps on a touch screen), and secondly *Tapas* the implementation (both the algorithm designed by Marriner (2007) and the concrete implementation as designed by Jolley (2008)).

The experiment clearly identified a number of problems with *Tapas* the implementation. Bugs in the software were discovered both in the planning stage and during the experiment. Some were fixed, yet others remain unresolved – specifically the installation and button locking bugs. Another fairly considerable flaw that requires further investigation is the ability to accidentally call emergency services. The short-term fix to simply disable the feature is not feasible in the long-term, as it does not satisfy the sensible requirement specified by Microsoft regarding emergency service phone calls.

The current implementation of *Tapas* clearly needs further work to address these superficial issues, however there are more fundamental issues discovered above that would still exist. Touch screen sensitivity is a problem that limits the usability of the system to certain devices. As touch screen technology develops, for example with multi-touch screens such as on the Apple iPhone or Google's Android platform, this issue may disappear and distinguish in importance, but at this current time touch screen technology does not appear to be consistently accurate which severely damages the usability of *Tapas*. A further possible mitigation for this would be an improved algorithm that does not depend on the detected taps being exactly the same length, yet still secure.

For *Tapas* as a concept, we can further separate out two aspects – security and usability. Section 9.2 shows disappointing results for the security of *Tapas*, however I believe that further work, specifically with an improved matching algorithm (the algorithm developed by Marriner (2007) was developed with successful matching as a main criteria with limited regard to security) and user education can improve on the results discovered here.

With regards to usability, the results shown are mixed. Tap sequences appear to be at least as memorable as passwords, yet the results for the more subjective measures appear to be not so good. This will be discussed further below.

Considering these points, the questions asked in the introduction can now be addressed.

10.1 HOW SECURE IS *TAPAS*?

As the results given in section 9.2, and the further analysis performed shows, *Tapas* is below the expected level of security for a mobile device, and unlocking the locked device was trivial.

For *Tapas* to succeed as a replacement for a PIN/password system, this needs to be corrected, either through further work to improve the algorithm, or by investigating what makes a strong tap sequence

10.2 HOW MEMORABLE ARE TAP SEQUENCES?

As the results given in section 9.3 shows, tap sequences are at least as memorable as passwords, and these results are positive. Although the sample size is small, the diary study results from section 9.1 confirm this, as no participants had issues with recalling their tap sequence.

10.3 HOW DOES *TAPAS* COMPARE TO THE TRADITIONAL PIN OR PASSWORD BASED AUTHENTICATION MECHANISMS?

This question is answered in part above. *Tapas* fares poorly with regards to security, but well with regards to memorability when compared to a PIN/password system.

In the measured aspects of effort, mental demand and frustration as part of the diary study, *Tapas* performed worse in every aspect. With the usability study *Tapas* appeared to score poor marks with regards to these aspects, however there is no good control data to verify this.

Another interesting point is that both users in the diary study rated *Tapas* worse in all aspects than the mean in the usability study. This suggests that *Tapas* possibly increases in frustration from heavy use, yet this may be a statistical anomaly of the small sample size of the diary study.

10.4 POTENTIAL FUTURE WORK

An obvious immediate area for future work is the investigation and fix of bugs discovered in the current implementation of *Tapas*, however this would not address the underlying issues. Further research needs to be done into the matching algorithm in order to improve, or replace, it with an algorithm that can improve security, yet is not so restrictive as to be frustrating to the user entering taps.

In addition to this work on *Tapas* the implementation; there are still many aspects of *Tapas* as a concept that are currently unknown. This study focussed mainly on the evaluation of the system compared to a traditional PIN or password system, however as discussed in section 6.2, there are other alternatives to password based systems. Graphical passwords are well-researched and implemented as an alternative to traditional password systems. An evaluation of *Tapas* against graphical password systems may provide an interesting topic for future research.

I believe that the poor results for security could be increased by user education of what makes a “strong” tap sequence. In section 9.2 I discussed some possible characteristics a strong password may have, however experimental analysis to determine if these characteristics are actually important has not been performed, and analysis on the affect on memorability this advice may have on users is also important.

As discussed in the literature review, security systems should be a “*minimal inconvenience to the users of the system*” (Morris & Thompson, 1979) and one possible metric for

inconvenience is the time taken to authenticate. Dhamija & Perrig (2000) shows that graphical passwords compare badly to PIN/password based systems in time taken to log in, and many users also reported they felt *Tapas* to take longer than a PIN/password system. Similarly, investigation into the tap timeout value, as discussed in section 9.3.6, may also be beneficial.

A final potential area of research comes from the growing technology of multi-touch devices, such as the Apple iPhone and the T-Mobile G1 which typically use fingers on the touch screen, rather than a stylus. Research could be performed into any difference between tapping a tune with a finger, rather than a stylus.

As section 9.3.3 and Clarke et al. (2002) showed, most people do not use security on their mobile device, with inconvenience cited as a major factor for not doing so. This larger problem of creating a convenient, yet secure, authentication mechanism is still one waiting to be solved, but, at least according to this study, *Tapas* does not appear to be the solution.

11 BIBLIOGRAPHY

BBC. (2008, November 2). *Previous cases of missing data* . Retrieved November 16, 2008, from BBC News: <http://news.bbc.co.uk/1/hi/uk/7449927.stm>

Blandford, A., Adams, A., Attfield, S., Buchanan, G., & Gow, J. (2008). *PRET A Reporter: evaluating Digital Libraries alone and in context*.

British Computing Society. (2006). *Code of Conduct*. British Computing Society.

Cairns, P., & Cox, A. (2008). *Research Methods for Human-Computer Interaction*. Cambridge University Press.

Canalys. (2008, November 6). *Global smart phone shipments rise 28%* . Retrieved March 13, 2009, from canalys.com: <http://www.canalys.com/pr/2008/r2008112.htm>

Clarke, N. L., Furnell, S. M., Rodwell, P. M., & Reynolds, P. L. (2002). Acceptance of Subscriber Authentication Methods For Mobile Telephony Devices. *Computers & Security* , 21 (3), 220-228.

Davis, D., Monrose, F., & Reiter, M. K. (2004). On user choice in Graphical Password Schemes. *Proceedings of the 13th USENIX Security Symposium*. San Diego.

Dhamija, R., & Perrig, A. (2000). Déjà Vu: A User Study Using Images for Authentication. *Proceedings of the 9th USENIX Security Symposium*.

Edwards, A. (2008). *ADNE03: Evaluation of the Tapas user verification system*. Retrieved from <http://www-users.cs.york.ac.uk/~alistair/projects2008.html>

Edwards, A. (2005). *ADNE09: Tap-pass*. Retrieved from <http://www-users.cs.york.ac.uk/~alistair/projects2005.html#adne09>

Elftmann, P. (2006). *Secure Alternatives to Password-based Authentication Mechanisms*. RWTH Aachen University, Laboratory for Dependable Distributed Systems, Aachen.

Eschenburg, F., Lylykangas, J., Krämer, N., Surakka, V., Troitzsch, H., Vuorinen, K., et al. (2005). User acceptance: the BioSec approach. *Biometric Technology Today* , 13 (7), 8-10.

Espiner, T. (2008, April 14). *Microsoft: Vista UAC designed to 'annoy users'*. Retrieved November 08, 2008, from ZDNet: <http://www.zdnet.com.au/news/software/soa/Microsoft-Vista-UAC-designed-to-annoy-users-/0,130061733,339288150,00.htm>

Gutmann, P., & Grigg, I. (2005). Security Usability. *IEEE Security and Privacy* , 3 (4), 56-58.

Harrington, V., & Mayhew, P. (2001). *Mobile phone theft*. Home Office Research, Development and Statistics Directorate.

Information Commissioner's Office. (2008, October 29). *Privacy watchdog calls on CEOs to take responsibility for data protection safeguards*. Retrieved November 22, 2008, from

Information Commissioner's Office:

http://www.ico.gov.uk/upload/documents/pressreleases/2008/data_breaches_29_october_2008.pdf

Jolley, M. (2008). *Implementation of a novel secure access mechanism for Windows Mobile*. University of York, Department of Computer Science.

Leyden, J. (2003, July 8). *PDA security slackers, the lot of you*. Retrieved November 16, 2008, from The Register:

http://www.theregister.co.uk/2003/07/08/pda_security_slackers_the_lot/

Limosani, R. (2008, October 16). *LAP Plugin freezes at startup on Windows Mobile 6 Classic*. Retrieved December 10, 2008, from Mobile Development "Support Side Story":

<http://blogs.msdn.com/raffael/archive/2008/10/06/lap-plugin-freezes-at-startup-on-windows-mobile-6-classic.aspx>

Marriner, C. (2007). *Investigation of an alternative security mechanism for pen based computer devices*. University of York, Department of Computer Science.

Microsoft. (2006, March 22). *Strong passwords: How to create and use them*. Retrieved March 5, 2009, from Microsoft Security:

<http://www.microsoft.com/protect/yourself/password/create.msp>

Morris, R., & Thompson, K. (1979). Password Security: A Case History. *Communications of the ACM*, 22 (11), 594-597.

Oates, J. (2008, October 29). *Thomas tells CEOs told to sort out data protection*. Retrieved November 16, 2008, from The Register:

http://www.theregister.co.uk/2008/10/29/ico_figures/

Passfaces. (2009). *Passfaces*. Retrieved February 24, 2009, from <http://www.passfaces.com/>

Perlow, J. (2008, October 06). *New in Labs: Stop sending mail you later regret*. Retrieved March 02, 2009, from The Official Gmail Blog: <http://gmailblog.blogspot.com/2008/10/new-in-labs-stop-sending-mail-you-later.html>

Peters, G., Anthony, C., & Schwartz, M. (2005). Song Search and Retrieval by Tapping. *Proceedings of the National Conference on Artificial Intelligence*, (pp. 1696-1697).

Petitcolas, F. (2008). *La cryptographie militaire (English translation of Kerckhoffs principles)*. Retrieved February 24, 2009, from <http://www.petitcolas.net/fabien/kerckhoffs/>

Rivest, R. L., Shamir, A., & Adleman, L. (1978). A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Communications of the ACM*, 21 (2), 120-126.

RSA Security. (n.d.). *RSA SecurID*. Retrieved February 24, 2009, from RSA: <http://www.rsa.com/node.aspx?id=1156>

- Saeveanee, H., & Bhatarakosol, P. (2008). User Authentication Using Combination of Behavioral Biometrics over the Touchpad Acting Like Touch Screen of Mobile Device. *International Conference on Computer and Electrical Engineering*, (pp. 82-86). Phuket, Thailand.
- Sanchez, R. (2008, March 13). *Counselling Service admits breach of trust after releasing over 300 emails*. Retrieved November 16, 2008, from Nouse.co.uk:
<http://www.nouse.co.uk/2008/03/13/counselling-service-admits-breach-of-trust-after-releasing-over-300-emails/>
- Sasse, M. A., Brostoff, S., & Weirich, D. (2001). Transforming the 'Weakest Link' — a Human/Computer Interaction Approach to Usable and Effective Security. *BT Technology Journal*, 19 (3), 122-131.
- Schedlbauer, M. J. (2007). *A Survey of Manual Input Devices*. University of Massachusetts, Department of Computer Science.
- Suo, X., Zhu, Y., & Owen, G. S. (2005). Graphical Passwords: A Survey. *Annual Computer Security Applications Conference*.
- The Register. (2004, March 26). *Securing the mobile enterprise*. Retrieved November 16, 2008, from The Register:
http://www.theregister.co.uk/2004/03/26/securing_the_mobile_enterprise/
- ThrottleLauncher. (2008). *Throttle Lock 0.4*. Retrieved February 29, 2009, from
<http://www.throttlelauncher.com/portal/downloads/34-public-releases/87-throttle-lock-04>
- Whitten, A., & Tygar, J. (1999). Why Johnny can't encrypt: a usability evaluation of PGP 5.0. *Proceedings of the 8th conference on USENIX Security Symposium*. 8, pp. 14-14. Washington, D.C. : USENIX Association.
- Yan, J., Blackwell, A., Anderson, R., & Grant, A. (2000). *The memorability and security of passwords - some empirical results*. University of Cambridge, Computer Laboratory.

Appendix A DISCLAIMER

Your participation in this experiment is entirely voluntary; there will be no remuneration for the time you spend evaluating it. All data gathered from the usability study will be treated in a confidential fashion: It will be archived in a secure location and will be interpreted only for purposes of this evaluation. When your data are reported or described, all identifying information will be removed. There are no known risks to participation in this experiment, and you may withdraw at any point. Please feel free to ask the researcher if you have any other questions; otherwise, if you are willing to participate, please sign this consent form and proceed with the experiment.

Date: _____

Signature: _____

Researcher's contact details:

Name: Chris Northwood
Address: 38 Starkey Crescent, YORK. YO31 0SY
Email: cjn503@york.ac.uk

Supervisor's contact details:

Name: Alistair Edwards
Address: Department of Computer Science, University of York, Heslington, YORK. Email:
alistair@cs.york.ac.uk

Appendix B EXTENDED USE EXPERIMENT PARTICIPANT INSTRUCTIONS

B 1 SOFTWARE DISCLAIMER

By installing the software you agree to the University's standard disclaimer for software as listed in the *Tapas* Readme. Please note the software has been tested to work, but you are still recommended to make a backup of your device before installing the software.

This Software is provided "as is", without warranty of any kind. To the maximum extent possible the Chris Northwood and the University of York disclaims all expressed or implied warranties as to its merchantability or fitness for any specific purpose.

B 2 PART ONE

First, please fill in the following questionnaire:

B 2.1 PERSONAL DATA

Please note this will be held in confidence and will only be used to correlate with your logbook:

Name: _____

Age: _____ Profession: _____

B 2.2 PREVIOUS EXPERIENCE

Please give a brief history of previous experience with touch screen devices:

Do you currently use a PIN or password with your PDA or phone?

Yes [] No []

If not, why not?

B 2.3 PASSWORD/PIN

Now, if you do not already do so, please enable PIN or password protection for your device and use it for the next 3 weeks.

In 3 weeks, please then move onto

Part Two of the experiment.

B 3 PART TWO

First, please can you complete this questionnaire about your past 3 weeks experience with the default password/PIN system.

B 3.1 PASSWORD QUESTIONNAIRE

Name: _____

For the following questions, please mark the box on the scale that represents your closest answer between the two scales

How mentally demanding did you find having to remember and enter your password/PIN?

Very demanding [] [] [] [] [] [] [] [] [] Undemanding

What was your success rate with entering your PIN/password?

Accurate every time [] [] [] [] [] [] [] [] [] I rarely managed to log in first time

How much effort did you feel that adding a PIN/password login required?

A lot of effort [] [] [] [] [] [] [] [] [] No effort

How frustrating did you find the PIN/password system?

Very frustrating [] [] [] [] [] [] [] [] [] Not frustrating at all

How secure do you feel the PIN/password system is?

Secure [] [] [] [] [] [] [] [] [] Insecure

Do you have any other comments regarding the password/PIN login system?

B 3.2 INTRODUCING TAPAS

Now, for the next 3 weeks you will be asked to evaluate an alternative to the PIN/password system called *Tapas*. *Tapas* works by remembering a series of taps made on your device's touch screen and then asking you to recall them to unlock the device. This can be considered similar to a 'secret knock'.

For the purposes of this experiment, *Tapas* also includes a backup password feature in order to allow you access in the event of you being able to remember or correctly reproduce your

taps. For this experiment, please only use the backup password feature as a measure of last resort.

During the experiment you are asked to keep a logbook of experiences using the system. Although you do not need to record every log in attempt, please do fill in the logbook at least once a day. In this logbook, you should record your experiences with the system, for example, whether or not you found it particularly difficult to use on that day, or if you had no issue remembering your taps. Examples of specific situations would be useful. Additionally, a note of environments the software is used in should also be made, especially if the environments are somewhat unordinary (for example, in a bar after some alcohol, on a moving bus or train, or in a particularly busy street).

If you have any questions or require any part of the experiment clarifying, please contact the researcher at cjn503@york.ac.uk, or on 07880862720.

The *Tapas* Readme contains instructions on how to install the *Tapas* software. This is included in the `readme.txt` file in the software bundle you were provided with and below for reference.

You are asked to keep this logbook and the software installed for at least 3 weeks. At the end of the 3 weeks, please move on to

Part Three.

B 3.3 TAPAS README

B 3.3.1 ABOUT

Tapas is a replacement authentication system for Windows Mobile devices. Instead of using a PIN or password to authenticate, *Tapas* uses a series of taps on the touch screen to log in.

B 3.3.2 SYSTEM REQUIREMENTS

Tapas requires any of the following Windows Mobile variants to run:

- Windows Mobile 5.0 for Pocket PC
- Windows Mobile 5.0 for Pocket PC Phone Edition
- Windows Mobile 6 Classic
- Windows Mobile 6 Professional

B 3.3.3 INSTALLING TAPAS

Tapas comes with two files: `Certs.cab` and `lap.CAB`. Both must be installed with `Certs.cab` being installed first.

Failure to install `Certs.cab` first will leave your Windows Mobile device unbootable and will require a hard reset.

In the file explorer, tap the `Certs.cab` file and okay the security warning.

Check that the installation destination is the device (not the storage card) and continue.

Once Certs.cab is installed, tap on lap.CAB to install that.

If you get a security warning, it means Certs.cab has not been installed. Go back and verify Certs.cab has been installed.

Once again, follow the prompts and install *Tapas* onto the device. If, whilst installing, you get an alert about the software being created for previous versions of Windows Mobile, this can be safely disregarded.

Once installation is completed, do a soft reset of the device, following the instructions of your particular device. You must now configure *Tapas* as specified in the *Configuring Tapas* section.

B 3.3.4 USING *TAPAS*

If *Tapas* is not configured, the *Tapas* verification system will be disabled until correctly configured, as specified using the *Configuring Tapas* section.

Upon rebooting the system, or after a period of inactivity defined in the Control Panel, the *Tapas* verification screen will be displayed. In this screen, you can tap your pattern in the central box to log in using taps, or use the top box to enter your backup password.

If you're using Windows Mobile 5.0 for Pocket PC Phone Edition, or Windows Mobile 6 Professional, then you will be able to make an emergency phone call (to the EU-wide emergency number 112) by tapping the 'Emergency Number Dial' button. Although this button appears on Windows Mobile 5.0 for Pocket PC and Windows Mobile 6 Classic systems, pressing it will have no effect.

B 3.3.5 CONFIGURING *TAPAS*

Before *Tapas* can be used it must first be configured.

To configure *Tapas*, you must go into the Control Panel and select the Lock option. Inside this screen there are multiple options.

To configure the taps, you must tap the 'Please enter a new tap pattern' button, and then enter your tap pattern successfully twice, following the instructions.

To configure the timeout before locking the device, you can use the 'Number of minutes until lock' text box.

To set a backup password, you must use the 'New Backup Password' text box and the 'Change Password' button.

To quit the *Tapas* Settings screen, you must click on the 'Apply Lock Out Time and Exit' button.

B 4 PART THREE

After evaluating the *Tapas* software, please submit your logbook back to Chris Northwood for analysis, and fill in this final questionnaire.

B 4.1 TAPAS QUESTIONNAIRE

Name: _____

For the following questions, please mark the box on the scale that represents your closest answer between the two scales

How mentally demanding did you find having to remember and enter your taps?

Very demanding [] [] [] [] [] [] [] [] [] Undemanding

What was your success rate with entering your taps?

Accurate every time [] [] [] [] [] [] [] [] [] I used the backup password feature frequently

If you had a low success rate in entering your taps, what was this due to?

Difficulty recalling the taps [] Taps not being consistently recognised []

How much effort did you feel that adding the taps login required?

A lot of effort [] [] [] [] [] [] [] [] [] No effort

How frustrating did you find the tap system?

Very frustrating [] [] [] [] [] [] [] [] [] Not frustrating at all

How complex was your tap sequence?

Complex [] [] [] [] [] [] [] [] [] Simple

How many taps did your sequence consist of? _____

If required to by your employer, which system would you prefer, and why?

PIN/password based [] *Tapas* []

How secure do you feel Tapas is?

If you were unable to log in, why do you think this was?

I had forgotten my tap sequence The system did not recognise my tap sequence

How mentally demanding did you find having to remember and enter your taps? Please tick the box in between the two indicated ends of the scale that matches your experience.

Very demanding Undemanding

How much effort did you feel that adding logging in using taps required? Please tick the box in between the two indicated ends of the scale that matches your experience.

A lot of effort No effort

How frustrating did you find Tapas? Please tick the box in between the two indicated ends of the scale that matches your experience.

Very frustrating Not frustrating at all

If required to by your employer, which system would you prefer?

PIN/password based Tapas

Why?

How secure do you feel Tapas is? Please tick the box in between the two indicated ends of the scale that matches your impression.

Secure Insecure

Appendix D USABILITY CONTROL EXPERIMENT PARTICIPANT INSTRUCTIONS

D 1 PART ONE

First, please can you fill in this questionnaire about past experiences with PDAs, phones and security.

D 1.1 EXPERIENCE QUESTIONNAIRE

Participant ID: _____

Age: _____ Course: _____

On the scale below, how would you rate your experience with touch screen devices? (e.g., touch screen phones, PDAs, iPod Touch, etc). Please tick the box in between the two indicated ends of the scale that matches your experience.

Used daily over an extended period of time [] [] [] [] [] [] [] [] [] [] No experience

Do you currently use a PIN or password with your PDA or phone?

Yes [] No [] I don't own a PDA or phone []

If not, why not?

If you do use a PIN or password, when are you prompted?

When the phone is turned on [] After a period of inactivity []

Do you play a musical instrument, or engage in another activity where remembering tunes or rhythms are important (e.g., Morse code)?

Yes [] No [] Not currently, but have in the past []

D 1.2 PASSWORD CAPTURE

Now, please think up a password. You can make it as simple or as complex as you'd like. You will now be asked to enter your password into the device, and once again to verify the password.

Once you have done this, please fill in the questionnaire on the next page.

If you were unable to log in, why do you think this was?

I had forgotten my password The system did not recognise my password
How mentally demanding did you find having to remember and enter your password? Please tick the box in between the two indicated ends of the scale that matches your experience.

Very demanding Undemanding

How much effort did you feel that adding logging in using a password required? Please tick the box in between the two indicated ends of the scale that matches your experience.

A lot of effort No effort

How frustrating did you find the password system? Please tick the box in between the two indicated ends of the scale that matches your experience.

Very frustrating Not frustrating at all

How secure do you feel the password system is? Please tick the box in between the two indicated ends of the scale that matches your impression.

Secure Insecure

Appendix E SECURITY EXPERIMENT PARTICIPANT INSTRUCTIONS

E 1 SETTING UP THE EXPERIMENT

E 1.1 USER INSTRUCTIONS

Please set up your tap profile, using the complexity given to you by the researcher as a guide. The researcher will help guide you through the tap process.

E 1.2 ATTACKER INSTRUCTIONS

It is important that you do not see or hear the setup stage – please leave the room until the setup is complete.

E 2 THE EXPERIMENT

In the main experiment, through a series of scenarios giving the attacker increased knowledge, the attacker will attempt to break into the PDA using the *Tapas* functionality. The attacker should not attempt to break in through the backup password, and the number of attempts for each attack will be limited to 20. Once the attacker has successfully logged in, the experiment will be over and no further stages will need to be completed.

E 2.1 DUMB ATTACK

The motivation for this experiment is to replicate attempting to break in to the PDA as if it had been found on a train, that is, the attacker has no prior knowledge in any form of the tap sequence.

E 2.1.1 USER INSTRUCTIONS

Please “lose your PDA on a train” and give the device to the attacker.

E 2.1.2 ATTACKER INSTRUCTIONS

Please attempt to break through the PDA by guessing tap sequences.

Was the attacker successfully able to log in?

Yes

No

If so, how many attempts did it take the attacker? _____

E 2.2 BUSY OVERHEARING ATTACK

The motivation for this experiment is to replicate attempting to break in to the PDA after hearing the user tap into the PDA in a realistic, busy environment, in the queue at a shop, but where the attacker can not see the user.

E 2.2.1 USER INSTRUCTIONS

Please attempt to log into the PDA in a suitably busy area with a decent amount of background noise (for example, Vanbrugh dining hall during lunch), ensuring that the attacker can not see your taps. Once you have logged in, please “lose your PDA” and give it to the attacker.

E 2.2.2 ATTACKER INSTRUCTIONS

Please stand near the attacker, but without being able to see the PDA or their arms. Listen for the taps. Once the user has logged in, you will be given the PDA and again asked to attempt to log in.

Was the attacker successfully able to log in?

Yes No

If so, how many attempts did it take the attacker? _____

E 2.3 QUIET OVERHEARING ATTACK

The motivation for this experiment is to replicate attempting to break in to the PDA after hearing the user tap into the PDA in a more quiet environment, perhaps in a library.

E 2.3.1 USER INSTRUCTIONS

Please attempt to log into the PDA in a reasonably quiet area (e.g., the Computer Science foyer) ensuring that the attacker can not see your taps. Once you have logged in, please “lose your PDA” and give it to the attacker.

E 2.3.2 ATTACKER INSTRUCTIONS

Please stand near the attacker, but without being able to see the PDA or their arms. Listen for the taps. Once the user has logged in, you will be given the PDA and again asked to attempt to log in.

Was the attacker successfully able to log in?

Yes No

If so, how many attempts did it take the attacker? _____

E 2.4 OBSERVATIONAL ATTACK

The motivation for this experiment is to replicate attempting to break in to the PDA after hearing and seeing the PDA user log in to the device, perhaps on an off-peak train journey where the two participants are sat opposite each other.

E 2.4.1 USER INSTRUCTIONS

Please attempt to log into the PDA in a reasonably quiet area (e.g., the Computer Science foyer) facing the attacker. Once you have logged in, please “lose your PDA” and give it to the attacker.

E 2.4.2 ATTACKER INSTRUCTIONS

Please stand facing the attacker, able to see the PDA. Listen for the taps and watch the sequence being entered. Once the user has logged in, you will be given the PDA and again asked to attempt to log in.

Was the attacker successfully able to log in?

Yes No

If so, how many attempts did it take the attacker? _____

E 2.5 SOCIAL ENGINEERING ATTACK

The motivation for this experiment is to replicate attempting to break in to the PDA after social engineering the user into giving away information about the device.

E 2.5.1 USER INSTRUCTIONS

In this case, the attacker is impersonating a member of your employer's IT staff, and you are helping them to log in to your PDA. You should describe your tap pattern to the attacker.

E 2.5.2 ATTACKER INSTRUCTIONS

The user will describe to you their tap pattern. Ask them for clarification where necessary, and then attempt to break into the PDA using the described tap pattern.

Was the attacker successfully able to log in?

Yes No

If so, how many attempts did it take the attacker? _____

Do you have any other comments to make?
