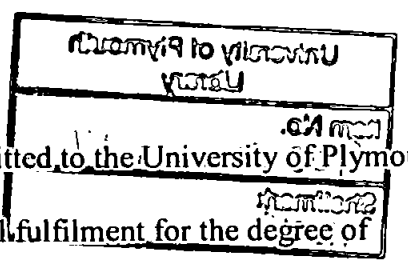
User Authentication and Supervision in Networked Systems

by

Paul Steven Dowland

B.Sc. (Hons), PGCE



A thesis submitted to the University of Plymouth

in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Computing, Communications and Electronics

Faculty of Technology

March 2004

User Authentication and Supervision in Networked Systems

Paul Steven Dowland
B.Sc. (Hons), PGCE

This thesis considers the problem of user authentication and supervision in networked systems. The issue of user authentication is one of on-going concern in modern IT systems with the increased use of computer systems to store and provide access to sensitive information resources. While the traditional username/password login combination can be used to protect access to resources (when used appropriately), users often compromise the security that these methods can provide. While alternative (and often more secure) systems are available, these alternatives usually require expensive hardware to be purchased and integrated into IT systems. Even if alternatives are available (and financially viable), they frequently require users to authenticate in an intrusive manner (e.g. forcing a user to use a biometric technique relying on fingerprint recognition). Assuming an acceptable form of authentication is available, this still does not address the problem of on-going confidence in the users' identity – i.e. once the user has logged in at the beginning of a session, there is usually no further confirmation of the users' identity until they logout or lock the session in which they are operating. Hence there is a significant requirement to not only improve login authentication but to also introduce the concept of continuous user supervision.

Before attempting to implement a solution to the problems outlined above, a range of currently available user authentication methods are identified and evaluated. This is followed by a survey conducted to evaluate user attitudes and opinions relating to login and continuous authentication. The results reinforce perceptions regarding the weaknesses of the traditional username/password combination, and suggest that alternative techniques can be acceptable. This provides justification for the work described in the latter part of the thesis.

A number of small-scale trials are conducted to investigate alternative authentication techniques, using ImagePIN's and associative/cognitive questions. While these techniques are of an intrusive nature, they offer potential improvements as either initial login authentication methods or, as a challenge during a session to confirm the identity of the logged-in user.

A potential solution to the problem of continuous user authentication is presented through the design and implementation of a system to monitor user activity throughout a logged-in session. The effectiveness of this system is evaluated through a series of trials investigating the use of keystroke analysis using digraph, trigraph and keyword-based metrics (with the latter two methods representing novel approaches to the analysis of keystroke data). The initial trials demonstrate the viability of these techniques, whereas later trials are used to demonstrate the potential for a composite approach. The final trial described in this thesis was conducted over a three-month period with 35 trial participants and resulted in over five million samples. Due to the scope, duration, and the volume of data collected, this trial provides a significant contribution to the domain, with the use of a composite analysis method representing entirely new work. The results of these trials show that the technique of keystroke analysis is one that can be effective for the majority of users. Finally, a prototype composite authentication and response system is presented, which demonstrates how transparent, non-intrusive, continuous user authentication can be achieved.

Contents

List of Figures

List of Tables

Glossary of Abbreviations

API — Application Programming Interface – a series of programming interfaces (functions) that provide access to the underlying operating system for an application.

C-I-A — Confidentiality, Integrity and Availability – the three core measures of system security. Commonly referred to as the CIA of security.

CSI — Computer Security Institute – a US organisation who, together with the FBI, conduct the Computer Crime and Security Survey.

DNS — Domain Name System – a distributed system storing information to allow the association of string domain names with numeric IP addresses.

DTI — Department of Trade and Industry – a UK government agency set up to assist trade and industry within the UK.

EER — Equal Error Rate – a metric used to evaluate authentication products. The EER rate is the point at which both FAR and FRR are equal.

FAR — False Acceptance Rate – a metric used to evaluate authentication products. The FAR rate indicates the proportion of impostors who would be falsely authenticated by the system.

FBI Federal Bureau of Investigation – a US agency set up to defend against terrorism, espionage and to enforce criminal law.

FRR False Rejection Rate – a metric used to evaluate authentication products. The FRR rate indicates the proportion of valid users who would be rejected by the system.

HMSO Her Majesty's Stationary Office – a UK government agency responsible for the supply of government information to the public and other interested parties.

ITSEC Information Technology Security Evaluation Criteria – security assessment criteria developed in Europe.

KPMG Klynveld Peat Marwick Goerdeler – a world-wide accountancy firm who conduct regular surveys covering the security issues of medium-large organisations.

RSA RSA Security Inc – a worldwide organisation specialising in IT security solutions

SMTP Simple Mail Transport Protocol – the protocol standard for email transmission across the Internet.

Acknowledgements

I would like to acknowledge the contributions of the following people:

- Dr Steven Furnell, my Director of Studies, for his continued support throughout the research programme, guidance in my research direction and his remarkable inability to fix even the most basic of IT related problems that caused no end of amusement for the research group. Steve provided inspiration and guidance throughout this project, without which I would not have completed this thesis.

- Professor Paul Reynolds, my supervisor, who provided advice and guidance throughout the project. I would also like to thank Paul for securing much needed funding from Orange through a number of research projects.

I would also like to thank specific members of the Network Research Group, namely Nathan Clarke and Harjit Singh for their support.

In addition to those who directly supported the research programme, I would like to thank the survey participants who contributed to the investigation presented in chapter 3. I would also like to thank all the trial participants who assisted in the collection of the keystroke data used in chapters 6, 7 and 8; without whom it would not have been possible to evaluate the techniques implemented. In particular, I would like to acknowledge the

assistance of the staff and researchers of the Network Research Group, the Department of Psychology and the staff of TMA Global and John Nicholls Builders Limited.

I would also like to thank and acknowledge my friends and family for their continued support over the last few years. In particular, to my parents for their encouragement throughout the research programme. Their support has ensured that I continued with my academic career from my first days at school through to the submission of this thesis.

Finally, I would like to thank my fiancée, Simone, who has provided encouragement and endless cups of tea during the final year of the research programme. Her continued support during the write-up stage of the PhD has ensured that this thesis was submitted on time. The final two months of the research programme have been a busy, stressful time for us both and, as such, she has been a tower of strength for me. It is not possible to express in words my appreciation. Thank you.

Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

Relevant conferences were regularly attended (at which work was frequently presented). Details of publications (conference and journal) can be found in the appendices.

Signed: Paul Dowland

Date: 11/06/04

Chapter 1

Introduction and Overview

1.1 Introduction

The last two decades have witnessed the use of computer technologies in a wide range of business and domestic scenarios. As such, there are few people in Western society whose lives are not affected by the use of Information Technology (IT). More recently, the explosive growth of both the Internet and the World Wide Web (WWW) has meant that IT has had yet further impacts upon our everyday lives. However, with society's widespread use of and, in some cases, reliance upon technology, significant opportunities now exist for both mischievous and malicious abuse via IT systems. While it is difficult, if not impossible, to prevent all forms of IT misuse, there are a number of methods of addressing the variety of risks that modern computer systems face. One of the key problems in IT misuse is authenticating the identity of end-users in order to both prove the identity of a valid user as well as identifying impostor activity and illicit use of computer resources.

In typical IT systems, protection against unauthorised user activities is usually provided via login authentication. Unfortunately, the majority of authentication schemes are based upon traditional password methods. The weaknesses of passwords are well-known (Jobusch and Oldehoeft, 1989), but their simplicity (from both user and developer perspectives) serves to ensure their continued use. A significant issue with passwords is that they typically provide a one-off authentication judgement at the beginning of a user session. Basing security measures on the identification of the user at the start of the session may prove unsatisfactory, as a user with lower user privileges or an outsider may gain access to the session and masquerade as the original, authenticated and, hence, authorised, user. The normal means of monitoring and identifying this is via audit trails, which maintain a record

of nominated security-relevant activities within the system and can be inspected at a later time in order to identify anomalies. The problem with this approach is that any detection of unauthorised activity will be retrospective, when significant damage may already have been done. If audit trails are not monitored, security breaches may potentially remain unnoticed for some time. What is, therefore, required is an automated, proactive means of detecting and responding to unauthorised access/activity. The research described in this thesis attempts to address a number of these issues.

1.2 Aims and objectives of the research

The objectives of this research programme can be categorised into two parts. Firstly, the range of methods for user authentication were identified and evaluated in order to determine currently available techniques and consider alternatives. This informed the second objective; namely the design of new methods and the development of associated practical experiments conducted to evaluate the alternative techniques, as well as considering user attitudes and opinions following exposure to the methods under trial.

The thesis begins by analysing the current methods of user authentication as well as considering the concept of user supervision. The user preference for secret-based authentication techniques is considered, together with the inherent weaknesses that these approaches present. User behaviour is considered by looking at the use (and misuse) of current authentication techniques, before a survey was conducted to determine the attitudes and perceptions of computer users to existing authentication techniques as well as introducing alternative methods.

The research continued by considering the practical implementation of a number of alternative user authentication mechanisms. In implementing these alternative approaches a key aim was to improve the level of security without causing the user to perform explicit actions (e.g. to provide authentication details) or to have to modify software (or their own behaviour). While the earlier trials considered improvements in the initial login authentication utilising secret-based methods, in order to achieve these objectives the later trials developed into transparent, continuous, real-time user supervision. The practical implementation of these techniques is evaluated via a series of prototype implementations.

The objectives for the research programme can be summarised as follows:

1. to investigate the current methods of user authentication/supervision within computer systems;

2. to assess user attitudes towards current authentication systems, as well as the acceptability of alternative authentication approaches and the concept of continuous user supervision;

3. to design new methods for improving user authentication and continuous user monitoring;

4. to evaluate the methods implemented and produce recommendations on necessary improvements;

5. to examine how the techniques can be applied in a wider, more comprehensive security system;

6. to recommend future development and propose further work relating to the research programme.

The objectives outlined above relate to the sequence of material presented in the following chapters in the thesis, the outline structure for which is presented in the next section.

1.3 Thesis Structure

This thesis presents the outcome of research conducted to investigate and evaluate alternative user authentication and supervision techniques in a modern PC environment. The thesis begins by considering the general area of user authentication before identifying potential approaches for further investigation. The chosen techniques are then evaluated in detail, and experiments conducted to evaluate the effectiveness of the approaches.

Chapter 2 presents a general overview of current issues relating to computer security and draws upon a number of examples of recent incidents to demonstrate the risks faced by computer systems. This is followed by a summary of the classifications of those responsible for computer abuse incidents in order to gain a better understanding of the different types of offenders. Having considered the background to computer abuse, the chapter then focuses upon the limitations of current user authentication techniques, considering the classic classifications of what the user *knows*, *has* and *is*. The chapter then

proceeds to discuss the problems with the current preferred technique, namely the password, before considering the alternatives. A significant section of the chapter is then dedicated to the consideration of a range of biometric based techniques, and finally summarising the relative merits.

Having evaluated a range of alternatives to the simple password, chapter 3 presents the results of a survey evaluating user attitudes and perceptions regarding a range of authentication techniques. This survey evaluated a range of issues in order to determine the acceptability of alternative techniques to the end users. The chapter begins by summarising current password practices (from a user perspective), before presenting the participants with a range of alternatives. The results of this survey informed the selection of techniques for further evaluation.

Following the selection of a subset of potential approaches, chapter 4 presents the results of two trials conducted to evaluate a range of secret-based authentication techniques. As these techniques are popular among users, this chapter focuses on methods that utilise secret knowledge. The chapter begins by presenting the technical implementation followed by the results of two trials aimed at evaluating both user recall of secret information and the long-term perceptions of user-friendliness and acceptability of these approaches.

Chapter 5 progresses beyond the use of secret-based techniques to present an overview of the concept of keystroke analysis. This chapter begins with a discussion of the range of metrics that can be obtained through keystroke analysis, and the ways in which these can be interpreted. Following this, the chapter identifies a range of ways to obtain the

identified metrics under the Windows operating system together with discussion of the potential integration of keystroke analysis directly into the Windows security model. The discussion then moves on to consider the role of filtering and post processing before summarising previous work in the area. Finally, the chapter identifies a series of new approaches to keystroke analysis, considering the use of trigraph, keyword and application-specific profiling.

Chapter 6 takes the proposed techniques from chapter 5 and describes an experimental implementation of keystroke dynamics aimed at evaluating the methods previously identified. The chapter begins with a detailed description of the software developed for the experiment and the features of the underlying operating system that were utilised. This is followed by a discussion of the analysis of the data, before considering the results of the trial. The chapter concludes with a brief discussion of a novel use of data mining for keystroke analysis, before presenting an initial experiment of the use of an application-specific approach.

Following the small experiment described in chapter 6, chapter 7 presents the results of a long-term experiment evaluating digraph, trigraph and keyword-based keystroke analysis. The chapter begins by presenting the technical implementation of the keylogging software and utility programs used to filter the raw data, generate the profiles and compare the samples. Following this, the chapter presents the results for each of the separate metrics.

Chapter 8 extends the work presented in chapter 7, and considers alternative approaches and concepts. This chapter presents the results of a composite approach that combines the three metrics evaluated in the previous chapter, before introducing a composite

authentication and response system based upon the software described in chapters 4, 6, 7 and 8. This is presented as a prototype implementation to demonstrate the effectiveness of these techniques in an operational context.

Finally, chapter 9 presents the conclusions drawn from the research conducted and presented in the thesis. The key achievements are emphasised, together with the limitations on the research programme. This chapter also suggests a number of potential extensions to the research described in the earlier chapters, and identifies a number of possible improvements to the experiments conducted.

The thesis also includes a number of appendices containing additional information to support the discussion presented in the main chapters. In addition a CD is provided that includes source code from the experiments described in chapters 6, 7 and 8, as well as the raw results (too large to include in the body of the thesis). Finally, a number of published papers arising from the project are included, as well as a list of papers produced during the same period that are less directly related to the PhD research.

Chapter 2

Evaluation of Current Authentication Measures

2.1 Introduction

The last two decades have witnessed the integration of personal computer technologies into a wide range of business and domestic scenarios. As such, there are few people in Western society whose lives are not affected in some way by the use of IT. More recently, the growth of the Internet and the WWW has meant that IT has had yet further impacts upon our everyday lives. This is set to increase over the coming years with the increased availability of new technologies – especially with an increasingly mobile and technically minded public. However, with society's widespread use of, and, in some cases reliance upon, technology, significant opportunities now exist for both mischievous and malicious abuse of IT systems.

Over the past 20 years, the UK Audit Commission has conducted a series of surveys to assess the scale of crime and abuse within the IT community. The results of these surveys show a significant upward trend in overall crime levels during this period (Table 2.1). The audit commission surveys are not alone in these findings with surveys conducted by the CSI in the USA showing similar rises in computer crime (CSI, 2003).

It should be noted that the categorisation of various types of computer crime cases has varied slightly over the twenty years of the Audit Commission surveys. In particular, the category of 'viruses' was not included until 1990 and cases of offences involving pornographic material were not reported until 2001 (the latest survey results currently available). In the 1984-1990 surveys the definition of the category "theft" was quite broad and covered the use of unlicensed/illicit software, private work and theft (of equipment or

data) while the category "hacking" covered hacking, sabotage and invasion of privacy. For the comparisons of later surveys, these categories have been maintained for consistency.

	Fraud	Viruses	Theft	Hacking	Pornography	Other	Total
1984	60	-	17	-	-	-	77
1987	61	-	22	35	-	-	118
1990	73	54	27	26	-	-	180
1994	108	261	121	47	-	-	537
1998	67	247	88	56	-	52	510
2001	50	200	19	44	193	119	625

Source: UK Audit Commission 1984-2001

Table 2.1 - Reported incidents of computer crime and abuse

It is clear that over the last 20 years there has been a significant increase in the number of reported incidents. A clear factor influencing this increase is the explosion in virus incidents that can be observed from the 1990s and the subsequent increase in cases relating to the access or distribution of pornographic materials in the most recent survey. It is worth noting that, in the latest results, 'pornography', 'use of unlicensed software' and 'private use of company IT resources' are the only categories of abuse in which the reported incidents have risen (in both real terms and as a percentage of incidents reported) when compared to the previous surveys (193 cases relating to pornographic material, 35 cases relating to unlicensed software and 72 cases relating to private use of company resources).

These findings are echoed in the surveys conducted in the USA by the CSI/FBI through their computer crime and security surveys. The most recent report (CSI, 2003) showed 56% of organisations suffering from unauthorised use of corporate computer systems (slightly down on the 60% reported in the previous year). Of the 490 organisations

surveyed, 45% reported unauthorised access to computer resources by insiders and 80% reporting insider abuse of network access. This shows a high level of computer misuse appearing from inside an organisation – misuse which may be prevented by improved user authentication and/or monitoring of user actions.

Over the last few years there have been numerous incidents that have been reported in the media that have reaffirmed the susceptibility of IT systems to abuse. Examples of these incidents include the MyDoom virus (and its variants) and the mass defacement challenge (a number of recent incidents are outlined in Table 2.2).

Incident	Details
MyDoom worm	January 2004 Mass mailing and peer-peer file-sharing worm • contained an SMTP server to send emails (spam/replication) • contained a backdoor to allow IP spoofing or remote code execution • contained a Denial of Service payload targeting the SCO (and Microsoft in the MyDoom.B variant) web sites At the peak of the infection it was estimated that 20-30% of all worldwide email traffic was generated by MyDoom. (F-Secure, 2004)
Mass defacement	July 2003 Global hacking competition conducted to deface web sites. Target of 6,000 web sites hacked in 24 hours There is no confirmed count of defacements but the figures are estimated to be in the region of a few thousand. (ZoneH, 2003)
Global DDoS	June 2003 Global distributed denial of service (DDoS) attack. The web sites of Clickbank and Spamcop were victims of a global DDoS attack suffering over 1000 hits per second. (Schultz, 2003a)

Table 2.2 - Example IT security incidents

Before we can look at ways of improving IT security, it is first necessary to understand the issues that affect computer systems. There are four main issues relating to computer

security (ITSEC, 1991) and overall IT system security relies on the preservation of all of these factors.

- Confidentiality

  The term confidential is indicative of a level of secrecy and is clearly understood. This generally refers to the prevention of unauthorised disclosure of information and has a familiar comparison with a "need to know", military-style, security model. The consequences of a breach of confidentiality are usually dependent on the context of the breach. From an organisational perspective, if confidential content is accessible to a third party the consequences are probably more significant than if someone inside the organisation accessed the same material.

- Integrity

  Integrity of data can be more difficult to analogise, as it relates to the consistency, completeness and correctness of data. An impostor or masquerador could potentially make minor changes to data files or programs (e.g. to siphon off small amounts of money *lost* in account transactions through rounding errors) that would not necessarily be identified immediately. Not all breaches of data integrity are malicious, however, even accidental modification/deletion of data can cause serious problems (e.g. a user mistakenly deleting an important file). Viruses represent one of the commonest threats to IT system integrity with payloads that can modify/delete files.

- Availability

  With an increased dependence on IT systems, their availability (or uptime), is increasingly important. Users expect systems to be available whenever and wherever they need them. While this used to concentrate on defence against mechanical/logical failure, it is now equally (if not more) important to consider the threat from malicious activity which can render a system inaccessible. Known as a Denial of Service (DoS), this form of malicious activity was used to significant effect against several major Internet sites (Yahoo, Amazon and eBay to name a few) in the worldwide attacks in early February 2000, effectively holding systems hostage (McCullagh and Arent, 2000). A recent survey (CSI, 2003) identified DoS attacks as the second most expensive form of incident affecting respondent organisations and costing industry in excess of $65m.

- Accountability

  Whilst not usually considered a part of the C-I-A trio, accountability is vitally important to allow actions and intrusions to be tracked. Without some form of accountability it is impossible to directly attribute an action to an individual or to be able to prove that an individual did *not* perform a specific action (i.e. through authentication we should be able to hold an individual accountable for their actions or, alternatively to be able to defend an individual or organisation). Accountability is usually achieved through historical logs, however this only allows action to be taken after the event, therefore some form of interactive monitoring is needed to audit (and respond to) user actions

in real-time. Action in response to illegitimate activity needs to be taken proactively rather than reactively.

2.2 Categorising system intrusions and misuse

Whilst the previous section discussed the four specific issues relating to IT security, it is also necessary to understand the sources from which computer abuse is likely to be encountered. This is important, as although it is necessary to appreciate the differing forms of computer abuse, it is also important to consider the nature of the person undertaking the misuse. By examining the perpetrators of computer crimes, it may be possible to evaluate the motives and hence to reduce the risks faced by IT systems. Forms of human abuse have already been comprehensively categorised by Anderson (1980), and are described in Table 2.3.

Abuser Type	Description
External Penetrators	Outsiders attempting or gaining unauthorised access to the system. E.g. a hacker trying to download the password file(s) from a server or a rival company trying to access the sales database.
Internal Penetrators	Authorised users of the system who access data, resources or programs to which they are not entitled. Sub-categorised into: • *Masqueraders* Users who operate under the identity of another user. E.g. someone using another users' PC whilst they are absent from their terminal, or, someone using another's username/password. • *Clandestine users* Users who evade access controls and auditing. E.g. someone disabling security features/auditing etc.
Misfeasors	Users who are authorised to use the system and resources accessed, but misuse their privileges. E.g. someone in the payroll department accessing a colleague's records or misappropriating funds.

Table 2.3 - Categories of system abuser

These groupings are considered appropriate for describing the different types of user-related abuse within an intrusion-monitoring framework and will, therefore, be adopted for the remainder of the discussion. Whilst it is also possible to develop a deeper profile of potential intruders, by considering factors such as the common motivations behind abuse (e.g. money, ideology, egotism etc.), these are not explored here as knowledge of them would not contribute to the process of detection. However, this subject was discussed in a paper published in 1999 (Furnell et al) in which the motivations, ethics and perceptions of computer criminals are explored and the role of the media is considered.

It should be noted that Anderson's categorisations do not take into account any of the types of abuse that may result from software activity (e.g. viruses, Trojan Horses etc.). This is understandable given that the analysis was made in 1980 before such incidents had become commonplace. However, there has been a significant increase in such attacks over the last decade and evidence suggests that viruses are one of the major causes of security breaches in both networked and standalone PC systems (CSI, 2003).

2.2.1 Internal Penetrators and Misfeasors

At the highest level, intrusions or misuse will be the result of actions by authorised users or processes, which operate on one or more targets that may include data (files), system devices and other users or processes. It has been shown that the most significant source of computer system abuse is from within the organisation (Dinnie, 1999; ISBS, 2000). Therefore, if we can secure systems against internal abusers, we will be targeting the most likely perpetrator. Unfortunately, internal abusers are quite likely to be authorised to use

the systems they are abusing, hence reliance on the basic username/password will only serve to, retrospectively, attribute blame once the abuse has been detected. In order to detect and prevent abuse in real time, it is necessary to introduce an element of user supervision.

The purpose of introducing supervision will be two-fold:

- to ensure that systems are only accessed by authorised users;
- to ensure that systems are only used for authorised purposes.

By introducing user supervision it is possible to monitor, in real time, the actions of individual users at a variety of levels. Monitoring could take place at operating system level; monitoring key files, directories or resources (e.g. printers and CD writers) or individual applications could be monitored (e.g. a database application could be monitored for export of data or deliberate deletion of data). Alternatively, a higher-level approach could be taken. A supervision application could be loaded into a system that would monitor specific characteristics (akin to a lifeguard watching a swimming pool from a high vantage point). When an uncharacteristic pattern is monitored (e.g. a person panicking in the deep end of the pool) an appropriate response can be initiated. This could potentially identify a user acting in an unusual or unexpected manner (e.g. conducting tasks outside of their defined role) or could identify a user account being used by an impostor.

User actions can be categorised as being either legitimate or illegitimate. However, it is useful if a more detailed breakdown than this can be derived for the different potential

classes of illegitimate activity. For example, all of the following scenarios represent types of illegitimate activity that should be monitored:

- an illegitimate action that is still within the normal authorisation of a valid user (i.e. abuse of privileges);
- an action by a valid user which is outside the normal limits of authorisation;
- any action by an unauthorised user.

2.2.2 External Penetrators

Whilst sources of abuse within the organisation are the most likely form of abuse, they are usually the least damaging to a company and its reputation. The reason for this is that being internal attacks, they are usually dealt with by the company concerned and rarely have any external impact. External penetrators are quite different. Due to the intrusive nature of such attacks, they are more likely to have a visible, external impact. Examples of this could include company blackmail, attacks against web sites and financial loss from e-commerce systems. Although most of these attacks can be described as prank or hoax attacks, they will often have a significantly negative effect as they undermine the company's public image. A particularly apt example of this was the hack of the RSA security web site in February 2000 (2600, 2000). (It should be noted that the server on which the RSA web site was held was not actually hacked. Instead, the perpetrator, a hacker named Coolio, was able to modify the DNS records to redirect traffic destined for the RSA web site to a different server.) The consequence of this particular hack was the rather unfortunate headlines indicating that a company purporting to provide high security

solutions was itself hacked – the media choosing to overlook the nature of the attack that was actually outside the control of RSA.

2.3 Limitations of current technology

Having identified and categorised the main sources of computer abuse, this section considers the current methods of protecting IT systems and identifies their shortcomings. There are three main approaches to user authentication, something the user knows, something the user has and something the user is (Wood, 1977). Table 2.4 summarises these 3 main forms of authentication and provides examples as well as indicating the known weaknesses of each method. These methods will all be discussed in more detail in the next section(s).

Something the user...	**knows**	**has**	**is**
Example	• Password • PIN • Other secret knowledge (e.g. personal information)	• Magnetic card • Smart card • Proximity device (e.g. RF tag)	• Physiological (e.g. fingerprint, retinal scan) • Behaviour e.g. keystroke analysis
Weaknesses	• Can be guessed • Often chosen inappropriately • Vulnerable to social engineering or shoulder surfing	• Can be stolen • Potentially replicated • Dedicated hardware needed (e.g. magnetic card reader).	• Usually requires dedicated hardware to profile and authenticate • If breached, users cannot replace their identifying characteristic (e.g. cannot grow a new fingerprint).
Strengths	• Easy to implement • Easy for users	• Can be combined with knowledge (e.g. pin plus swipe card)	• Depending on measure is highly unique to the individual. • Cannot be forgotten by user

Table 2.4 - Examples of authentication measures

The most commonly used means of authentication in IT systems is the password. Passwords are conceptually simple for designers and users and can provide effective

protection if used correctly. However, their protection is often compromised by users. Typical problems include forgetting passwords, writing them down, telling other people your password and selecting easily guessed passwords.

Several studies have been carried out over the last 30 years looking at the ease with which passwords can be determined. In 1979, 86% of the 3829 passwords gathered, could be guessed by a PC in less than 1 week (Morris & Thompson, 1979). This was repeated in 1990 by Klein and in 1992 by Spafford (Klein, 1990; Spafford, 1992). Whilst the results from these later experiments showed that password selection had improved (only 21% could be guessed in 1 week), so have the tools that can be used to guess passwords. In 1998, L0pht Heavy Industries developed L0phtCrack (Heskett, 1998) - later renamed and re-released as LC4 when bought out by @stake (2004). @Stake is a utility that allows Windows NT Server Message Block (SMB) password packets to be captured during network authentication sessions (although the product is advertised as a password auditing and recovery tool). This utility not only allows the encrypted passwords to be captured directly off the network, it can also perform a dictionary and brute force attack against the encrypted passwords. Similar utilities are also available for other operating systems (notably CRACK which runs under a number of flavours of UNIX) (Cherry et al, 1992). More recently, researchers from the Swiss security organisation Lasec developed "Advanced Instant NT Password Cracker" (Schultz, 2003b). This prototype application (only available as a demo from the web site) claims to be able to crack Windows user account passwords in hours rather than days.

There are a number of measures that can be taken to improve password security for example:

- Non-Dictionary Words

  Forcing users to select non-dictionary passwords prevents the use of dictionary based attacks. A dictionary attack can identify a password in less than 20 minutes even on word lists with up to one million words. The only way to identify non-dictionary passwords is by using a brute-force approach (testing every combination of characters for every length of password).

- Passwords with mixed case/symbols

  Including both upper/lower case and symbols (!£$% etc.) in passwords requires any attack to use a brute force method. The use of these extra characters exponentially increases the time taken to determine the password by increasing the key-space that must be searched through.

Password Characters	**4 Character Password**	**8 Character Password**
a-z	456,976 (26^4)	208,827,064,576 (26^8)
a-z, A-Z	7,311,616(52^4)	53,459,728,531,456 (52^8)
a-z, A-Z, 0-9	14,776,336 (62^4)	218,340,105,584,896 (62^8)
a-z, A-Z, 0-9, and special characters e.g. !, “, £, $, %, ^, @, # etc. [Approx 82 characters]	45,212,176 (82^4)	2,044,140,858,654,976 (82^8)

Table 2.5 - Character combinations for passwords

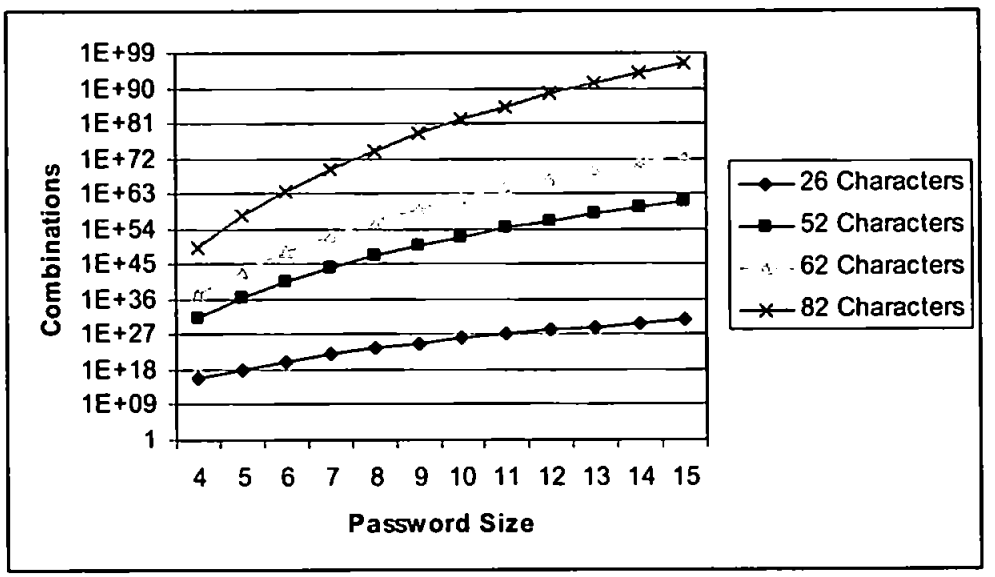


Figure 2.1 - Exponential increase in password complexity

- Password Ageing

  Should an intruder obtain a valid username/password combination most systems will allow the intruder to continue to access the system until the intrusion is noticed. If a password ageing policy is in place users can be forced to change their passwords regularly, thus forcing the intruder to identify the new password.

Whilst these suggestions will help to make a password-based system more resilient to an intruder they are by no means secure. A determined intruder can utilise password-cracking utilities to determine even the most random password in a matter of weeks. With the advent of more powerful processors, intruders will be able to crack passwords in a more realistic time – a matter of days for some PCs. A more worrying issue is the failure by some users to protect their own passwords. None of the measures outlined above help to address the human-element of the problem. This was demonstrated in a recent survey conducted at the InfoSec conference at Olympia (IT Week, 2003). The survey revealed

that of the 150 people questioned, more than 100 were prepared to give their passwords to a complete stranger – tested most effectively by simply asking people what their password was. While not very scientific, the survey did highlight a fundamental problem with IT security – we still rely on users! To counter these problems with password based systems, we need to consider alternative approaches to user authentication. Another factor suggesting the use of alternatives is that any of the above means of strengthening passwords ultimately reduce their simplicity and friendliness for users.

2.4 Methods of Improving Authentication

To meet the demand for alternative methods of user identification and authentication, there are a growing number of companies offering both hardware and software based products. Solutions range from basic (and cheap) keystroke analysis software (BioPassword, 2004) to sophisticated iris and retinal scanning devices (Secure Computing, 1995; Sherman, 1992; Cope, 1990). Many of these devices can be incorporated into desktop PC's and can usually be configured into a network-wide security policy. However, most of these products utilise proprietary technology in both the hardware and software making it impossible (or at the very least difficult) to integrate products purchased from a variety of manufacturers. There is also considerable variability in the extent to which the software provided can integrate with the operating system's security model. Some products simply provide an additional layer of security, requiring the user to authenticate themselves in addition to providing a valid username/password pair, whilst others provide replacement logon interfaces and fully integrate into the security model of the OS. However, some work has been done by OS vendors to provide a standardised application programming

interface (API) to allow product vendors to integrate their identification/authentication and monitoring devices/products into the OS security kernel (BioAPI, 2004; Microsoft, 2000).

As indicated earlier in Table 2.4, the three main methods of user authentication have a number of weaknesses. Passwords can be written down, forgotten and shared; tokens or cards can be stolen, copied or lost whereas the alternative, biometrics provides a seemingly near-perfect solution. If the problems of cost, user acceptance and integration can be overcome (not insignificant problems on their own) the use of biometrics could be a solution to the problems of user authentication. Biometric characteristics cannot be (easily) lost, stolen or duplicated; are usually stored in a non-reproducible manner and offer a high level of authentication confidence (depending on the methods chosen).

Methods of biometric-based user authentication fall into two distinct categories, specifically physiological and behavioural characteristics.

- Physiological characteristics represent those traits that describe who we are based on physical attributes, for example fingerprints, hand geometry, retinal and iris scanning. These characteristics usually require additional equipment to be connected externally to the computer to provide the necessary data capture.

- Behavioural characteristics cover attributes such as typing style, voice pattern and signature recognition. Most behavioural characteristics can be acquired without the need for additional equipment (keyboard & mouse)

however others do require specialised hardware solutions (signature recognition).

Most biometric devices offer a compromise between high security/low user acceptance and low security/high user acceptance. This trade-off can be measured as the False Acceptance Rate (FAR) and False Rejection Rate (FRR) of the devices. Figure 2.2 shows the relationship between the FAR and FRR rates and can be seen to be mutually exclusive.

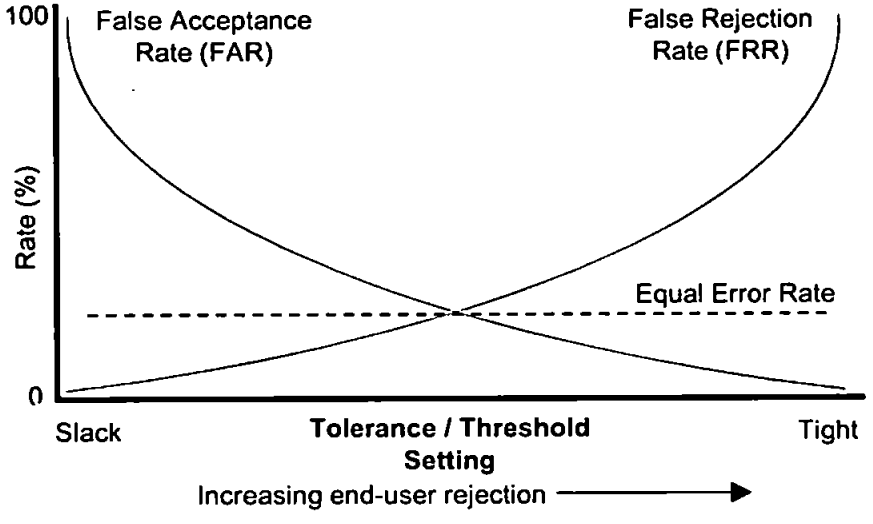


Figure 2.2 - FAR/FRR Graph

FAR represents the proportion of invalid users who will be successfully authenticated by a system (i.e. falsely accepting an unauthorised user). FRR represents the proportion of valid users who will be identified as illegitimate by the system (i.e. incorrectly rejecting an authorised user). The error rates of an authentication system are assessed using a fixed population of test users and usually involves cross testing of users against the profiles of all other users in the trial. This allows consistent testing of error rates, but results in rates based on a subset of the population – i.e. it does not consider a true intruder (an individual unknown to the system). Despite this flaw, the use of FAR/FRR rates is widely used as a

metric for authentication techniques and, as such, is used within this thesis to allow appropriate comparisons to be made to existing techniques. It has so far proved impossible to achieve a system where the FAR and FRR rates have both been reduced to 0. Most systems, instead, select an appropriate level at which inconvenience to the legitimate user, through denial of access (rejected logins), is acceptable, without allowing too many intruders to gain unauthorised access. All systems have an Equal Error Rate (EER), the point at which the FAR and FRR rates are equal. Whilst this rate represents the theoretical "best-fit" for security measures, it is rarely ideal in a secure environment where a preference for either low FAR or FRR exists (Cope, 1990).

2.4.1 Continuous supervision

While the techniques outlined in the previous section are usually used to identify an individual at the start of a session (in a similar way to passwords) this does not address the problem of on-going authentication. In most systems, once a user has identified themselves to the operating system/application, it is usually impossible to determine if the active user is authorised. For example, a typical user session commences when they arrive in the morning and will end when they leave in the afternoon/evening. Throughout the day, the user will leave their terminal (often without any form of locking or password protected screen-saver) for breaks, lunch, meetings etc. During this period of absence, the system, without additional controls, will continue to assume that the user is still the original authenticated user – who could have been replaced at any stage by an impostor or masquerador. Whilst this may be adequate for some users, in environments where a higher level of security is required, some form of ongoing user authentication may be desirable.

To remedy this, authentication can be extended beyond the login stage and effectively supervise the user throughout a logged-in session. It should be noted that not all of the measures described above (and in the following sections) are suitable for continuous user monitoring. For example, signature recognition would be entirely unsuitable, as the user would be forced to stop his/her work to sign their name before proceeding. Instead, measures such as continuous keystroke analysis, combined with strong one-off authentication measures may prove effective.

2.4.2 Finger scan

Fingerprint scanning requires a hardware-based device linked to a PC or access control system and is one of the most established forms of biometric based user authentication. The use of fingerprints as a means of identification has been established over many years through their use in criminology. Whilst this measure is a good discriminator of identity, there are social issues to consider. It has been suggested (IBG, 1999; Sherman, 1992) that the use of fingerprints for identification has criminal overtones and that there may be some resistance by users to use such systems. Despite this, there have been successful trials with organisations like MasterCard and the US Department of Defence (Identix, 2004) and these devices are now available at relatively low cost from a range of suppliers.

2.4.3 Hand geometry

Hand geometry is, like fingerprints, a good discriminator of identity and also requires some form of dedicated hardware. It is one of the least used of all biometric measures, although

there are a number of establishments that have adopted it (Recogsys, 2003). Although there is little resistance to this form of user identification, it still requires a specific action on the part of the user, i.e. the user must place their hand on the scanner, there are also significant cost factors to consider for any deployment. While this seems to be accepted for access control for buildings, offices etc., it is unlikely to be acceptable on individual PC's or other devices.

2.4.4 Retina scan

Retinal scanning projects a laser light onto the user's retina and identification is based on the retinal vascular pattern i.e. the vein pattern on the retinal surface. There is only one commercially available system (Eyedentify from Access Controls International), their device claims to match a users retinal scan from a user base of 300,000 in less than 15 seconds (ACI, 2004). The accuracy of this device is claimed to be very high, this allows a user to simply present their eye to a device without requiring a card, pin or username, i.e. one to many identity matching. The main drawback to this device is cost and acceptability. The cost factor has restricted the use of such identification methods primarily to the military sectors, where security is of greater importance than cost. The acceptability factor is related to public perception of the safety of lasers. There are also some technical problems when identifying user's wearing glasses (especially if darkened) (ACI, 2004).

2.4.5 Iris scan

Iris scanning is also a hardware-based technique, using high-quality cameras and natural light to detect and identify the unique patterns of a users' iris. Iris scanning has several

advantages over retinal scanning, notably, cost and acceptability. The entry-cost for iris scanning is significantly cheaper than retinal scanning as it relies on simpler cameras e.g. Panasonic's Authenticam (Panasonic, 2004). As these cameras become more widespread (even potentially built into monitors), the use of iris scanning as an acceptable method of user identification may increase. Secondly, because the iris can be identified without the use of laser light, it may prove more acceptable from a safety viewpoint. Recent improvements have resulted in much more reliable products that are able to cope with users who wear contact lenses or glasses (Khew, 2002)

2.4.6 Facial geometry

Facial recognition is another economically viable method currently available. Facial recognition can be achieved using standard cameras e.g. a webcam. The video stream is then encoded and the face located and identified. The face can be authenticated using a range of methods but can be represented by two techniques. Feature extraction and recognition identifies the location of key features of the face (e.g. eyes, ear, nose and mouth) and calculates a geometrical relationship between the key features. This relationship is then stored as a mathematical model of the face. This method has the advantage that it relies on relative positioning of facial features and is therefore more tolerant of variations in head position. An alternative is to use a holistic approach where the face is evaluated as a whole and use neural networks or statistical techniques to evaluate a face-print and form a profile. As with iris recognition, there are some technical problems with user's appearance, e.g. wearing glasses, changing hairstyle etc., such deviations from the stored images may require the user's profile to be updated (Woodward et al, 2003).

2.4.7 Key stroke dynamics

Keystroke dynamics can provide an additional level of security for the traditional username/password by analysing the typing style during logon, or, can be used to identify the user based on more dynamic freestyle text entry e.g. allowing the user to type a known string or phrase. This method is of particular interest as it is one of the few totally non-intrusive authentication mechanisms available as the user does not have to explicitly *do* anything to be authenticated by this method (i.e. the user would perform a normal login, or, in a supervised environment, would continue with routine tasks whilst being monitored in the background). There is only one available product that utilises this method, namely BioPassword (2004). This product is able to supplement login authentication with keystroke analysis – however, this is only at login and there is no further improvement in security beyond the initial authentication.

2.4.8 Mouse Dynamics

Mouse dynamics is an extension of the keystroke analysis concept. The principle of mouse dynamics is that, like our typing, our mouse movements may be a characteristic trait that can be monitored and compared against a historic profile. Like the keystroke dynamics technique described previously, this method could also be transparent (depending on implementation) and could act as a continuous authentication mechanism. There are currently no products available using this technique.

2.4.9 Speaker verification

Speaker verification requires nothing more than a basic microphone connected to the PC's sound card and the appropriate software. Speaker verification is a distinct area from that of voice interpretation that is commonly found in commercially available dictation software often (mistakenly) referred to as voice recognition (voice recognition/interpretation software is used to convert spoken language into written text – i.e. a speech-text translator and has no recognition of the individual speaker). Whilst this is a conceptually simple form of user authentication, it may be inappropriate in certain (noisy) environments.

2.4.10 Dynamic signature

Of all the biometric measures, signature recognition is probably the most familiar to the end-user. The principle of providing a signature as a means of authentication is a historical part of western culture – thus potentially avoiding some of the problems of acceptability. However, it should be note that this would be one of the most intrusive biometric-based authentication techniques if applied in a continuous context, requiring a user to stop work and sign their name. A significant advantage of signature recognition over keystroke analysis or voice recognition is avoiding the dependency on technical ability or language. However, there is also the need for a signature capture device – typically provided as a graphics tablet.

2.4.11 Summary of techniques

Each of these methods of enhancing user authentication has advantages and disadvantages. Perhaps the most significant of these is the user acceptance of such measures. To determine the level of user acceptance, a survey was carried out, the details of which are listed in chapter 3. There are also other issues to consider, specifically, effectiveness, transparency and cost. These were not considered as part of the survey as the majority of the respondents would not have the adequate knowledge in the area to provide appropriate responses. The benefits and disadvantages of each of these methods are briefly described in Table 2.6. False Acceptance/Rejection Rates are also shown (where available). Unfortunately it is often difficult to gather sufficient quantitative data from the biometric manufacturers. In particular, mouse dynamics has been omitted from this table, as there are no commercially available products to evaluate.

It is interesting to note that hardware/software vendors seem to ignore the only statistical analysis of the effectiveness of biometric authentication methods. In fact a number of vendors stated the error rates for their products as "low", an entirely subjective assessment. This apparent ignorance of the appropriate measures may further hinder the acceptance of these methods into both commercial and private sectors.

There are few standardised methods of comparison between biometric authentication measures. However, a method used by the Independent Biometric Group (IBG) is the Zephyr™ chart, which shows the authentication methods around the outside with the assessment criteria ranked inside the chart. The evaluation criteria used are listed in Table 2.7.

Method	FAR	FRR	Advantages	Disadvantages
Password	Not Applicable		No hardware requirements. Transparent to user.	Users forget, share and write down passwords. Selection of passwords is usually poor (names, places, dictionary words etc.)
Keystroke analysis [BioPassword]	0.68% Equal Error Rate		No hardware requirements. Transparent to user.	May not be suited to certain typists (touch). Not appropriate in certain industries (graphics).
Face recognition [Miros]	"Low"	<0.2%	Simple to use. Camera can be integrated into monitor	Camera required.
Voice verification [Motorola]	1.07% 1-Phrase 0.2% 4-Phrase [Equal Error Rates]		May be ideal for switchboard/VoIP based systems. May be effective for telesales/phone banking.	Requires audio hardware. May not be suitable in office environments (intrusive to other users) or industrial (background noise).
Signature analysis [PenOp]	Variable Not Quoted		User familiarity with concept of signatures for ID may help acceptance	Requires signature device. Intrusive.
Iris scanning [IriScan Inc.]	0.0001% [Equal Error Rate]		High level of user discrimination.	Requires scanning device. Intrusive. Acceptability may be hindered by confusion with retinal scanning (laser).
Retinal scanning [Eyedentify Inc.]	0.0001%	0.1%	High level of user discrimination.	Requires scanning device. Intrusive. Uses laser light which may worry users.
Hand geometry [Recognition Systems]	0.15%	0.15%	Simple to use.	Requires scanning device. Intrusive.
Fingerprint analysis [Compaq]	0.001%	0.5%	Simple to use. Hardware becoming cheap (<£100).	Requires scanning device (relatively cheap). Intrusive. Criminal connotations (fingerprinting)

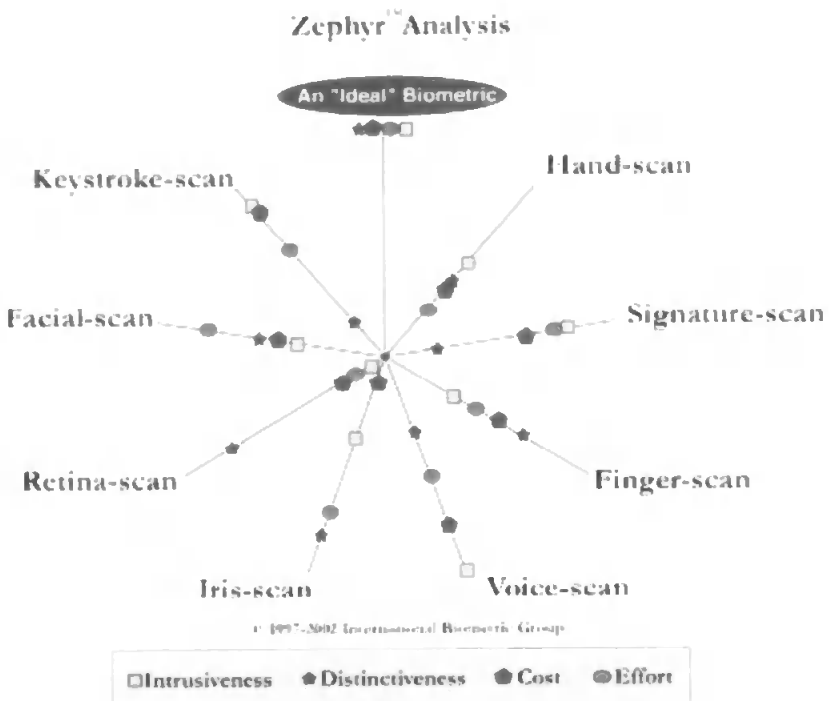
Table 2.6 - Comparison of commercially available authentication methods

User Criteria Aspects which relate to the user.	**Technology Criteria** Aspects which relate to the technology
1. **Effort** - How much time and effort is required on the part of the user i.e. registration, enrolment, authentication.	1. **Cost** - Cost of any necessary hardware capture device together with any software, support and training required.
2. **Intrusiveness** - How intrusive the user *perceives* the system to be.	2. **Accuracy** - How well the system identifies individuals. [FAR/FRR]

Source: International Biometric Group (2002)

Table 2.7 - Table of evaluation criteria used in Zephyr chart

A Zephyr chart allows the comparative strengths and weaknesses of each measure to be examined visually. As an example, Figure 2.3 shows a Zephyr chart available from the International Biometric Group (IBG,2002). The ideal criteria for a successful biometric measure is shown at the top of the chart, (this only occurs when all four criteria are at their optimum settings and are, therefore, located at the outer perimeter of the chart).



Source: International Biometric Group (2002)

Figure 2.3 - Zephyr™ Chart for common biometric-based authentication methods

If we consider the use of keystroke dynamics, the criteria are distributed across the entire range (it should be noted that this is based on one-off authentication rather than continuous monitoring). Of the four criteria, keystroke analysis satisfies two, namely non-intrusiveness and cost. Many biometrics fail on the intrusiveness criteria as they require the user to perform some explicit action to authenticate themselves. Static keystroke analysis offers the ability to analyse keystroke patterns during the entry of the username/password pair (or some other form of identification string) therefore reducing the

impact on the user. Cost is minimised (as far as hardware issues are concerned) as almost all PCs will have a keyboard, although there will be some additional expense as a result of additional costs (i.e. training, maintenance etc.). The effort involved to authenticate via keystroke analysis is still small in comparison to some other methods as the user is only required to type in a small sample of text. Ideally, this would be enhanced to monitor free-style text, thus further reducing the effort involved to achieve authentication by this method. The final criterion is that of accuracy. Static keystroke analysis is ranked comparatively poorly against the other methods; this is mainly due to the limited amount of data with which user authentication can take place and the variability likely with such a small sample size.

2.5 Conclusions

This chapter has presented the increase in levels of detected computer crime and abuse cases over the last twenty years (as reported within the UK). Although it is not possible to attribute the increase in reported crime to any underlying specific cause (i.e. was the rise due to increased criminal/malicious activity, increased reporting of crimes or the increased success in detecting and solving such cases) it is clear that there is an increased level of incidents with a respective increase in costs.

One of the problems in addressing the increase in crime is that of attributing blame. The surveys presented have indicated a high level of cases originating from within organisations however, in order to identify the person or persons responsible for the incident, it is necessary to have absolute authentication of a user's identity. A misfeasor

would be acting under their own account with appropriate permissions and hence would not be highlighted by the usual authentication measures, while a masquerador may only be detectable through improved authentication techniques. Given that many users leave their terminals logged in and unlocked for long periods of time (usually all day), simple login authentication is not sufficient.

This chapter has also discussed the limitations of current authentication techniques, highlighting in particular the shortcomings of the commonest method – the username/password combination. To overcome these problems a number of alternatives have been presented together with an overview of their advantages and disadvantages.

It is clear that some form of improved authentication is necessary to ensure accurate authentication at login. It is also important for the concept of monitoring to be considered to ensure that the active user is the original user who logged in – i.e. continuous user authentication/monitoring.

Before investigating potential improvements to user authentication, a survey was conducted to determine the need for improved user authentication and supervision, and to determine the user acceptability of such measures (which would help to inform the design and implementation of an improved user authentication and supervision system). This survey is described and discussed in the following chapter.

Chapter 3

A Survey of User Attitudes and Perceptions

3.1 Introduction

The previous chapter identified the weaknesses of current authentication systems and proposed a number of alternatives to the current secret-based approaches. Before attempting to implement and evaluate alternatives to the traditional username/password approach, it is first necessary to determine the attitudes of computer system users to the current methods of authentication, as well as their perceptions of the alternatives presented in the previous chapter. This is increasingly important as users frequently find ways to bypass security mechanisms that are considered to be inconvenient or intrusive.

This chapter presents the results of a survey that aimed to determine awareness, and acceptance of, a range of alternative authentication and supervision methods. The survey respondents were asked to comment on current techniques, identifying bad practise (e.g. sharing passwords), as well as mitigating circumstances (e.g. large numbers of passwords for different systems), before considering a number of alternative authentication mechanisms.

3.2 Survey Overview

In order to determine the acceptability of user authentication and supervision techniques, a survey was conducted to assess the attitudes and awareness of current IT users. The survey aimed to assess the following issues:

- attitudes towards different forms of user authentication;
- the attitudes towards the concept of continuous monitoring.

The survey questionnaire consisted of 53 main questions, the majority of which were multiple choice, with the remainder requiring short written responses (a copy of the paper-based questionnaire can be found in Appendix A). Many of the questions contained multiple sections, resulting in a maximum of 130 possible answers per respondent. The survey was split into a number of categories, each focussing upon a specific area of interest. Questions 1-7 gathered general details, to determine the gender, age, education, and level of computer use; these provided demographic information on the survey response base. Questions 8-14 considered the use of computers within the respondent's work environment, whilst questions 15-19 considered the use of computers at home. These helped to provide information on the spread of IT into the home and work contexts, as well as the likely IT awareness of the respondents. Questions 20-34 were intended to determine individual opinions and knowledge in the area of computer crime and abuse. The findings relating specifically to these questions are presented in Dowland et al (1999) and will not be discussed here. The final section (encompassing questions 35-53) looked at the respondent's views on user authentication and supervision.

The survey was distributed to a wide range of individuals and organisations with the intention of gaining a diverse variety of opinions. The questionnaire was made available in two forms, a printed copy and an online version. Approximately 300 printed surveys were distributed with 148 completed responses being received, representing a response rate of 49%. A further 27 responses were submitted via the web site resulting in a total of 175 responses. It should be noted that, whilst questionnaires were sent to companies, the focus of the questionnaire required respondents to reply from an individual rather than organisational perspective (i.e. respondents were asked to consider the questions posed from a personal perspective rather than considering an organisational security policy –

should one exist). As such, these responses were still representative of a personal rather than business viewpoint on the issues.

3.3 Demographics

The survey demographics showed a male dominance in all age groups, with 80% of the total respondents being male. In terms of age, 74% of the respondents were below 35, indicating that the vast majority of the responses were likely to be from people who had grown up with IT to some extent. The overall breakdown of respondents by age group is given in Figure 3.1.

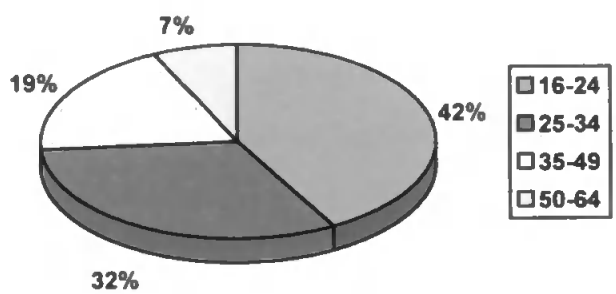


Figure 3.1 - Survey respondents by age

In terms of employment background, a high number of responses were received from the technology fields (with 103 out of the 175 responses claiming to be from the computing, communications or engineering domains). Academically over 70% of the respondents claimed to hold A-Levels or above, with 44% having a degree level education. This represents a generally high level of academic achievement and reflects the fact that the distribution of a large proportion of surveys occurred via academic channels.

The vast majority of respondents had considerable familiarity with IT, with over 98% having used a computer for over 1 year, 88% using a computer at work and 84% using one

at home. In terms of the level of use, the results indicated that, in both home and work environments; over half of the respondents used their systems for four hours per week or more. The respondents were also asked about the availability of Internet access. 129 respondents (88%) claimed to have access at work, while 69 respondents (48%) claimed to have access at home. The latter statistic indicated that the respondent group was clearly ahead of the UK average in terms of Internet adoption, as the penetration into UK homes (at the time of the survey) was considered to be around 14% (ICM, 1999). It should be noted that more recent surveys have estimated UK Internet adoption at around 50% (Ofcom, 2003), which would be more in-line with the respondents from this survey.

The general information above shows that the respondents had considerable experience using computers in both home and work environments. As such it was considered that the respondents had adequate background knowledge in order to comment on a range of authentication techniques and the issues relating to them.

3.4 Password based authentication

Given that they represent the most common, and therefore familiar, form of authentication, the survey began by assessing respondent attitudes towards passwords. The results indicated that over 91% of respondents relied on passwords for access control to their computers – a figure that is generally compatible with the 1998 KPMG security survey which showed 97% of organisations using them (KPMG, 1998). This situation appears to have changed relatively little since the survey was completed, with the latest DTI survey (DTI, 2002) reporting between 82% and 99% of organisations still using passwords for user authentication (the precise figure varies by organisational size). The high reported use

of passwords among respondents ensured that the subsequent questions about password implementations and password limitations would be answered based upon practical experience.

Due to the dominance of passwords, most users have multiple passwords for different systems and applications. When asked how many different systems or applications they use which require passwords, 26% of respondents claimed to have five or more, with 18 people claiming in excess of ten (Figure 3.2). This is becoming an increasingly problematic issue due to the increased usage of web-based systems. With the spread of on-line banking, e-commerce and login details needed for a wide range of sites, an average user may have numerous username/password combinations. Added to this, many web sites have specific requirements for the format of usernames/passwords. These requirements may include the familiar minimum length requirements, but they are increasingly requiring mixed case, punctuation or numeric characters to occur in login details (especially in the on-line banking sector). It is likely that the number of systems that users will be required to authenticate themselves to will only increase over the coming years.

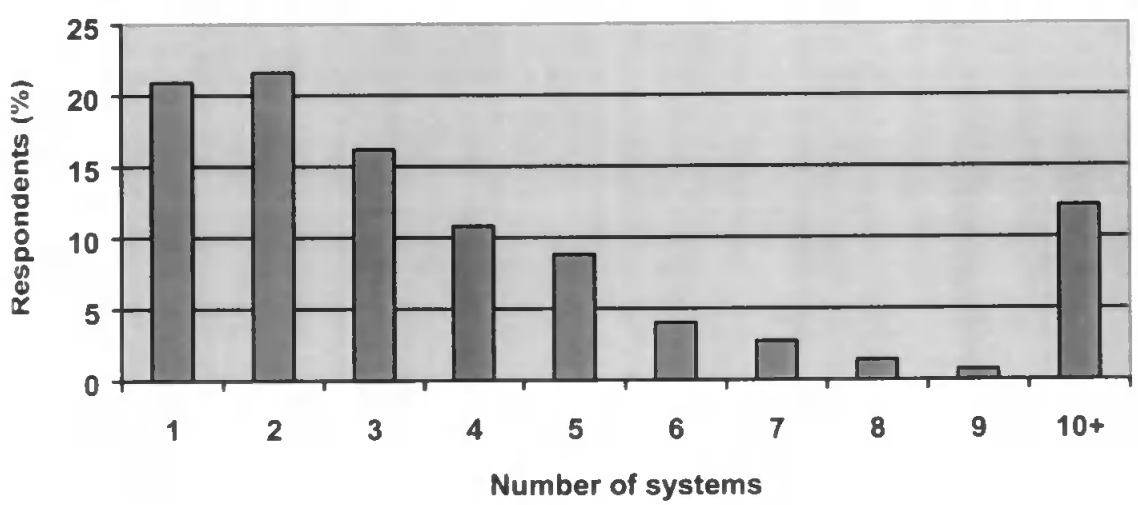


Figure 3.2 - Number of different systems/applications used requiring passwords

The requirement to remember such a large number of passwords can cause a major problem for users. It is, therefore, no surprise that users frequently select dictionary words or personal names as the basis for their passwords, as these are easier to remember. Having said this, only 15% of respondents felt that their passwords could be easily guessed. The phrasing of the question in this case gave examples of information that, if used as a basis for selection, could render the password more easily guessed (i.e. "is it part of your address, name, partner's name?"). Although the majority of users considered themselves to be safe on this basis, the question did not provide an exhaustive list of what might constitute obvious choices. As such, many respondents may still have been using insecure passwords, such as dictionary words (which the L0phtCrack tool mentioned in the previous chapter can determine in less than a minute). It should, however, be noted that even when forced to select more complex (and hence less meaningful) passwords, users often compromise these by writing them down (15%) or sharing them with other users (29%). This is not always done without any thought to the consequences, but for some users, the benefits of having a note stuck to the monitor of their PC with their password in clear sight for all to see is better than the problems caused by forgetting the password every four weeks due to forced password changes!

Not only do users often choose insecure passwords, they also frequently select the same password for multiple accounts, with 40% of respondents re-using the same password. As such, should an intruder gain access to one protected account, it is quite likely that they would be able to reuse that same password for other machines and applications. This could be extrapolated further when considering the impact upon web site logins – if users choose the same (insecure) login details across multiple sites, it is likely that a successful intruder on one site would have unrestricted access to a wide range of sites. Responses to

subsequent questions revealed that 31 (21%) of the 151 respondents who used computers at work claimed to have used another person's password without their consent or knowledge. It is surprising that so many people have illegitimately used other users' accounts and also to find that so many of the respondents were prepared to admit to doing so. It is possible that some of the respondents lied about such illegitimate activity and, as such, the real figure could be somewhat higher. It could be argued that if respondents had taken better care of their passwords (by following basic guidelines on selection and use) it is likely that there would have been less opportunity for such abuse.

A further issue is that of the password's lifetime. Once a password is illegitimately acquired then, without time limits, restricted logins or account monitoring, it is possible that the intruder would remain unnoticed until he/she committed an act that caused some form of disruption that would consequently be detected. If we again consider the use of web sites, a compromised account could provide not only unrestricted access (if no time limits were imposed on accounts) but, more worryingly, could provide unlimited access to other accounts sharing the same login details. The respondents were asked how frequently they changed their passwords and if they were forced to change their passwords by the system or the system administrators. As indicated in Table 3.1, an alarming 34% claimed to never change their passwords. Furthermore, the responses to the subsequent question revealed that 51% were not forced to change their password by the system. The former represents bad practice on the part of the users, whereas the latter reflects poor system administration. From an administration point of view, it is more encouraging to observe that 70% of users claimed to use systems in which a minimum password length is enforced. Having a minimum length of seven or more characters helps to ensure that passwords are more resilient to brute force attacks.

Frequency of password change	Respondents
Weekly	2%
Fortnightly	1%
Monthly	25%
Six-monthly	18%
Less frequently	20%
Never	34%

Table 3.1 - Frequency of password changes

These results serve to underline some of the known problems with passwords and provide the justification for the subsequent questions, which asked users about other forms of authentication.

3.5 Alternative authentication and supervision methods

One of the main objectives of the survey was to evaluate user's opinions regarding different authentication methods. In order to achieve this, the respondents were asked to rate the acceptability of a variety of initial login and continuous supervision techniques on a 5-point sliding scale from 'totally acceptable' to 'totally unacceptable'. A total of nine methods were cited, ranging from passwords to a variety of physiological and behavioural biometric methods. Each of the methods was briefly described on the questionnaire sheet to ensure that the respondents understood the context (using the text shown in Table 3.2).

Method	Description
Keystroke analysis	Research has shown that users have different typing styles and that they can be identified by measuring the times between keystrokes.
Face recognition	A snapshot of the user, taken by a camera positioned on the monitor, is compared with a previously stored 'faceprint'.
Mouse dynamics	Similar to keystroke analysis, users can be identified by the way in which they use the mouse.
Voice verification	A user's voice, when speaking a word or phrase into the computer's microphone, is compared with a previously stored 'voiceprint'.
Signature analysis	A user signs their name using a special pen and pad, the signature is digitised and compared with a previously stored version.
Iris scanning	A snapshot of the user's iris, taken by a camera, is compared with a previously stored image.
Hand geometry	This technique measures the physical dimensions of the hand using a small camera and compares these with previously stored values.
Fingerprint analysis	An automated version of the fingerprint identification system similar to that traditionally used in criminology.

Table 3.2 - Biometric methods, as presented to survey respondents

Table 3.3 summarises the ranked results, which are also illustrated graphically in Figure 3.3. The responses have been normalised to reflect the variable response rate to each question, as there was a higher response rate to questions on initial login authentication (probably reflecting a lack of understanding of the concept of continuous supervision amongst some respondents). The positive responses ('totally acceptable' and 'acceptable') were summed and then the total number of negative responses ('unacceptable' and 'totally unacceptable') were subtracted, thus producing a rank of user preference.

Method	Initial login authentication	Continuous supervision
Password	95.7%	-10.2%
Keystroke analysis	29.8%	25.5%
Face recognition	49.1%	3.2%
Mouse dynamics	21.3%	21.8%
Voice verification	53.4%	-0.6%
Signature analysis	40.1%	-35.9%
Iris scanning	47.2%	-16.8%
Hand geometry	44.4%	-19.9%
Fingerprint analysis	48.8%	-16.0%

Table 3.3 - Ranked user preference of security methods

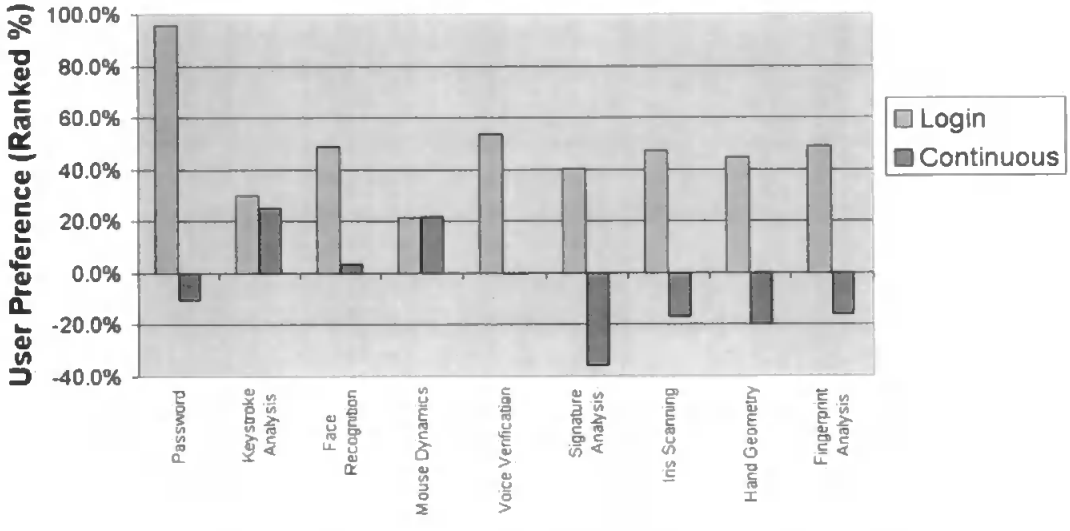


Figure 3.3 - User preference of authentication methods

As expected, the most popular form of initial login authentication was the password, with 90% of respondents rating it as 'totally acceptable' (scoring more than twice as many votes in this category than most other methods). However, it should be noted that if all respondents were forced to adopt correct password selection and use procedures, it is considered unlikely that the level of acceptance would be quite so high. This high level of acceptance for passwords did not mean the outright rejection of alternative methods and many also achieved respectable scores.

It is clear that there is a reasonably significant level of user acceptance for all the initial login authentication techniques suggested. Methods such as face recognition, voice verification, signature analysis, iris scanning, hand geometry and fingerprint analysis were all considered favourably. It is interesting to note that all of these techniques (with the exception of signature analysis) have had significant media coverage, especially through film and television. It is possible that familiarity with these techniques influenced the respondents' choices. The acceptance of signature analysis cannot be readily explained by the familiarity with the technology through the media; however the concept of a signature as a means of identity verification is well established in our society.

After passwords, the most acceptable forms of login authentication were considered to be voice verification and fingerprint recognition, scoring raw overall acceptability ratings of 68% and 67% respectively (full table of results in Appendix A). The latter result is somewhat surprising, in that conventional wisdom suggests that the association of fingerprints with criminal identification may represent a potential barrier to user acceptance (Observer, 2002). However, it is clear from these results that the majority of respondents are comfortable with the concept. It can, however, be noted that, in the normalised results (Table 3.3), face recognition scored higher than fingerprints once negative responses had been taken into account (un-normalised results can be found in Appendix A). It should be considered that while these results show a positive response to the use of such techniques, the respondent base was predominantly represented by IT-literate, well-educated individuals working regularly with PC's. As such, it is perhaps unsurprising that a high level of acceptance was achieved. If a wider (more inclusive) survey were conducted, it is possible that certain techniques (e.g. fingerprint scanning or retinal scanning) may have achieved lower acceptance for login authentication due to public distaste for such mechanisms; while others (e.g. keystroke analysis and mouse

dynamics) may have had a lower acceptance due to lack of exposure or knowledge of the techniques.

One of the significant questions posed in the survey was whether respondents would be comfortable with the concept of continuous supervision. This would provide a means for authentication to become an ongoing process within a logged in session, rather than being merely a one-time judgement at the beginning. This, in turn, would guard against situations such as an impostor replacing a legitimate user at the terminal, or an impostor who may have been able to fool the initial login authentication system. In general, the respondents were positive towards the idea of monitoring, with 43% considering it acceptable, though 29% were unsure. It is, perhaps, unsurprising that there was some reluctance – if continuous monitoring were deployed in an organisational context, employees may be concerned about the use of data gathered by such techniques. For example, if keystroke analysis were used to monitor a users' typing, the statistics could be used to profile a users' productivity or to confirm when the user was at his/her desk. It is also possible that an employer (or administrator) could use keystroke analysis software to actually record the typing of an employee and then obtain potentially sensitive/private information (e.g. on-line banking details). Safeguards and other concerns are discussed later in this section.

It is recognised that the concept of continuous supervision also introduces ethical considerations. Indeed, 70 respondents (40%) stated that they would consider monitoring as an invasion of their privacy (interestingly, 18 of these had stated that they considered this to be acceptable), with a further 18% being unsure. It is clear that if continuous supervision of users is to be implemented, then certain safeguards should be considered. In

particular, users should be aware of the intended uses of the information collected. 45% of respondents felt that they could not trust their organisation to use the supervision data for security-related purposes only, and were concerned that it could be utilised for an ulterior motive, such as monitoring work productivity. 85% stated that users should be aware of any monitoring being used, potentially demonstrating an inherent distrust of the concept of a big-brother scenario. The simplest way to ensure these requirements are met is to involve the users in the planning and implementation of these systems and provide clear policies on the uses for the gathered information. With the introduction of data protection laws covering personal information stored electronically there will also have to be provision for securing the profiles and session data obtained in any continuous monitoring system as well as ensuring access (upon request) to all personal data stored in the system (in the same way that an employee may ask for his/her personnel records).

The respondents considered only three techniques to be acceptable for continuous monitoring; namely keystroke analysis, mouse dynamics and face recognition (the latter being with a very low preference). Whilst the overall ranked results reflected sensible views, some of the individual responses in the underlying data did provide a few surprises. In particular, 34 respondents rated the use of signature analysis for continuous monitoring to be 'acceptable'. This is most likely to be a misunderstanding, as few computer users would be prepared to stop work and sign their name intermittently (a view borne out by the fact that 90 respondents rated this as 'unacceptable').

Respondents were also asked to consider how long they would be prepared to spend creating a behaviour profile that the monitoring system would then use to authenticate them. The responses are shown in Table 3.4. It is clear that the majority of users would

not be tolerant of explicit profiling activity for any long periods. Equally, the time that most of them would consider acceptable is 15 minutes or less – which would be unlikely to be adequate for some measures (e.g. whilst face and fingerprint recognition systems would allow adequate registration within this time, accurate measures relating to typing and more general system usage would require longer periods – potentially in the order of days rather than minutes). As such, elements of profiling would need to occur as a transparent background task in order to ensure user acceptance.

User-profile set-up time	Respondents
No time	11%
Up to 5 mins	36%
Up to 15 mins	24%
Up to 30 mins	13%
Up to 1hr	12%
> 1hr	5%

Table 3.4 Acceptable duration of profiling activity

Once a profile has been created, there is still the possibility that a monitoring system may falsely reject a legitimate user, believing them to be an impostor. The questionnaire made the respondents aware of this and asked them how frequently they would be willing to tolerate such errors. The results are presented in Table 3.5 and illustrate that any deployed system would need to have a very low error rate in order to avoid alienating the user population.

Frequency of false rejection	Respondents
Hourly	7%
Daily	27%
Weekly	36%
Never	29%

Table 3.5 Perceived tolerable frequency of false rejection by monitoring system

Finally, the respondents were asked to indicate which fields/sectors would benefit most from supervision of users by computer, rating the benefit from 'great benefit' to 'no benefit at all'. As with the results in Table 3.3, the responses have been normalised to provide a ranked overall benefit. The positive responses ('great benefit' and 'some benefit') were summed and then the total number of negative responses ('little benefit' and 'no benefit at all') was subtracted, thus producing a rank of overall perceived benefit. These results are shown in Figure 3.4.

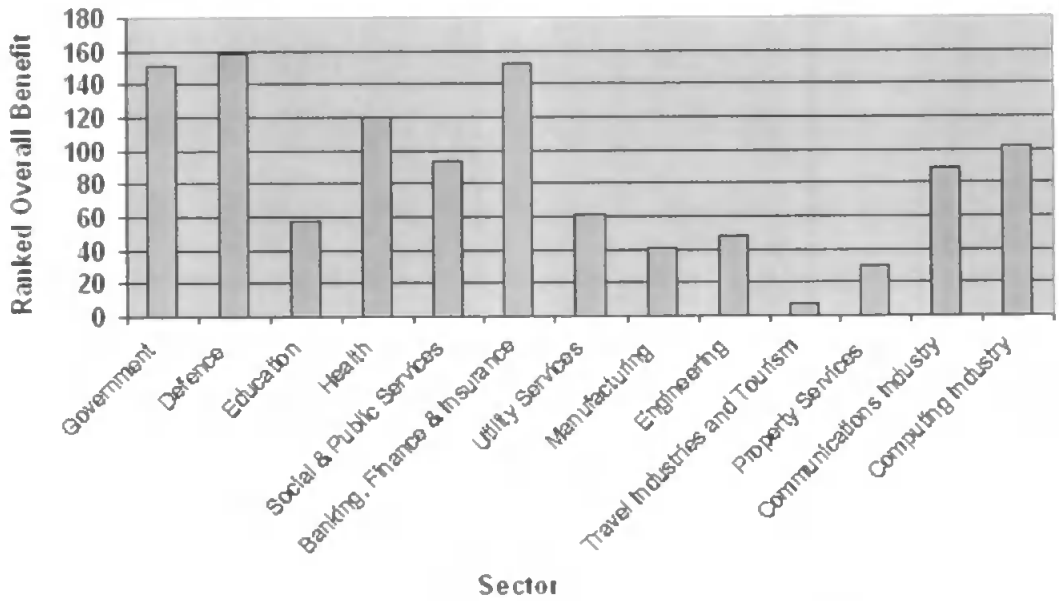


Figure 3.4 - Benefit from monitoring by sector

As expected, the majority of respondents considered the areas of government, defence, health and banking to benefit most from user supervision (these being the areas with the most obviously sensitive systems and data to protect). However, the respondents felt that all areas could benefit from improved supervision, showing that there is still considerable concern over the perceived level of computer security across all sectors.

3.6 Discussion of results

The results reaffirm the accepted shortcomings of password-based authentication (poor selection, sharing and forgetting passwords etc.), as well as the fact that, in spite of these, it remains the dominant form of user authentication. However, the fact that the respondents have shown a willingness to use alternative authentication techniques can be considered to be encouraging. It should be noted, however, that in the majority of cases, it is unlikely that the respondents had actually used the techniques that they were being asked to comment upon. As such, it is possible that their views may change if presented with the practical experience.

Given that a strong preference was expressed for passwords, consideration should be given to retaining them as the means of login authentication, whilst identifying means to compensate for their weaknesses. Suitable strategies in this respect could include (with numerous other combinations):

- Utilising password login in conjunction with transparent keystroke analysis of the information entered. In this way, the user would be authenticated not only by what they type, but also how they type it. This should not have any significant influence on user acceptance, as the primary authentication mechanism will still appear to be the password. This would, however, not overcome a fundamental problem of password-based system; that once logged in, there is typically no further authentication required throughout a users' session. A user may login at the beginning of the working day, providing a valid match is made between the typed password and the users' historical typing profile but

then be able to continue throughout the rest of the day without any form of challenge. While this approach enhances the initial login authentication, it provides no further session protection. Given that many users leave their workstations logged in all day (with some not even logging out when they leave in the evening), this would only be able to potentially catch another user misusing a colleague's details at login. While this offers no protection during the logged-in session, it does represent a significant improvement over login authentication.

- Retaining password-only authentication at login, but supplementing it with on-going supervision during the user session. (NB the term continuous is used to distinguish between static monitoring; e.g. monitoring key presses during login, and dynamic monitoring which has the ability to monitor a users' behaviour throughout a session – the time interval for which could be truly continuous or intermittent.) The survey results suggest that techniques such as keystroke analysis and mouse dynamics would be conceptually acceptable to users in this regard. This approach could still retain the user-accepted login mechanism (i.e. username and password), but would switch to a supervision/monitoring role once the user had been authenticated; to monitor the users' typing behaviour, application usage or mouse dynamics (or some other metric) throughout the logged-in session. This would effectively remove the ability for a third party to hijack the users' session providing the resolution of the monitoring was sufficient to identify an impostor within an acceptable time frame.

The respondents preference for passwords is in agreement with the previously published results from the Australian TRUST project, which (based on an experiment with 76 participants) found users' principal preference to be for passwords, followed by physiological biometrics and, finally, behavioural measures (Deane et al, 1995). The latter finding is, however, in contrast to the results from this study in that (for continuous monitoring) the behavioural techniques of keystroke and mouse dynamics were chosen in preference to the physiological technique of face recognition. Indeed, in the TRUST study, keystroke analysis and pointing device based verification scored the lowest of the seven biometrics assessed.

Although many considered the concept of continuous supervision to be acceptable for security purposes, the respondents showed concern over the potential wider use of such data. As such, it is important for organisations to establish agreed working practices to employees before proceeding with such methods (this may assist in reassuring those such as the 29% of respondents who were undecided over the acceptability of the monitoring concept). If such practices are not naturally adopted by organisations, it is possible (maybe even preferable in some cases) to legislate on acceptable supervision practices. This could be implemented in a similar way to that which restricts the rights of an employer to intercept and/or read an employee's email correspondence (HMSO, 2000).

Overall, a significant factor in the acceptance of alternatives to the password will be that of education. If people can be shown that newer authentication techniques are safe, reliable and secure, then their acceptance is likely to be improved. This will be best demonstrated in the UK over the next ten years where a trial is being undertaken to determine public acceptance of biometric identification. This scheme (Home Office, 2003), announced by

the government in November 2003, is due to begin in early 2004, with the aim of introducing a national identity card scheme comprising driving license, passport details and incorporating a biometric ID (iris or fingerprint scan). This scheme, more than any other in recent years, will depend on public acceptance to determine the widespread used of biometrics. If the public reaction to biometric identification is negative, it will be increasingly difficult to implement such systems – even when forced through employment contracts or government legislation.

Of course, the UK is by no means alone in considering the use of widespread biometric systems. The US has been trialling biometric systems at border points for some time now (ZDNet, 2002) and has recently announced plans to use fingerprint scanning for overseas workers entering the country (USDoS, 2003).

3.7 Conclusions

The survey has shown that, although demonstrably weak, the password remains the most popular form of authentication in the minds of users. However, a number of other methods emerged as possible contenders, and it is possible that practical experience of using them, combined with improved awareness of the vulnerabilities of passwords, would increase their perceived acceptability as alternatives.

Another conclusion that can be drawn from the survey results is that the use of continuous supervision is, in general, acceptable. However, the viability of such a scheme would be dictated by the methods chosen, and subject to suitable assurances being given to the monitored population regarding the planned uses of the collected data. This is an

important issue as user acceptance of such a radical approach as continuous user authentication will probably be just as important as overcoming the technical issues of implementation. There are also potential problems with the level of user support needed to make these systems work. There may still be reluctance on the part of users to undertake the necessary profiling to make the systems used sufficiently reliable for day to day use, as well as resentment should the system reject valid users too frequently. The survey has also indicated a lack of understanding over authentication in general and the specific issues of continuous user supervision. It is unlikely that re-educating users will avoid the problems with passwords, however, it is possible that increased awareness and understanding could improve acceptance of alternatives.

Given the acceptance of some form of supervision during a logged-in session and the preference for a keystroke-analysis based method, the latter parts of this thesis will discuss the application of transparent keystroke analysis in a modern operating system. Chapter 5 will introduce in more technical detail the concepts of keystroke analysis and identify the range of metrics that can be assessed, while chapters 6 and 7 evaluate the results of two trials conducted to ascertain the feasibility of a keystroke analysis based supervision system.

Before committing to keystroke analysis, the next chapter presents the results of some further investigations that were conducted in order to determine whether the weak, password-based approach could be more easily replaced by an alternative secret-based knowledge approach to user authentication.

Chapter 4

Assessing Alternative Methods of User Authentication

4.1 Introduction

The survey described in the previous chapter identified the need for improved user authentication, with the respondents indicating a continued preference for secret-based authentication (e.g. the use of passwords). Unfortunately, although users may still have a preference for passwords, the weaknesses of this method are well known. In order to provide improved security, while still maintaining the users' preference for a secret-based approach, this chapter considers the results of two trials conducted to investigate alternative forms of user authentication using software-based methods (which do not incur any additional expenditure on hardware technologies, and as such are likely to be considered favourably by system administrators).

Previous research into alternative software-based methods has identified a number of potential approaches:

- Cognitive and associative questions – a question and answer based approach using easily memorable (but nonetheless secret) information (Haga and Zviran, 1991). This approach has a distinct disadvantage as it can require a lengthy exchange between the user and the system in order to be authenticated.
- Graphical authentication – this uses graphical images that the user must memorise and identify on screen. There are a number of implementations, ranging from a picture with a number of pre-defined regions (Blonder, 1996 and Jermyn et al., 1999) to the Déjà Vu system that used randomly generated electronic art images (Dhamija and Perrig 2000).

These methods share the common advantage of easy implementation – i.e. they are software based methods that have no hardware dependencies (unlike most biometric-based authentication methods that require dedicated hardware).

While previous trials have shown the individual techniques to be viable authentication measures, there have been no comparative experiments to evaluate user acceptance of these measures side by side. As such, two trials were conducted in order to evaluate the techniques. The first, described in the next section aimed to test users' recall of personal (but secret) information. The second trial, described in section 4.3, took a longer-term approach to evaluate user friendliness and acceptability. These trials allowed a comparison to be made between these techniques with users having similar levels of exposure to each method.

4.2 An Experimental Study Of Alternative Methods

The first trial was devised to evaluate five secret-knowledge based techniques. The methods selected were PINs and passwords (familiar methods, included to provide a baseline for reference), alongside two question and answer methods (using cognitive and associative questions respectively), and a graphical technique using an image-based PIN (hereafter referred to as an ImagePIN). The study sought to assess the practical effectiveness of the techniques, as well as the friendliness and perceived level of security from the user's perspective.

The effectiveness was gauged by means of a practical trial, using specially designed profiling and authentication software to present the various techniques to a series of participants. Opinions relating to the friendliness and security of the methods were then obtained using a written questionnaire – completed by participants after they had participated in an authentication phase. The design of the experimental tools and the follow-up questionnaire are described in the subsections that follow.

4.2.1 The Profiler

The Profiler required participants to identify themselves and then provide appropriate responses for each of the methods under test. The profiling procedure for each of the methods is summarised below.

- Passwords and PINs. The implementation of these methods was fairly standard, with each participant being asked to supply a 4 digit PIN and a password of at least 8 characters. Participants were requested not to select a password or PIN that they already used on other systems, as the aim of the exercise was to assess their ability to recall new details, and thereby put these more familiar methods on an equal footing with the other techniques when it came to assessing ease of information recall. Nonetheless, as later results will indicate, some participants did not follow this guideline.

- Cognitive questions. Participants were asked to provide answers to a series of twenty questions, each requiring factual or opinion-based answers. The

questions requested information that was personal to the participant, and would therefore be difficult for a potential masquerader to guess in an operational scenario. The questions used are listed in Table 4.1.

What is your mother's maiden name?
Where were you born?
What is your favourite colour?
What was the name of your best friend at school?
What is your favourite music?
What is your favourite food?
What was the name of your first pet?
Which primary school did you go to?
What is your favourite sport?
Where was your first house?
What make was your family's first car?
How old were you when you had your first kiss?
What is your favourite film?
Where was the first place you remember going on holiday?
What was your favourite subject at school?
What is the most important part of your body?
What is your favourite type of animal?
What is the name of your favourite relation?
How many cousins do you have?
What is your favourite shape?

Table 4.1 - Cognitive questions

Even in cases where the participants might not have had a genuine answer (e.g. they may never have had a pet), it was expected that they would still be able to provide a response that could later be reproduced if prompted to answer that question.

- Associative questions. Participants were then asked to provide word association based responses to a set of twenty keywords. The keywords are listed in Table 4.2, and were carefully chosen to ensure that a number of

different responses were theoretically possible in each case. For example, for the associative word "seven", responses might include "wonders", "dwarves", "sins" or "days".

Blue	House	Table	Computer	Friend
Peace	Glass	Marriage	Sea	Love
Cat	Music	Fire	Seven	Video
Father	Food	Remote	Fast	Door

Table 4.2 - Associative keywords

- ImagePIN. The user had to select five images from a number of icons, by clicking on them with the mouse. Later authentication would work by the user reselecting the same images in the correct sequence (images are shown in Figure 4.1). Limiting the selection to five images meant that the resulting ImagePIN would not be considered as resilient to brute force attack as the 6-8 character passwords that are typically recommended. However, when viewed as an alternative to a traditional 4-digit PIN, as used with ATM cards and such like, the concept was considered to offer some useful potential.

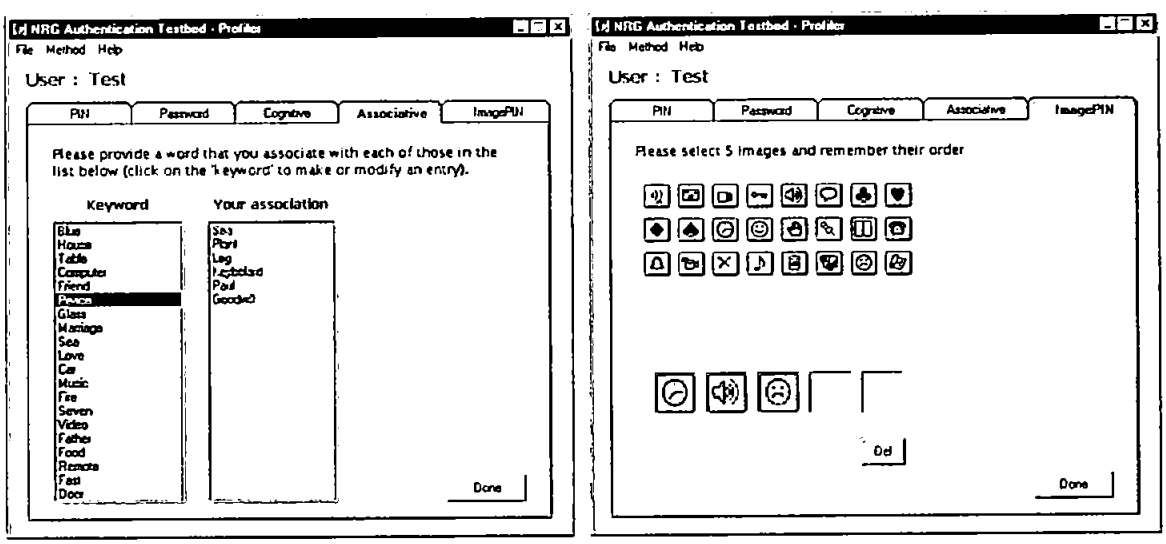


Figure 4.1 - Profiler system (showing associative questions and ImagePIN screens)

After the profile had been created, a short training exercise was performed using the second program, the Authenticator, in order to familiarise the users with how the later authentication test would work. After this, it was up to the participants to attempt to remember the details they had provided in order to perform the later authentication tests.

4.2.2 The Authenticator

The authentication tests took place one month after the initial profiling, with the aim of assessing whether the participants were able to adequately recall the information that they had previously provided during profiling and thereby authenticate themselves successfully. The interface of this system was very similar to that of the Profiler, and two aspects are illustrated in Figure 4.2.

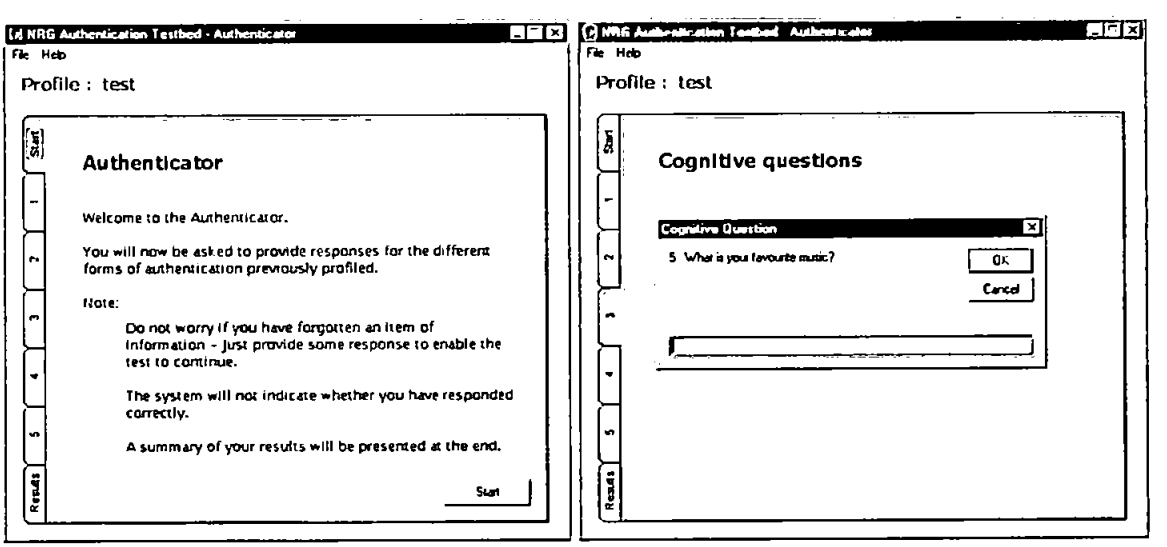


Figure 4.2 - Authenticator system (showing welcome and cognitive question screens)

In the case of the PIN, Password and ImagePIN methods, the participant was directly asked to provide the same information as originally profiled. For the cognitive and associative

methods, however, they were asked to answer five randomly selected questions out of the twenty that had been profiled in each case. This was considered to represent a good simulation of how such question and answer authentication techniques would be implemented in practice.

4.2.3 Participant questionnaire

Following the authentication test, all the participants were asked to complete a questionnaire, in order to determine their regular exposure to user authentication methods in other contexts and to assess their views about the different methods under trial. The key information collected from the participants was the ranking of the trialed methods according to the perceived user friendliness, level of security, and overall preference (although several other questions were also asked).

A total of 27 participants were involved in the profiling and subsequent authentication testing, and the results of the study are described in the next section. While the number of participants was small, this did allow for baseline comparisons to be made between the different authentication techniques.

4.2.4 Experimental Results

The results presented here consider the effectiveness of the techniques that were observed in the practical trial, as well as the participant's subsequent opinions in relation to the methods. It should be noted that, in the discussions and graphs that follow, the percentage figures have been rounded to whole numbers.

The practical evaluation began by examining the participant's performance in relation to the password and PIN methods. The results indicated that 70% of the participants had succeeded in authenticating themselves using passwords, and a similar proportion (67%) were successful using the PIN based method. Although these results initially appear very encouraging from the perspective of the participants being able to accurately recall the details after an absence of a month, the results of the accompanying survey revealed that a significant number of people had not followed the request to use different passwords and PINs than the ones normally used in other applications. In fact, only 56% used different passwords and 41% used different PINs. Within these subgroups, the authentication success was markedly lower - 53% of them succeeded in the password test and only 36% in the PIN version. By contrast, within the subgroups that used the same details as in other systems, 92% of them succeeded with passwords and 87% succeeded with PINs, so these figures can be considered to have artificially inflated the overall results.

In the cognitive and associative question tests, the participants were presented with a random selection of five questions out of the twenty that they were profiled for. Authentication was judged to be successful if all five questions were answered correctly. With the cognitive questions, a success rate of 59% was observed, whilst a number of further participants did succeed in answering a proportion of the questions presented to them. The distribution of correct answers in the cognitive test is shown in Figure 4.3.

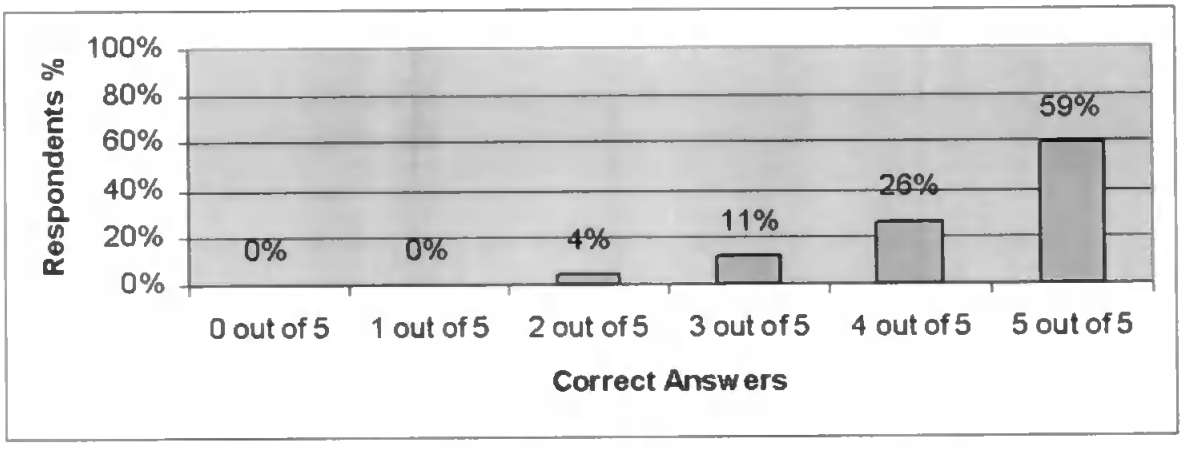


Figure 4.3 - Distribution of correct answers in cognitive questions

With the associative questions, the success rate was significantly lower. Only 4% (equivalent to one person) managed to correctly answer all five questions and the distribution of correct answers across five random questions is shown in Figure 4.4.

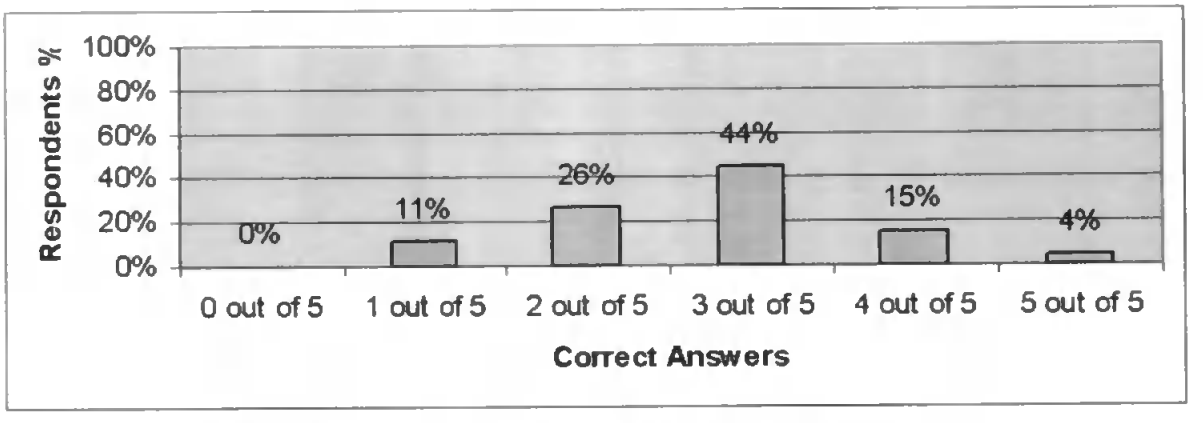


Figure 4.4 - Distribution of correct answers in associative questions

These results suggest that the associative question method is extremely problematic in relation to the correct recall of the information, and that participants are inconsistent in the words that they most readily associate with the keyword prompts. A further problem observed in the results of this study was that many participants chose the same associations for certain keywords, suggesting that the method could be easily targeted for masquerade attacks if used in practice. Table 4.3 summarises the cases in which the same associations were chosen for each keyword. The highest frequency of duplication was 44%, in which

respondents had chosen the word "control" as the associative response to the keyword "remote".

Keyword	Frequent word associations
Blue	Sky (41%), Sea (15%)
House	Big (15%)
Table	Food (22%)
Computer	Work (11%), Game (7%), Internet (7%)
Peace	War (15%)
Glass	Wine(22%), Broken (11%)
Sea	Blue (11%)
Love	Hate (11%), Marriage (7%)
Music	Rock (15%), Dance (7%)
Fire	Red (11%), Alarm (11%), Engine (7%)
Seven	Film (15%), Seven (7%), Days (7%)
Video	Games (11%), Movie (11%), Tape (7%)
Father	Mother (19%), Names (15%)
Remote	Control (44%)
Fast	Food (22%), Car (19%)
Door	Key (11%), Open (11%), Closed (7%)

Table 4.3 - High frequency associative responses

For the final technique, the ImagePIN, the participants had to recall their graphical PIN by reselecting the original icons in the correct order, with 63% being successfully authenticated. Even though the implementation of the method offered the participants the opportunity to somewhat undermine the security by selecting the same icon five times, only two participants actually did this.

Figure 4.5 summarises the overall results of the authentication tests, indicating the percentage of respondents who would have been successfully authenticated using each of the methods.

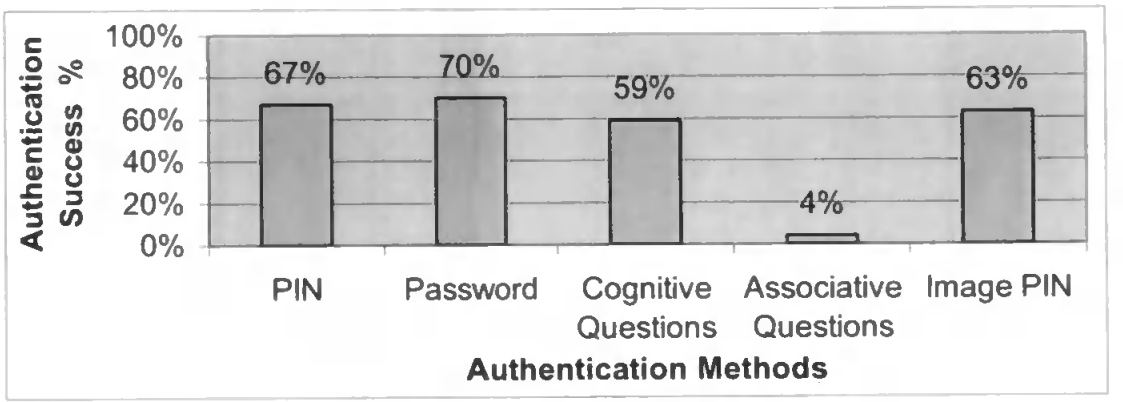


Figure 4.5 - Authentication methods success

Having experienced the techniques and witnessed their own performance, the participants were asked to rate the approaches on the basis of user-friendliness, security, and overall preference.

In terms of user-friendliness, participants were asked to assess the methods on a five-point scale, progressing from 'easy' to 'hard'. The best outright indicator of preference in this case was where methods were ranked as 'easy'. In this context, passwords were ranked first, receiving 48%, followed by the PIN method with 44%. The third position was shared by the cognitive question and ImagePIN methods, with 22% respectively. Last was the associative method with only 4%. Taking a wider view, and considering the total percentages for which methods were rated 'medium' or above, the password was still favourite, with 96%, followed by the PIN with 93%, cognitive questions with 81%, the ImagePIN with 59%, and associative questions with 48%. Looking from this viewpoint serves to place some separation between the cognitive and ImagePIN methods, and shows that more people tended to express concern over the friendliness of the latter technique. The full results are presented in Figure 4.6.

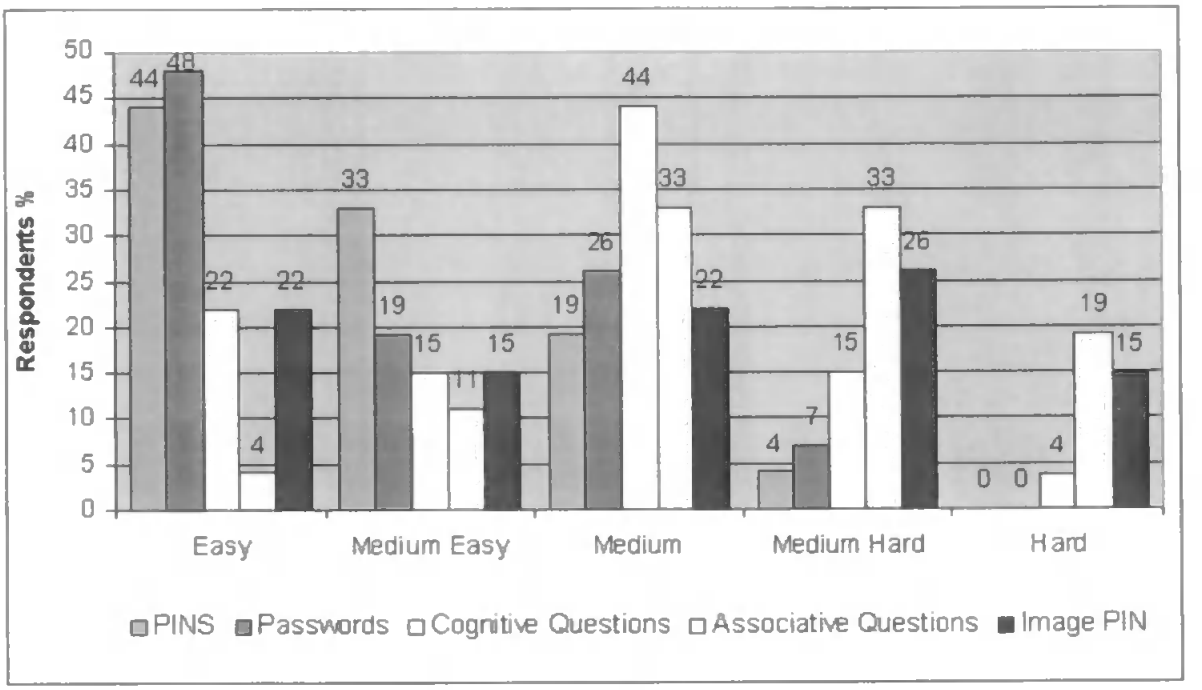


Figure 4.6 - Perceived user-friendliness

The second ranking addressed the perceived level of security. In this case, the password still fared well, with a combined total of 78% rating it to offer a 'medium' to 'high' level of protection. In this instance however, the popularity was also equalled by the cognitive and ImagePIN methods (and it can be noted that both of these methods actually exceed the results for passwords if only the 'high' and 'medium high' ratings are considered). Meanwhile, the PIN method attained 53%, and the associative approach was again ranked lowest, with 45% ranking it in the 'medium' to 'high' range. Figure 4.7 presents the perceived level of security for each authentication method.

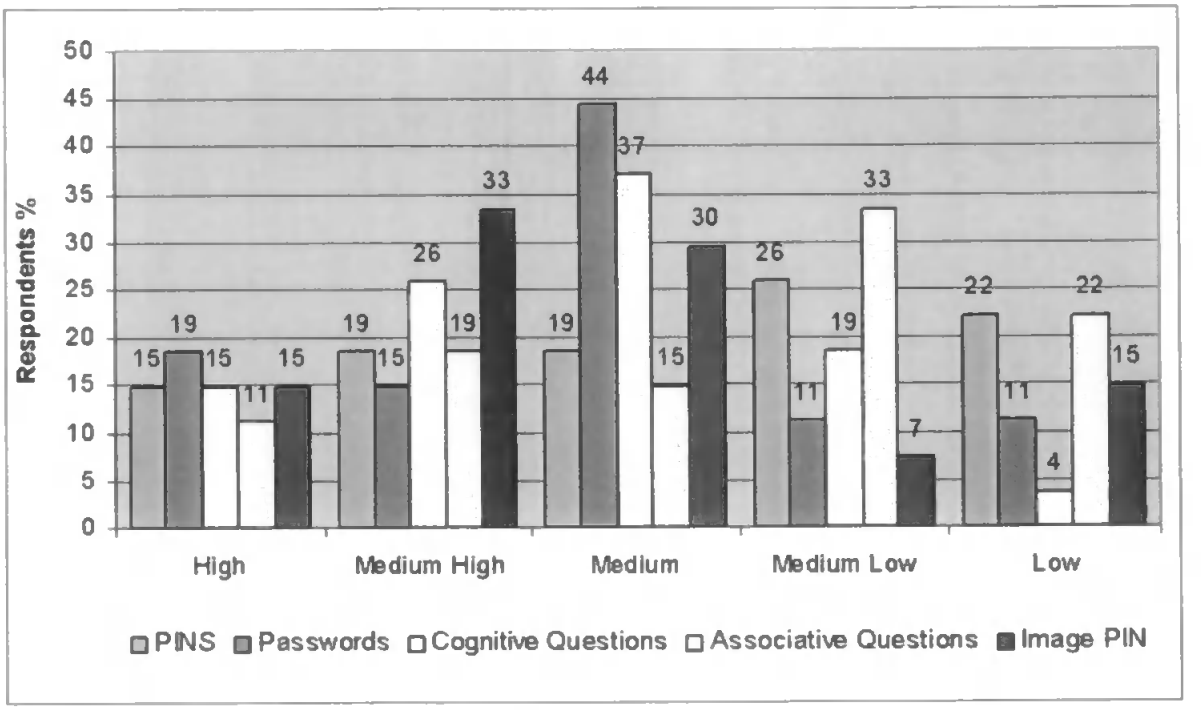


Figure 4.7 - Perceived security

The final question asked the participants to rank the methods according to their overall preference. The password method was again the most preferred form of authentication, with 44%, as shown in Figure 4.8. In second place is the PIN method with 22%, and third is the ImagePIN method with 19%. It is, therefore, demonstrable that the more traditional and familiar methods of authentication are still the most readily accepted. However, if the rationale behind the alternative methods is accepted (i.e. that passwords and PINs are open to compromise), then it is relevant to give further consideration to the results and responses in the other categories. This is not to suggest that the alternatives are themselves incapable of compromise – they will still be bound by the limitations of the implementation and the behaviour of the users.

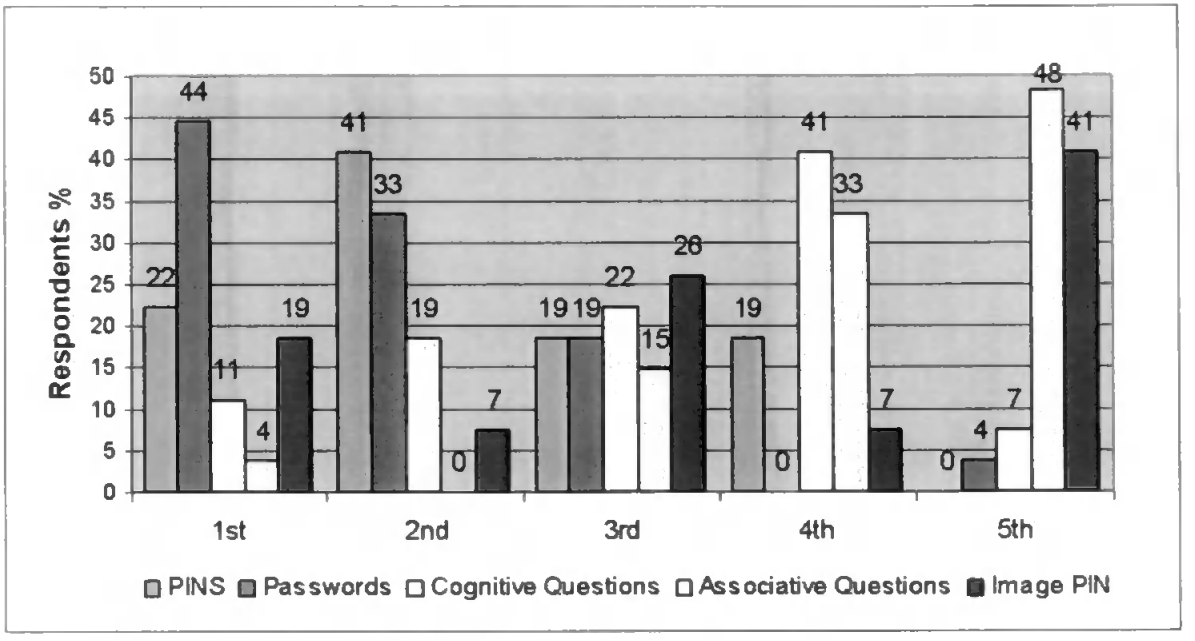


Figure 4.8 - Overall preference of trialled methods

4.2.5 Discussion

Although people clearly prefer passwords and PINs, the other results obtained continue to suggest concerns about the level of security they actually provide. For example, analysis showed that 48% of the participants selected passwords that might be easily guessed or cracked (e.g. based upon dictionary words, variations of their name, or foreign words written in English characters). Only 38% of participants used an alphanumeric combination, and fewer still (4%) introduced other symbols into their passwords. These results increase the attractiveness of the other methods, which may be less vulnerable to such unintentional compromise.

The participant's performance in relation to the cognitive questions was relatively strong, with 59% achieving successful authentication (interestingly, a previous study by Haga and Zviran (1991) reported better results, 74% success, for a broadly similar set of cognitive

questions). Further points noted about the cognitive technique were the relatively time consuming nature of the profiling phase, in which the participants had to provide answers for all 20 questions. In addition, several participants expressed concern about the nature of the information that was requested, and were reluctant to provide genuine answers to the questions during the trial for fear that the information might be accidentally divulged. Particularly notable questions in this respect were in relation to mother's maiden name (a commonly used identity verification question in other contexts, such as bank accounts), place of birth, and age of first kiss. Overall, however, this method was ranked relatively high in terms of perceived user-friendliness and security.

The associative approach proved to be weak as an authentication method, with the performance of the participants (4% success) suggesting that it cannot deliver an adequate level of effectiveness. It is considered that this poor performance can in part be explained by the fact that users still have to remember potentially abstract information (as opposed to the more recognition-oriented approaches of cognitive questions), placing more or less the same demand on their memory as the password method. In addition, the results raise questions over the level of security that the approach would provide – the fact that many participants chose the same word associations suggests that the method would be vulnerable to attackers attempting to guess the likely associations. At the very least, this requires that more care must be taken in the selection of the keywords, to ensure that none of them have obvious first-choice answers. A previous study of the same basic method reported a far higher success rate, with an overall average 69% recall after a period of three months (Haga and Zviran 1991). It must be noted, however, that there was a significant difference in the experimental procedure in this case, as participants were asked to select their own keywords, as well as the appropriate associative responses.

The ImagePIN approach demonstrated positive results in the authentication phase, with 63% success, placing it very close to the results observed for passwords. This result is partially explained by the findings from previous surveys, which have shown that people tend to have less difficulty in recognising previously seen pictures than they do in recalling passwords or phrases from the memory (Bensinger, 2000 & Sasse et al., 2001). In addition to its practical effectiveness, the ImagePIN scored well in terms of user acceptance, which was considered to bode well for the rating that it might receive if users were given additional time to familiarise themselves with it. Another point worth noting is that the ImagePIN method as implemented for the study was rather crude, with a set of standard Windows icons having been used as the selection of available images. With more consideration given to the number and range of images available, it was considered likely that the perceived user friendliness of the approach could be further improved. Having said this, there was also a fairly high proportion of respondents who put it as their clear least favourite, whereas most of the other methods did not elicit such strong negative opinions.

Although some techniques suggested themselves as potential alternatives to standard passwords and PINs, it does not necessarily follow that they would make good replacement methods in all contexts. For example, the use of cognitive questions could potentially be too time consuming as a regular means of login authentication. The technique could, however, provide a good secondary level of authentication, which could be invoked in a number of scenarios (e.g. when a user tries to perform a sensitive activity, in response to a suspected masquerade attack, or simply at random intervals). Image based authentication techniques could be more easily implemented as an initial login technique, but their

applicability would be limited to systems that are able to offer sufficient graphical displays (e.g. mobile phones, Personal Digital Assistants and Automated Teller Machines).

With the exception of the associative approach, the practical effectiveness of the techniques was closely comparable. However, in terms of the overall preference, the known and familiar methods of passwords and PINs were, perhaps unsurprisingly, favoured. However, as indicated previously, if passwords and PINs were used in a secure manner (in accordance with guidelines) it is likely that the acceptance rates for these methods would decrease. If the previous arguments and evidence regarding the weaknesses of these methods are accepted, then it may be reassuring to consider that the cognitive and ImagePIN methods are already comparably effective from a user recall perspective, and given further training and exposure these methods may gain greater acceptance.

Although the initial results are encouraging the judgements relating to user-friendliness of the methods were based on a relatively brief level of exposure to the question and answer approaches and the ImagePIN method. The next section describes a longer-term trial in which participants used two of the alternative methods in day-to-day operations, in place of their normal passwords or PINs.

4.3 A Longer Term Study Of Alternative Methods

While the initial trial had primarily assessed the effectiveness of the methods (i.e. the users' ability to recall the necessary information), the aim of the second trial was to

evaluate the long-term user acceptance of the most promising of the alternate approaches, namely cognitive question and Image-PIN, with each technique being used in practice for several weeks (the associative question approach was not considered due to its weak performance in the first trial). For example, some users who initially found one or other of the new methods attractive might change their view in the longer term, whereas others who might have initially preferred passwords could come to prefer one of the alternatives once they became more familiar with it.

4.3.1 The Profiler

The software implementation of the profiler was based upon that used in the previous study (described in section 4.2.1) with some cosmetic changes to the interface of the ImagePIN approach. The original implementation had used relatively small, iconic images, which some participants had commented were not very meaningful, and so it was considered that using larger, more colourful images would aid their recall. The use of larger images has some drawbacks; most significantly it increases the potential for "shoulder-surfing" (i.e. the act of observing a user *over the shoulder*). It would, therefore, be necessary to introduce further safeguards to ensure the security of the chosen images in a typical office environment. It would also be necessary to consider the selection of images carefully in a live environment to ensure there is no link between an end-users lifestyle/hobbies and the available image library. The images used for this trial are pictured in Figure 4.9.

Figure 4.9 - New image sets used for ImagePIN

4.3.2 The Authenticator

Having collected the necessary information for each participant, the two authentication phases worked in the same manner as the first trial (pictured in Figure 4.10). Following each login attempt, the program logged the user's results to a file. This enabled the research team to track the number of times that participants actually used the system, and how well they performed.

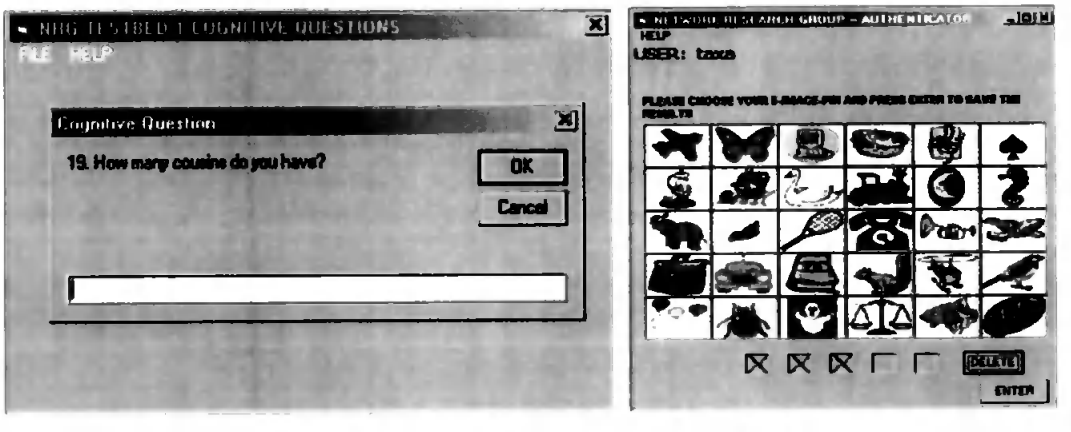


Figure 4.10 - Authenticator programs for cognitive questions and ImagePIN

For technical reasons (relating to the difficulty in interacting with the underlying security of Windows NT) it was not possible to replace the actual login authentication mechanism at the system level. However, in order to encourage users to make regular use of the

mechanisms on trial, the execution of the authenticator program was added as a scheduled task on participating systems. This ensured that the program would run automatically and remind the user to participate. However, in order to ensure that the program did not become unnecessarily intrusive (e.g. interrupting the user in the middle of important work), there was nothing to prevent users from terminating the program without actually entering any information. It is possible that some users may have chosen to close the program rather than respond to a specific, inconvenient, authentication challenge; however, the benefits of allowing users to terminate an inconvenient challenge in this trial were considered to outweigh the drawbacks of alienating test subjects.

Following the authentication trial periods, the acceptability and robustness of the techniques was assessed by means of a questionnaire. Key points of information collected included indications of:

- the perceived user friendliness of the techniques;
- the perceived security of the techniques;
- which, if any, of the trialed techniques could replace existing ones.

The trials were conducted for a total of ten weeks, split equally between the two alternative methods. 19 participants took part in the first stage, but unfortunately one user was not able to provide responses to the second stage and the subsequent questionnaire. The participants were primarily male post-graduate research students, in the 25-30 year age group. Although this represented a fairly limited sample from which to draw definitive conclusions about the methods, it was considered to provide a suitable basis for determining general results to accompany those from the initial study. In addition, the fact

that all participants came from a computing background meant that they were well-versed in the use of the traditional authentication mechanisms against which they would be asked to contrast the new approaches.

4.3.3 Experimental Results

Despite concerns, it was discovered that each of the participants made a suitable degree of use of each approach (rather than simply choosing to close the authenticator application), with the cognitive question method being attempted an average 25 times per user, and the ImagePIN being used an average of 21 times per user. These figures correspond to approximately one use per participant per day, and are considered to be a sufficient basis for the participants to offer more informed opinions about the suitability of the methods as they would have had a similar level of exposure to the new techniques as currently experienced with typical username/password authentication.

Before considering the opinions, it is relevant to consider how effective the methods proved to be from an authentication perspective. In the cognitive question trial, the criterion for successful authentication was to correctly answer all five of the questions asked randomly by the program. The results indicated that 64% of the participants had succeeded to authenticate themselves using the cognitive question method. This was actually a slight improvement on the effectiveness observed in the original study, in which the cognitive question approach scored 59%. Nonetheless, the results still showed that some users had problems in memorising the required information, even though it was related to personal details. Figure 4.11 shows how the distribution of correct answers broke down across the sessions. Unexpectedly, it was founded that factual questions were

the ones that participants often answered wrongly. For example, questions 2, 8, 10, 12, and 19 were mostly answered incorrectly (refer to Table 4.1). A possible reason could have been the case sensitivity of the program; for example, some people may have entered words in upper case, but then entered them in lower case during authentication and were rejected. A further explanation could be the potential for variation in answering questions, for example, the question "What is your favourite music?" could be answered with a group name, genre, track or artist.

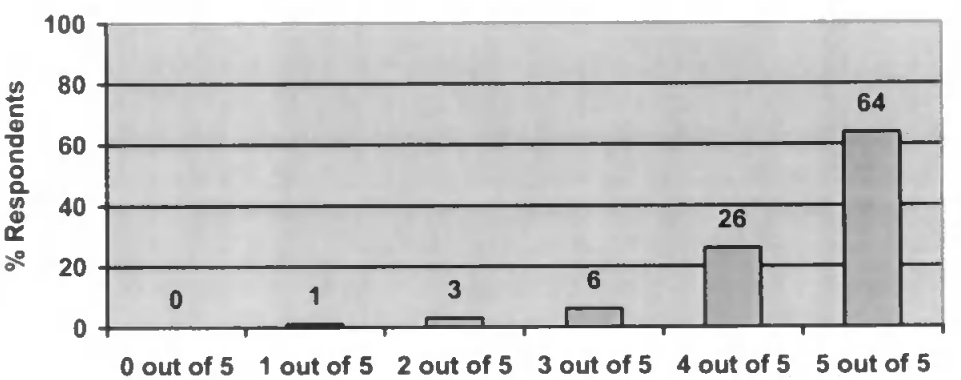


Figure 4.11 - Distribution of correct answers for the cognitive questions method

When considering the results for the ImagePIN, the criterion for successful authentication was the correct recall of the images in the same order they were selected in the profiling stage. The results, shown in Figure 4.12, indicated that 84% of the participants had succeeded in authenticating themselves.

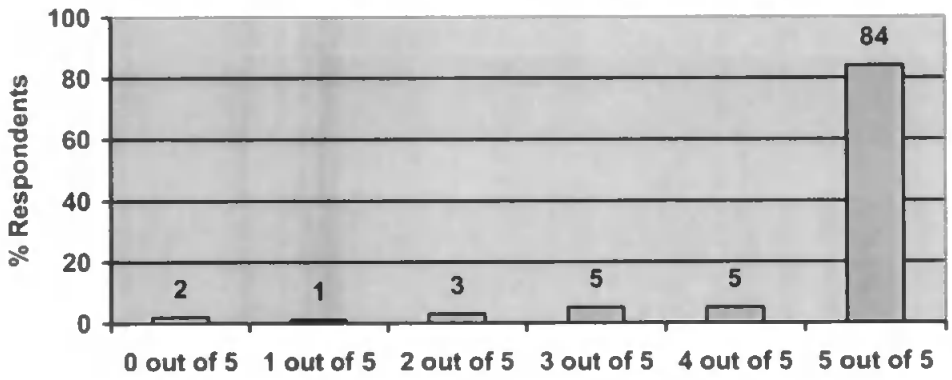


Figure 4.12 - Distribution of correct answers for the ImagePIN method

It can be noted that the participants' overall performance was again notably better than in the original study (84% versus the previous ImagePIN result of 63%). One reason for the improved result (for both this method and the cognitive questions) could have been that the participants made use of the methods very soon after creating the profile, whereas in the first study they had been forced to wait a month (in order to evaluate their recall of the information). However, another contributing factor to the high success of the ImagePIN was that two thirds of the participants had chosen potentially weak sequences. Examples here included a sequential line of images, in the same horizontal row or vertical column, or selecting five instances of the same image (approaches which clearly made the information easier to remember). Possible solutions to this would be to prevent selection of duplicates, and to reposition the images each time the program runs. However, there would be a strong possibility of this reducing the success rate.

Following the trial period, the eighteen remaining participants were asked to provide answers to the evaluation questions. They were firstly required to indicate whether they believed that either of the new techniques could replace the existing ones (i.e. passwords or PINs) for login authentication. The responses revealed that 56% of users supported this

idea. However, as Figure 4.13 shows, the two techniques did not meet with equal approval, with the cognitive questions approach being clearly preferred.

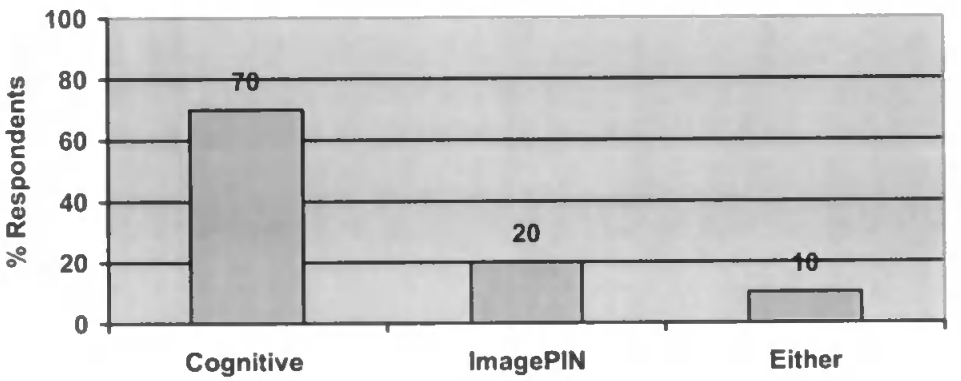


Figure 4.13 - User preference for replacement method

Continuing the evaluation, participants were asked to assess the techniques in terms of their ease of use, using a four-point scale, progressing from "Very easy" to "Very hard". The results are shown in Figure 4.14, conveying the impression that the overall difficulty associated with the cognitive questions approach is less than that for the ImagePIN (with none of the participants rating the cognitive approach as 'very hard'). By contrast, the ImagePIN results suggest that the method is more likely to elicit strong feelings from the users, with many more responses focusing at the two extremes of the scale.

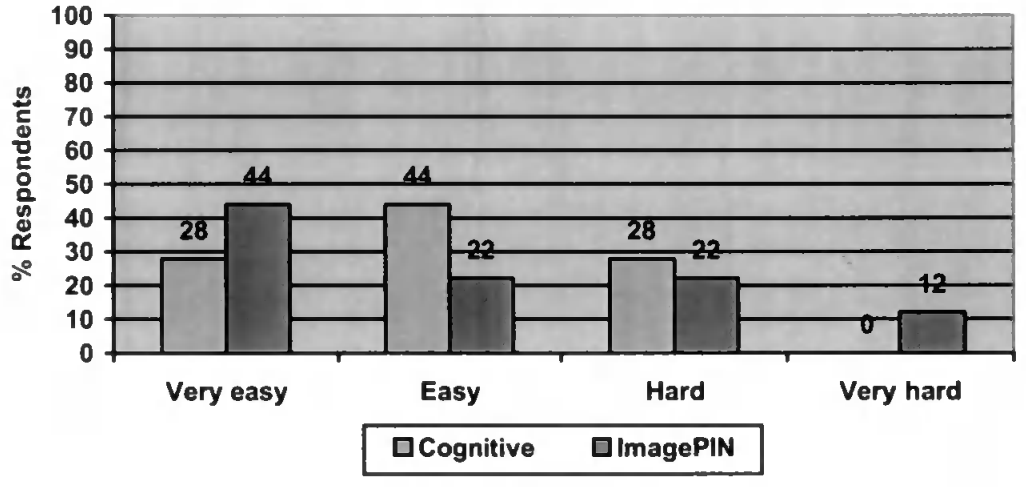


Figure 4.14 - User ratings for perceived ease of use

Given that one of the principal problems with traditional passwords and PINs is that people often find them hard to remember, it was relevant to consider this aspect with the new methods. When asked if they had difficulty in remembering the required information, the participants responded as shown in Figure 4.15. This indicates that users have found the cognitive questions most difficult, contrasting with the earlier findings from Figure 4.13, which indicated that 70% preferred this method. It should be noted that the results in Figure 4.15 are based on the full population of participants, whereas Figure 4.13 was only based upon 56% of them as participants were only asked to answer **if** they believed the existing password/PIN techniques could be replaced.

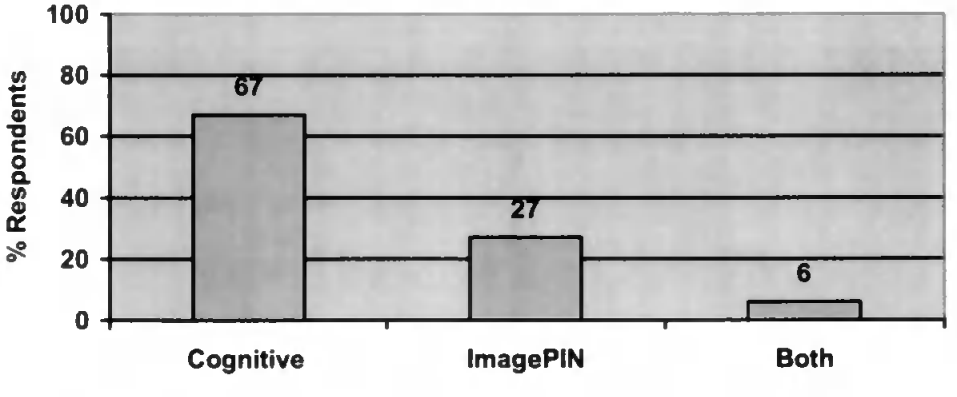


Figure 4.15 - Perceived difficulty to remember the required information

Another important factor was the resulting security that the replacement techniques were perceived to provide. The reason for considering the replacement of existing passwords and PINs is that they have been found to be vulnerable to compromise by legitimate users, and to attack by impostors. In order to evaluate the perceived security of the selected methods, users were asked to rate them on a four point scale, progressing from "Very easy" to "Very hard", when considering the possibility of an impostor being able to guess the required information. The distribution of the responses is shown in Figure 4.16.

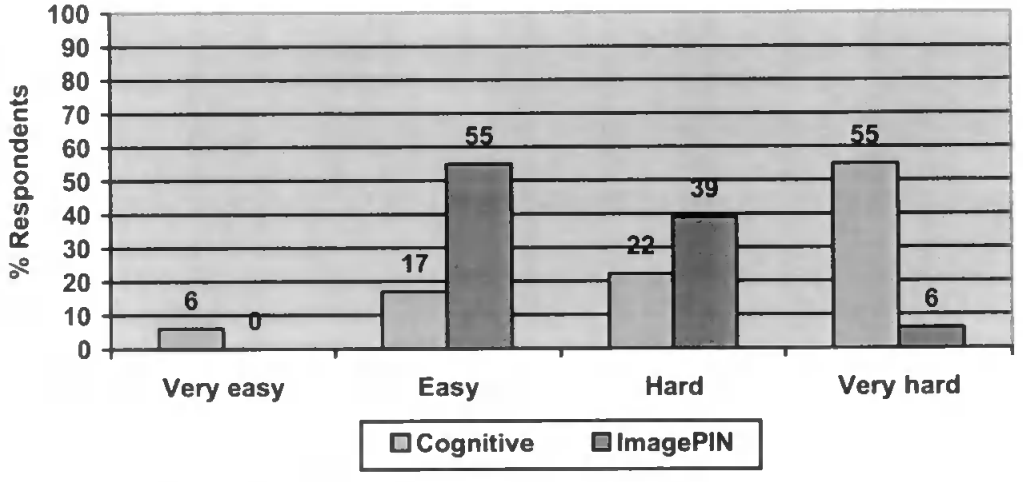


Figure 4.16 - User ratings for ease with which methods could be broken

The results show that 55% of the participants believed the ImagePIN to be the easier to guess which contrasts with the earlier findings (Figure 4.7) where 47% of the participants perceived ImagePIN's to have a high/medium-high level of security. Their feedback suggested that it would be easy for an impostor to find the appropriate sequence by sitting behind the user, or simply by guessing possible sequences. The latter observation was possibly reflective of the fact that many users had chosen weak sequences, and considered that other people would do likewise. A larger variation was found in the earlier results for cognitive questions where 55% perceived a low/medium-low level of security compared

with 77% who later considered the method to be hard/very-hard to be bypassed. This perhaps indicates a change of opinion following prolonged exposure to the method.

In order to further, evaluate the Image-PIN method participants were asked to assess the difficulty to remember the 5-images selected from the image grid provided. The participants were asked to assess the method on a four-point scale, progressing from "Very easy" to "Very hard". The results are shown in Figure 4.17. While over 60% of users considered the approach to be easy/very easy, a significant number considered the approach to be hard/very hard. This approach may be viable as an authentication technique, however, more work must be done to make this method more user-friendly.

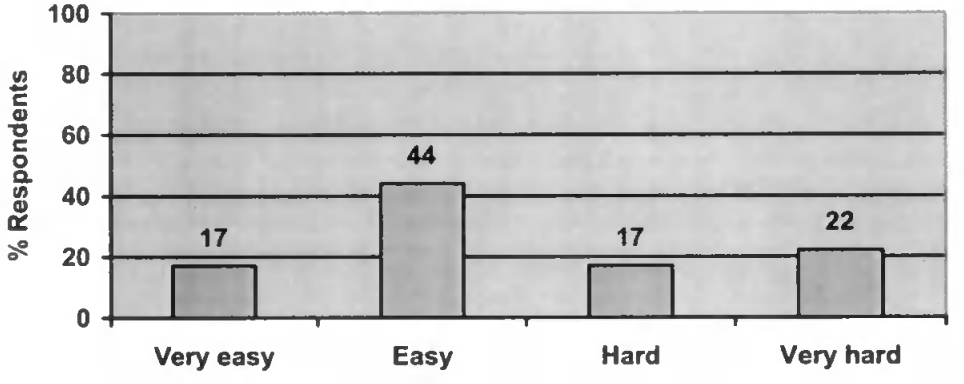


Figure 4.17 - Perceived ease of remembering a 5-image sequence

The final question asked the participants if they would be prepared to use the methods for login authentication. Findings indicated that 56% of users would be prepared to use the cognitive questions method for login, and 39% of users for the Image-PIN method respectively (Figure 4.18).

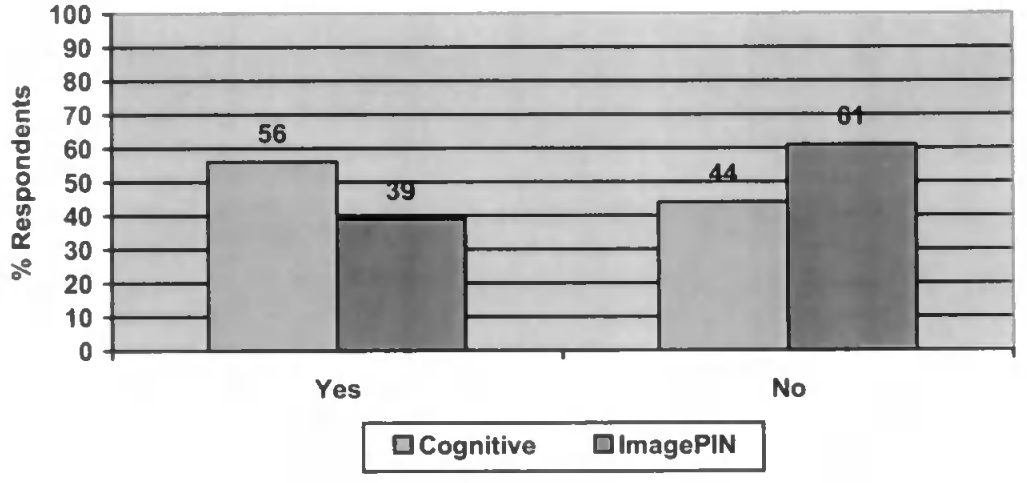


Figure 4.18 - Overall user acceptance of the techniques

Most of the users that rejected the cognitive question method for login authentication did so on the basis that the method was time consuming and that they had problems relating to the entry of the data. By contrast, most of the users that rejected the ImagePIN method expressed concerns regarding the interface design and the selection of the images, as well as security implications. With such a high level of user rejection for the evaluated techniques, it is clear that further work is required to improve the user perception/practical issues as well as addressing the technical aspects. It can be observed that for the user community as a whole, neither technique represents the natural successor to passwords. However, the results do suggest that they offer some potential in some contexts.

4.3.4 Discussion

The participant's comments about the methods highlighted some further interesting points that were not apparent from the statistics alone. In some cases, these opinions raised further legitimate questions about the viability of the techniques, whereas in others they

flagged issues that would be unlikely to be genuine considerations in practice. Nonetheless, the fact that such comments were made even within a small sample population suggests that similar misconceptions could also be reflected within a wider user community.

In the case of the cognitive questions method, individual opinions suggested that it would be easy for impostors to acquire the required information by social engineering (e.g. having brief conversations with the user, or friends of the user in order to find answers to the questions). Set against this, however, is the fact that an impostor must know all the answers in order to guarantee successful access to the system, because the questions are presented randomly to the user. In addition, even if the attacker found the answers, they must also know how they were typed in the profile stage (e.g. for the question "How many cousins do you have?", did a user with three cousins enter their answer as "Three", "three", or "3").

A problematic aspect of the cognitive question approach was that it was considered to be too time consuming, ruling it out for login authentication, and suggesting that it should only be used where higher security requirements justified the effort. Another valid point was that basing the authentication on factual information meant that answers would be impossible to change if the method was compromised (i.e. users cannot simply change their answer to a question such as "What is your mother's maiden name?" in the same way that they could change a password). This would mean that the questions themselves would have to change as well. As such, investigation of how well users are able to adapt to changing questions would be another valid avenue of further research arising from this study.

In the context of the ImagePIN, there were several comments about the user interface. Participants suggested that the method should have more image grids or different images for selection, as well as better images, and the option to choose individual images in order to complete the grid. Addressing these factors would have the potential to improve the level of user acceptance. Further potential refinements could consider the random positioning of images (as previously mentioned), or the use of a static subset of images from a much larger collection so that users could select their preferred images (i.e. the user would select 20 images from a large collection and then select their ImagePIN from the subset – for authentication, only the subset would be presented, this would allow a user to select similar images that are only distinct to them). If a much larger set of images were available, authentication could be conducted with a randomly selected range of images (i.e. 4 or 5 valid images with a number of random invalid images). Another refinement could involve selecting each image individually from a random group of images. This would increase the number of invalid images for each part of a 4 or 5 image PIN and thus increase the number of combinations (e.g. a 5x5 grid of images shown five times would give a total of nearly ten million combinations). However, randomising the images in this manner could adversely affect the usability and acceptance of the technique as there would be an increased delay while the user identified the position of the image(s). A potential solution to this would be to reduce the number of images presented (with a subsequent impact on the level of security of the method).

Although many participants appeared to consider the ImagePIN vulnerable to compromise by guesswork, it is worth noting that the probability of a properly selected sequence being broken by this means is 0.000000041. However, the clear problem from practical experience was that many users did not select appropriate sequences, leading to the

conclusion that the implementation of such a technique could be a high risk for security-unaware users. As with all access control systems, if the authentication mechanism is too difficult or inconvenient to use, end-users will find ways to subvert the system.

Given the inherent difficulties in the above methods (spelling mistakes, forgetting images etc.), all of the proposed methods could suffer from some of the same issues that make passwords weak, namely that users will forget their login details and may have to write them down. Preventing users from doing this may increase the loading on administrators with users requesting ImagePINs to be reset, or answers to be retrieved. To resolve this, users need education in appropriate selection of their authentication responses.

4.4 Conclusions

The two trials described in this chapter have provided interesting results regarding the use of alternative secret-based authentication techniques. The first trial suggested that longer exposure to the methods would allow a more accurate impression of user acceptance to be gained. The second study demonstrated that longer exposure to the methods can improve both acceptance and the rate of successful authentication, suggesting that each of the techniques offers some potential as a replacement or supplementary authentication mechanism. However, none of the evaluated techniques appeared to represent an ideal solution that could consequently be used to replace traditional passwords and PINs for all users in all contexts. In addition, full-scale implementation of the techniques could still be hampered by the ways that people can potentially misuse them.

In parallel with considering alternative authentication methods, there is also a need for improved user education, both in terms of selecting appropriate information to act as their authentication secret, and also in appreciating the level of security offered by different techniques. The solution to login authentication lies not just in technical implementations, but also in the way end-users interact with the measures employed.

The results of these trials suggest a number of further research projects that will be discussed in more detail in chapter 9.

Having evaluated these approaches, it is clear that software-based approaches are popular with users and could compare favourably with passwords if implemented properly. The results do, however, indicate that users still had difficulty in selecting both unique and non-predictable secret information. This again serves to reinforce the fact that users frequently undermine authentication systems through misuse (albeit often inadvertent). In order to build on the use of software-based techniques, whilst moving away from secret-based methods, the next chapter presents the use of keystroke analysis as an alternative authentication mechanism to the traditional username/password, before considering a similar trial to evaluate the use of keystroke analysis in chapter 6.

Chapter 5

Approaches to Keystroke Analysis

5.1 Introduction

Chapter 3 identified the need for a method of transparent continuous user authentication, with the respondents indicating a preference for a method based upon keystroke analysis. Before looking at the keystroke analysis methods, alternative software-based authentication techniques were evaluated and described in the previous chapter – this chapter determined that alternative secret-based authentication has potential but may best be applied as a response mechanism for specific authentication challenges. This chapter summarises the potential approaches to keystroke analysis, presents a novel method based on application-specific user profiling and considers the use of multiple metrics to create a composite supervision system.

The concept of keystroke analysis is by no means a recent development. Previous work, published in 1980, first identified the profiling of key-presses as a potential method of user authentication (Gaines et al, 1980). Since then a number of research projects have been conducted to evaluate different methods of data gathering (using a range of operating systems and considering a variety of metrics) and post-processing techniques (ranging from purely statistical to AI/neural network approaches). Later in this chapter these projects are summarised and compared.

Before looking at the results of previous work in this area, it is first necessary to determine which characteristics of typing are viable for profiling and authenticating against. Previous studies have identified a selection of data acquisition techniques and typing metrics upon

which keystroke analysis can be based. The following section summarises the basic methods and metrics that can be used.

5.2 Metrics

There are a variety of possible keystroke metrics that can be profiled as the basis for subsequent comparison. The main methods are based on timings between consecutive keystrokes, and consider either the latency between two consecutive keypresses (digraphs) or three consecutive keypresses (trigraphs). Other possibilities include the mean typing and error rates.

- **Digraph latency** - Digraph latency is the metric that has traditionally been used for previous studies. This measures the delay between two consecutive keypress events that are produced during normal typing (e.g. when typing the word 'THE' two digraph timings can be generated - T-H and H-E). Given a suitable volume of digraph samples the character distribution of the English language will ensure that a range of commonly occurring digraphs will be generated that can subsequently be profiled. In most cases, some form of low and high pass filter is applied to remove extraneous data from the session/profile data (discussed in section 5.5.1).

- **Trigraph latency** - Trigraph latency extends the previous metric to consider the timing for three successive keystrokes (e.g. T-H-E). Spaces are usually ignored – e.g. the word THERE could generate three trigraph samples, T-H-E, H-E-R and E-R-E, with the final trigraph R-E-[space]

ignored. As with digraph latencies, the application of a low and high pass filter is usually required. It should be noted that the occurrence of trigraphs (and digraphs) would normally comply with the rules of English – i.e. T-H and T-H-E are likely to be the most commonly occurring. However, as new words (and in particular acronyms) are introduced these distributions may change. For example, the introduction of the world-wide-web has introduced the trigraph W-W-W that would have been unlikely to occur twenty years ago. It is probable that over time, the distribution of digraphs and trigraphs will change and as such, any system that relies upon these measures will also have to change (i.e. profiles will have to undergo periodic refinement).

- **Keyword latency** - Keyword latencies consider the overall latency for a complete word, or may consider the unique combinations of digraph/trigraphs in a word-specific context. The use of keyword latencies allows not only profiling of commonly occurring words, but also the ability to monitor words that could be interpreted as commands to the operating system or applications running on the system – i.e. a command with a high misuse risk (e.g. delete or format) could be specifically monitored or a specific function in an application (e.g. entering an application-specific command with a high risk consequence).

- **Keystroke duration** – Keystroke duration considers a different metric to that of the digraph and trigraph latencies. This approach considers the duration of each individual keypress (i.e. the time between the key-down

and key-up events for a single key as opposed to the latency between two/three key presses). While this approach is worthy of investigation, it does introduce significant limitations. 26 letters of the alphabet provide far fewer profile-able characteristics than digraphs – with a maximum of 676 (26*26) discernible values. The actual number of digraphs encountered in typed language is likely to be less than this as certain digraphs do not commonly occur (e.g. ZZ, QA). It is also likely that the keypress duration will be less variable than digraph latency as there would be no hand movement or changes between hands in the keypress duration – unlike the digraph latency.

- **Keystroke pressure** – Keystroke pressure relates to the level of pressure applied for each keypress. It is not possible to obtain pressure values from a standard computer keyboard, and as such the use of a customised keyboard specifically designed to produce pressure values in addition to the usual keypress would be required. While this approach may have some potential, the need for a modified keyboard makes this inappropriate for practical use.

- **Mean typing rate** – Whilst this may not be user specific, it may be possible to classify users into a generic category, according to their typing ability, which can then be used as an additional authentication method or potentially to set filter thresholds.

- **Mean error rate** – Finally, the mean error rate can be used to provide an indication of the competence of the user during normal typing. It might also be feasible to evaluate a users' typing errors – e.g. certain users may mistype the same word or words consistently. As with the mean typing rate, individuals could be classified according to their typing ability and hence evaluated based on their average typing accuracy.

While the final two metrics indicated above are unlikely to provide a suitably fine-grained classification of users for direct authentication judgements, they may be used to provide a more generic set of user categories that can contribute to a combined measure. The environment in which they are used will determine the usefulness of such non-specific metrics. For example, monitoring typing speed and error rate within a pool of touch-typists would be of little value, as there is unlikely to be any significant deviation between each user's profile.

5.3 Collection methods

In addition to the variety of metrics that can be recorded, there are also variations in the methods of data collection. The following list presents a number of ways in which user-typing patterns can be acquired and subsequently used for authentication purposes.

- **Static at login** – Static keystroke analysis authenticates a typing pattern based upon a known keyword, phrase or some other pre-determined text. The captured typing pattern is then compared against a profile previously

recorded during system enrolment. Static keystroke analysis is generally considered to be an initial login enhancement as it can supplement the traditional username/password login prompt, by checking the digraph latencies of the username and/or password components (i.e. authenticating the user on the basis of both what they typed and how they typed it).

- **Periodic dynamic** – Dynamic keystroke analysis authenticates a user on the basis of their typing during a logged in session. The captured session data is compared to an archived user profile to determine deviations. In a periodic configuration, the authentication judgement can be intermittent; either as part of a timed supervision, or, in response to a suspicious event or trigger. This method provides distinct advantages over the static approach. Firstly, it is not dependent upon the entry of specific text, and is able to perform authentication on the basis of any input. Another factor is the availability of data; in static keystroke analysis, the range of digraphs and frequency of their occurrence is likely to be significantly limited compared with a dynamic approach. Even an inexperienced typist is likely to produce sufficient digraph pairs to allow an authentication judgement to be derived. This is an important factor as it is necessary to have a large volume of keystroke data in order to generate a user profile.

- **Continuous dynamic** – Continuous keystroke analysis extends the data capturing to the entire duration of the logged in session. The continuous nature of the user monitoring offers significantly more data upon which to base the authentication judgement. With this method it is possible that an

impostor may be detected earlier in the session than under a periodically monitored implementation. On the downside, however, the additional processing required will add to the computational overhead of the supervision system.

- **Keyword-specific** – Keyword-specific keystroke analysis extends the continuous or periodic monitoring to consider the metrics related to specific keywords. This could be an extra measure incorporated into a monitoring system to detect potential misuse of sensitive commands. For example, under a DOS/Windows environment it may be appropriate to monitor the keystroke metrics of a user attempting to execute the FORMAT or DELETE commands. This could represent a significant enhancement, as a command with a high misuse consequence (e.g. DEL *.*) is unlikely to cause sufficient profile deviation when observed from a system-wide context, due to the limited selection of digraphs. By contrast, static analysis could be applied to specific keywords to obtain a higher confidence judgement.

- **Application-specific** – Application-specific keystroke analysis further extends the continuous or periodic monitoring. Using this technique, it may be possible to develop separate keystroke profiles for distinct applications. For example, a user may be profiled separately for their word processing application and email client as a user may type sporadically in response to emails while word processing generally has

prolonged periods of continuous typing. The potential of this new technique is discussed in more detail in section 5.7.2.

It should be noted that all of the above techniques and metrics could be implemented on a standard PC platform, without the need for special hardware (e.g. the only requirements are a PC and the ability to run the monitoring software). This offers a significant advantage over other alternative authentication and supervision techniques (discussed in Chapter 2), as a standard keyboard is present on almost all PC's. However, this approach is not without its drawbacks and thought must be given to the ethical and legal questions that may arise when proposing such close supervision of computer users.

5.4 Methods of implementation

The actual methods of collecting and subsequently verifying user keystroke data vary depending on the operating system on which the collection is to take place and also upon the nature of the characteristics being monitored.

There are three potential methods to obtain keystroke data:

- Custom application – this method can be used to provide a custom front-end through which a user's typing patterns can be assessed programmatically or added to an existing application to provide context-sensitive analysis of user interaction. This could, for example, be used to demonstrate the concept of keystroke analysis by presenting the user with a user interface in which they

would be required to either type pre-determined text (e.g. a password) or allowed to type freely. Depending on the operating system, this may require substantial coding to obtain the necessary signals indicating the keypress events. This method is limited in its operation, as it will only monitor typed content within the specific application. As such, its use is restricted to enhancing security within an application or to act as a response to a security issue – i.e. to request the user to authenticate with a stronger confidence than password only.

- Modification of operating system – this method is considerably more involved than the previous approach as it requires an understanding of the underlying operating system. Using this approach it is possible to replace the login authentication presented under Microsoft Windows (by modifying the GINA.DLL that provides the graphical user interface for the login prompt as part of the Graphical Identification and Authentication services under Microsoft Windows). This process is applied in the Windows software product BioPassword described in chapter 2 and illustrated in Figure 5.1. This method has distinct advantages in that no specific user action is required in order to authenticate as the process of keystroke analysis is integrated into the traditional login prompt. As with the previous method, this approach is limited in its scope as it would only be activated on login (or when the terminal is locked or the screensaver deactivates – assuming such options are available and are enabled).

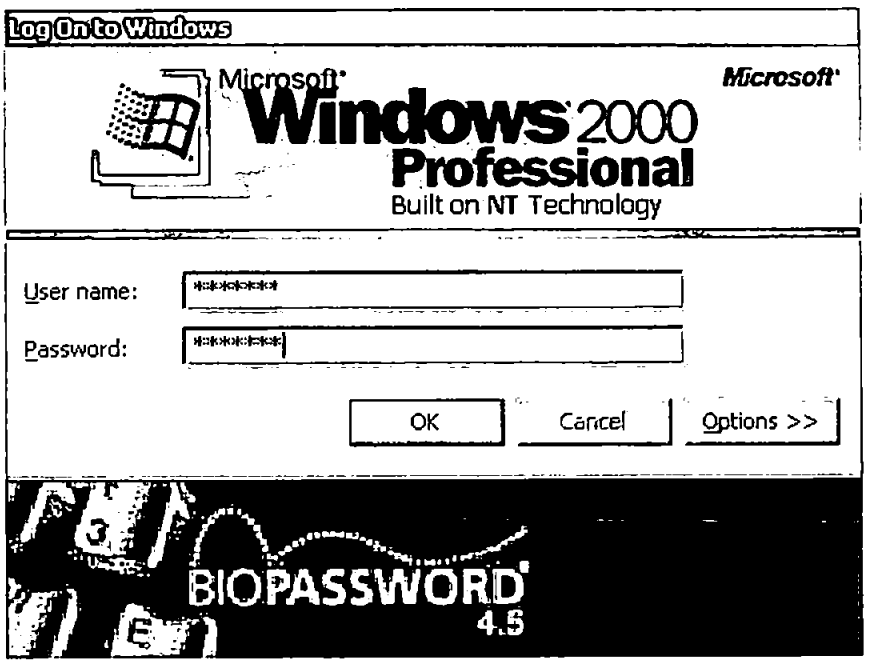


Figure 5.1 - Biopassword login screen

- Capturing keystroke data – this method uses a custom application to effectively *sniff* keystroke activity by intercepting keystroke messages (under Windows) or by redirecting pipes (under Unix/Linux). The appropriate software to enable logging of the keystrokes then processes these messages. This approach is considered the most flexible (and hence appropriate) of those outlined in this section because it can be implemented in a completely transparent manner. This approach will be discussed in more detail in chapter 6 where a prototype system is described.

5.5 Processing Keystroke Data

Once the keystrokes have been obtained (using one of the methods outlined in the previous section, it is necessary to filter the keystrokes (to remove low and high latencies) and to then perform post-processing to extract data on which to base a user profile.

5.5.1 Filtering

For most implementations, some form of low and high pass filter is required to remove extraneous samples from the session/profile data. Low pass filtering is important as it ensures that undesirable typing characteristics (e.g. stuck keys or repeated key-scans) are eliminated from the profile. The use of a high pass filter helps to ensure that excessively long digraph timings (which could be indicative of a distraction during typing) are also eliminated. The use of such filters could of course have negative side effects – i.e. an impostor or masquerador could attempt to bypass security (given appropriate knowledge of the system) by typing commands so slowly that the high pass filter would ignore the majority of keypresses. To remedy this, it would also be necessary to monitor the proportion of *filtered* digraphs to ensure that a number of consecutive incidents are flagged.

In addition to the filtering of low and high timings, intelligent filtering is usually needed (for all methods) to remove certain control keys (i.e. the function keys F1-F12, CTRL and ALT keys etc.) as these are unlikely to provide any meaningful contribution to the profiled samples.

5.5.2 Post processing and comparison

This section considers the issues of post-processing and comparison from a digraph-centric perspective. However, it should be noted that all the points raised are equally relevant to trigraph and keyword-based profiling.

Once the data is gathered it is necessary to process the raw keystroke samples to provide a comparative profile against which session data can be compared. The production of a profile is not an insignificant task. If the approach being taken is purely statistical, the data must first be cleaned to remove sample outliers that may affect the range of valid values for each profiled digraph/trigraph – to tighten the distribution for each profiled digraph (as described in the previous section). Once the data is cleaned, the profile is generated based on the valid data obtained from the user. One method (described in more detail in chapter 6) involves the production of a mean and standard deviation for each digraph recorded (although there are other statistical techniques employed by the studies described in the following section). Once a reference profile has been created, subsequent user data can be compared on a digraph-by-digraph basis against the reference profile, recording the deviation from the expected typing patterns. The output of such a comparison could be a simple count of accepted versus unaccepted digraphs, but could also consider the number of consecutive matched, unmatched, accepted and rejected samples (matched and unmatched refers to samples where a valid digraph profile does or does not exist – i.e. the user profile may not have a recorded entry for a specific digraph pair).

An alternative to the statistical approach is to use a neural network method. To use a neural network approach, the data is split into two parts. The first part of the data would be used to generate a profile – typically selecting the top *n* occurring digraphs for the user and then feeding each digraph timing through a neural network. Once the network is trained using the first data set, the second set can be used to verify and improve the network's performance. Finally, the data from other users can be fed through the same network to determine the false acceptance rate for the users' profile. The output of this method would

normally be a classification accuracy – i.e. the confidence that the sampled data matches the expected users' profile.

Both of these methods have been considered by a number of trials investigating the use of keystroke analysis utilising digraph samples (a selection are listed in Table 5.1). However, there are significant problems with the use of either method. A purely statistical approach relies upon sufficient data being available to allow a profile to be generated – if too little data is present, the accuracy of the profile will be jeopardised as the deviation for each profiled digraph could be very high (more samples are needed to make the high/low timings less significant). The neural network approach requires a reduced subset of digraphs (typically choosing the *n* most commonly occurring) as any increase in the number of inputs to the neural network (where one input is required for each profiled digraph) increases the complexity of the networks and the time taken to process the data. This results in a trade-off between reducing the number of inputs to speed up the data processing while ensuring enough data to accurately authenticate a user.

The neural network method, while limited by the volume of data collected in keystroke analysis, may be able to identify patterns in user profiles (i.e. identifying distinguishing digraphs for each user) that would be missed by a purely statistical approach. While not considered in the first trial described in chapter 6, the use of neural networks will be revisited in chapter 8.

5.6 Summary of previous work

The idea of using keyboard characteristics for authentication is not unique, and there have been a number of previous published studies in the area. To date, however, virtually all published studies have focussed upon static or context-independent dynamic analysis, using the inter-keystroke latency timing method. From the earliest studies in 1980 (Card et al & Gaines et al), the focus has been on the analysis of digraph latencies. Later studies, such as those by Joyce & Gupta (1990) and Mahar et al (1995) further enhanced the work, identifying additional statistical analysis methods that provided more reliable results.

In Legget et al. (1991), the concept of dynamic keystroke analysis was first proposed, with the introduction of a reference profile that could be used to monitor a live user session. Brown and Rogers (1993) also explored the idea of dynamic analysis, presenting preliminary results.

A summary of some of the main results from studies to date is presented in Table 5.1 below, which illustrates the effectiveness observed (in terms of false acceptance and false rejection errors), as well as the type of keystroke analysis technique employed (digraph/trigraph etc.) and the analysis approach taken (statistical or neural network).

As can be seen from Table 5.1 the range of results shows the inherent unreliability of the keystroke analysis approach – with large variations in both FAR and FRR rates. It can be observed that in almost all cases, the FAR rates have either been fixed at 0% (optimising the system under test to reject all known impostor activity) or have been less than 10% (less than 1 in 10 impostors would be authenticated by the keystroke analysis system).

However, with some experiments producing FAR rates in excess of 10% it is clear that the previous work in the area has still not entirely addressed the problem of false acceptance. As different organisations will have differing security requirements, there is no single FAR rate that will prove acceptable to all – while all organisations will aspire to 0% FAR, it is unlikely to be feasible due to the inverse relationship with the FRR rate (Figure 2.2) and the consequent affect on user attitudes and opinions when high rejection rates are encountered. Several studies have forcibly fixed the FAR to 0% by optimising the software to reject all impostor attempts during the experiment run. However, this can only be achieved at the expense of the false rejection rate. By fixing the FAR at 0%, the system has to be much more precise about accepting a login attempt – this results in an increased FRR.

Authors	Method	Participants	%FAR	% FRR
Umphress & Williams (1985) Static	Digraph Statistical	17	6%	12%
Legget & Williams (1988) Static	Digraph Statistical	17	5%	5.5%
Joyce & Gupta (1990) Static	Digraph Statistical	33	0.25%	16.67%
Bleha et al. (1990) Static	Digraph Statistical	10 (profiled) 26 (comparison)	2.8%	8.1%
Legget et al. (1991) [1]Static, [2]Dynamic	Digraph Statistical	17	5% [1] 12.8% [2]	5.5% [1] 11.1% [2]
Brown & Rogers (1993) [1]Group 1, [2]Group 2 Static	Digraph Combined Neural Network & Statistical	25[1] / 21[2]	0%	4.2% [1] 11.5% [2]
Napier et al. (1995) Static	Digraph Statistical	67	29.5% / 3.8% (FAR + FRR)	
Mahar et al. (1995) Static	Digraph Statistical	67	35% / 17.6% (FAR + FRR)	
Furnell et al. (1996) [1]Static, [2]Dynamic	Digraph Neural Network [1], Statistical [2]	26	8% [1] 15% [2]	7% [1] 0% [2]
Guven and Sogukpinar (2003) Static	Digraph Statistical	12	1%	10.7%

Table 5.1 - Previous keystroke analysis studies

The effect of an increased FRR is not so obvious when considering the classic C-I-A trio (Confidentiality, Integrity and Availability) as usually the only impact is on the latter, availability. Forgetting or mis-typing a password is a common occurrence and as such users may be somewhat forgiving of a system that asks them to re-type their password. However, if the supervision system is reliant on keystroke analysis it is not *what* the user types but *how* the user types that is important. A user is likely to become annoyed and confused at having to retype the same authentication details knowing that they are typing in the correct information. If the user is unaware that the system is also monitoring how they are typing this could result in poor acceptance of such techniques. As such, it is important to attain a balance between the two, or, to optimise the FAR to 0% whilst minimising the FRR to as near to 0% as possible.

There is also an issue of logged data size with the experiments summarised in Table 5.1. It is difficult to make direct comparisons based solely on the error rates without considering the size of the acquired data sets. While all the experiments listed provided details of the number of participants (ranging from 10 to 67), none provided details of the size of the sampled data. Without this information, the comparisons made between studies can only be based on the statistics available; namely the error rates, analysis method and number of participants. It is possible that the experiments achieving the best error rates could have been based on more samples than the others - in theory providing a broader range of samples on which to base a profile.

It should be noted that Table 5.1does not fully reflect the entirety of published work in this domain, as a number of papers did not provide results in an appropriate format (i.e. the outcomes of the trials did not specify overall acceptance/rejection rates).

5.7 New approaches

The previous section summarised a range of published trials considering the application of keystroke analysis using digraph samples. It has already been indicated that there are other metrics that can be considered and the following sections will discuss the use of trigraph, keyword and application profiling. While these new approaches are not used in the initial trial described in the next chapter, the larger trial discussed in chapter 7 considers these methods in a practical implementation.

5.7.1 Trigraph/keyword profiling

Previous works have concentrated on the use of digraph profiles to authenticate users' typing patterns with few suggesting the possibility of using trigraphs - combinations of three characters (Song et al, 1997, Bergadano et al, 2002) or keywords. The use of trigraph profiling presents a wider range of available profile samples (i.e. the English language provides more three character combinations than two character) and also improves the likelihood of obtaining a wider range of timings (i.e. short digraph timings would be removed using the low pass filter whereas with trigraphs it is probable that a short timing on the first and second characters could be offset by a longer timing on the second and third character timings. It would still be necessary to have a high/low pass filter to cut off the extreme outliers, but this should still leave enough samples on which to base a profile. The same process applies to the application of keyword profiling; this could either be based upon a subset of commonly occurring words or based on specific rules (e.g. username or password)

Once enough data is gathered, it is likely that trigraphs would be evaluated in the same manner as digraphs, either adopting a statistical approach or using a neural network method (with a reduced range of trigraphs).

5.7.2 Application profiling

A further variation in the data analysis can be introduced through the consideration of application specific keystroke profiles. If we accept from previous work that individual users have a distinct typing pattern, it can be hypothesised that an individual's typing pattern may also vary depending upon the application in use. For example, a user participating in a chat session may type in a fairly relaxed conversational style, while the same user may type in a significantly different way when producing a document. As such, it may be possible to base an authentication judgement based on keyword timings.

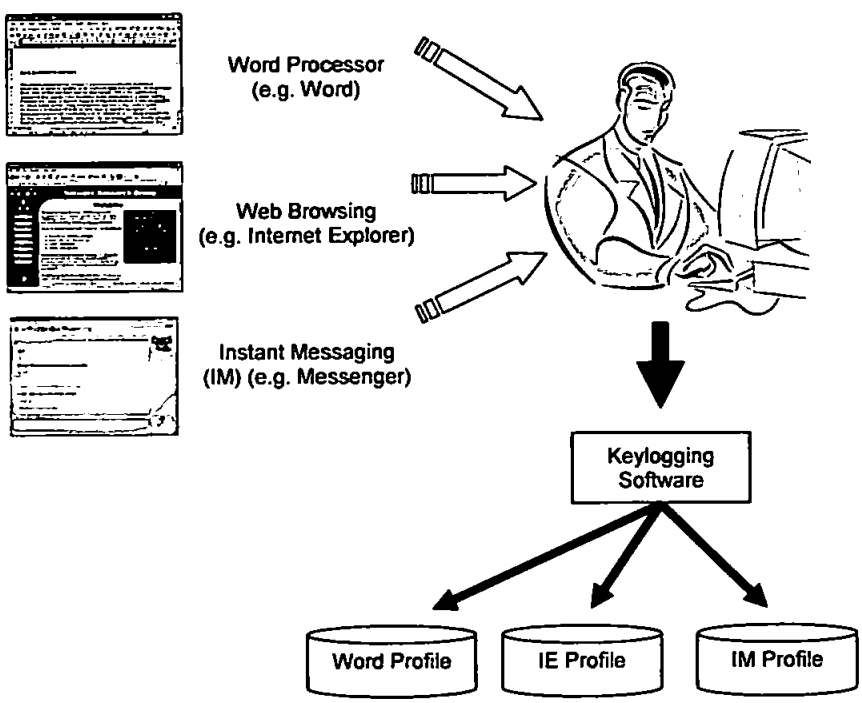


Figure 5.2 - Application profiling

5.7.3 Numeric profiling

It should also be noted that certain categories of user might use the numeric keypad for large quantities of data entry. Under these circumstances the volume and diversity of the keystroke digraphs will vary tremendously when compared to the more usual alphanumeric typing encountered with most user profiles. Previous research has been carried out in this area (Ord & Furnell, 2000), which has shown that analysis of numeric keystrokes can provide a viable authentication measure. This is an area receiving on-going attention through a separate research project in the Network Research Group (Clarke et al, 2003).

5.7.4 Composite keystroke dynamics

The results shown earlier in Table 5.1 demonstrated that the concept of keystroke analysis is feasible as an authentication mechanism. However, these trials were limited and provided variable results. While the approaches outlined in this chapter have focussed on the application of a single latency measure (i.e. using a digraph, trigraph or keyword method) it may be possible to obtain better results using a composite approach. By combining the confidence measures of multiple metrics (e.g. monitoring digraphs and trigraphs), coupled with monitoring specific keywords (e.g. the typing patterns for high-risk words – format, delete etc.), it may be possible to provide a higher level of confidence in the authentication of the user. The potential for this method will be considered further in a practical implementation in chapter 7.

5.8 Conclusions

This chapter has considered the range of metrics available for monitoring keystroke analysis and the methods for processing such data. A number of possible additional measures have been identified that could be employed to improve the performance of digraph-only keystroke analysis.

The following chapter describes an initial trial investigating the application for keystroke analysis (digraph based) within a limited number of trial participants over a one-month period. While this trial only used digraph profiling, a later trial (discussed in chapter 7) profiled users based on digraph, trigraph and word specific samples.

Chapter 6

System-Wide Keystroke Analysis

6.1 Introduction

In the previous chapter, the concept of keystroke analysis was introduced with a description of the various keystroke metrics that can be evaluated, together with an overview of the ways in which keystroke samples can be obtained. This chapter presents the results of a small-scale trial conducted to evaluate the viability of unconstrained, non-intrusive keystroke analysis (i.e. transparently monitoring a users' normal session). Before looking at the results, it is first necessary to describe the technical implementation of the software used to monitor the trial users, and to then consider the processing performed upon the acquired sample data.

This initial trial used an analytical approach for detecting deviation from a users' historical keystroke profile captured under a multi-tasking windowed environment. An alternative technique, a Data Mining (DM) approach, was also considered in order to determine the potential for improving user classification. These trials aimed to determine which approach provides the best basis for further research, and were not intended to produce a statistically valid conclusion (rather its aim was to provide a "proof of concept" that could then be used to demonstrate the viability of this approach).

6.2 Experiment Overview

While keystroke analysis has been investigated (and hence implemented) in previous studies, a Graphical User Interface (GUI) environment (e.g. Microsoft Windows) introduces new challenges. In previous published studies, the user has been required to

type and interact with a specific application (typing either pre-defined or free-form text). While this approach makes the development of the keystroke monitoring software simple, and maintains the consistency of the test environment, it is not representative of normal typing behaviour as the user becomes focussed upon the *task* of typing, rather than focussed upon a task that *involves* typing. If the aim is to produce static keystroke analysis for occasional authentication judgements (e.g. supplementing login authentication) then this approach will work well. However, to implement continuous supervision it is necessary to monitor the users' normal behaviour when interacting with their normal applications and operating system environment. Even providing a simulation of these environments may not be sufficient to obtain valid sample data on which to base a profile.

In order to address this problem, software was developed that would transparently monitor and log all typing activity. The system was designed to allow keystroke data to be collected under the Microsoft Windows NT environment (although the technique is equally applicable in all Microsoft Windows operating systems). In order to collect the required data, it was necessary to implement a mechanism for acquiring user typing patterns across all applications running within a users' active session. This is important as the experiment was designed to create a profile for each user based upon their typical typing patterns when using their computer (not constrained to a specific application or task). The implementation of the keylogger utilised several key features of the Windows operating system and the underlying chains of messages on which the operating system is built. These are briefly discussed in the following section.

6.2.1 Windows messages

The Windows operating system works on an event driven principle – when a key is pressed on the keyboard, or the mouse is moved, or any other *monitored* event occurs, the operating system generates a message that is sent to any application (or service) in the event chain.

If we imagine a hierarchy, representing the operating system and the applications running under it, messages are normally passed from the operating system, to each application registered to receive the appropriate messages. For example, almost all Windows applications process mousemove events – these are used to detect when the mouse has entered an application window. When an application is loaded, it notifies the operating system that it wishes to receive specific events when they occur, and from that point onwards the operating system provides notification of events to each application listed for each event. Windows applies a level of intelligence to message handling, such that when a notifiable event occurs (e.g. a keypress on the keyboard), the foreground application with the focus (i.e. the current application) will receive the message. This ensures that when typing into one application, others running in the background do not inadvertently receive the messages. However, if there are a series of applications that request the event notification, the events are passed through the event chain starting with the foreground application and working down to the most recent addition (Figure 6.1 shows this message handling in a simplified form).

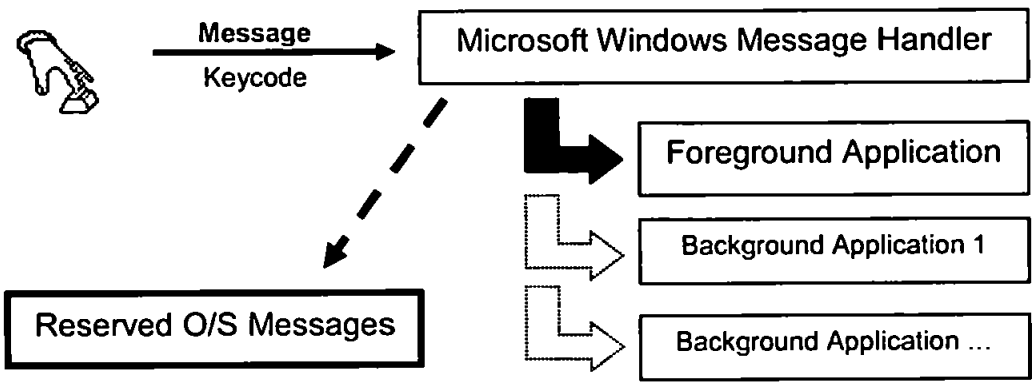


Figure 6.1 - Normal Windows messaging

When a key is pressed an event is raised in the operating system that is used to create a Windows message. Depending on the content of this message (the keycode in the event of a keypress), the message is routed to one or more applications or services. Generally (for keypresses) the message is directed at the current foreground application. However, if the message relates to a reserved operating system event (e.g. the CTRL-ALT-DEL sequence), this message is redirected to other system processes (e.g. the GINA DLL mentioned in the previous chapter). This is an important feature as, without it, an application would be able to obtain notification of the CTRL-ALT-DEL sequence and potentially override the underlying operating system's authentication routines. This could then be used (theoretically) to present a bogus login prompt with the aim of surreptitiously obtaining user login details.

In order to obtain keystroke messages under Windows (so that logging of keystrokes can occur across all applications) it is necessary to insert an application into the highest position of the event chain. This is important as once an application has received and processed a message (e.g. following a keypress event) it is removed from the hook chain and lower applications will not receive the message. To achieve this, a special form of

application must be written which implement a *system-wide* hook function. System-wide hooks allow a specified code block (the hook-function) to receive the required Windows messages irrespective of the target application (i.e. it is possible for a hook function to receive keystroke notifications for all currently running applications). This effectively allows application keystroke data to be *sniffed* and directed towards the data logger on the client workstation (Figure 6.2).

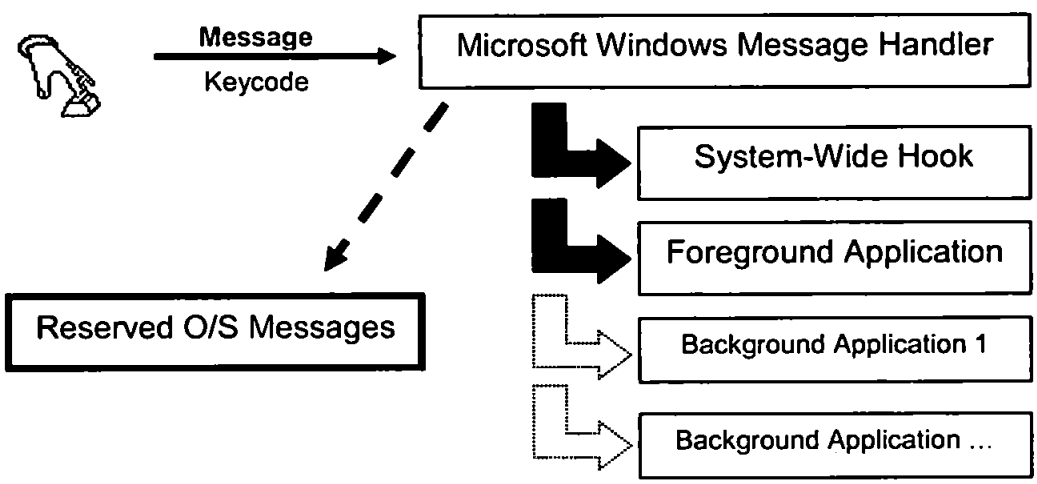


Figure 6.2 – Insertion of system-wide hook function

Once the hook function receives the message from the operating system it is important to ensure that the message is processed quickly and then sent on (dispatched) to the next application in the chain. If this is not done, the hook function (the keylogger in our case) will effectively absorb all keypresses and will not allow lower level applications to see any user activity – which is clearly undesirable.

6.2.2 System-wide hook implementation

The implementation of the keylogger required a system-wide hook to intercept all keystroke messages, and an application to filter and log keystrokes. The first attempt at implementing this utilised Visual Basic v6 Enterprise Edition to develop a simple application to log all keystrokes entered within the application window. Once the user had completed typing, the application would create a typing profile based upon typical typing patterns (illustrated in Figure 6.3). Figure 6.4 shows typing profiles from three different users with the lines indicating average typing speed per digraph. It should be noted that this application was a simple proof-of-concept program to determine the abilities of Windows to capture keystroke information and produce accurate digraph latencies transparently – no actual profiling was carried out based upon the results of this program and it is presented here simply to demonstrate the principle of capturing keystroke data.

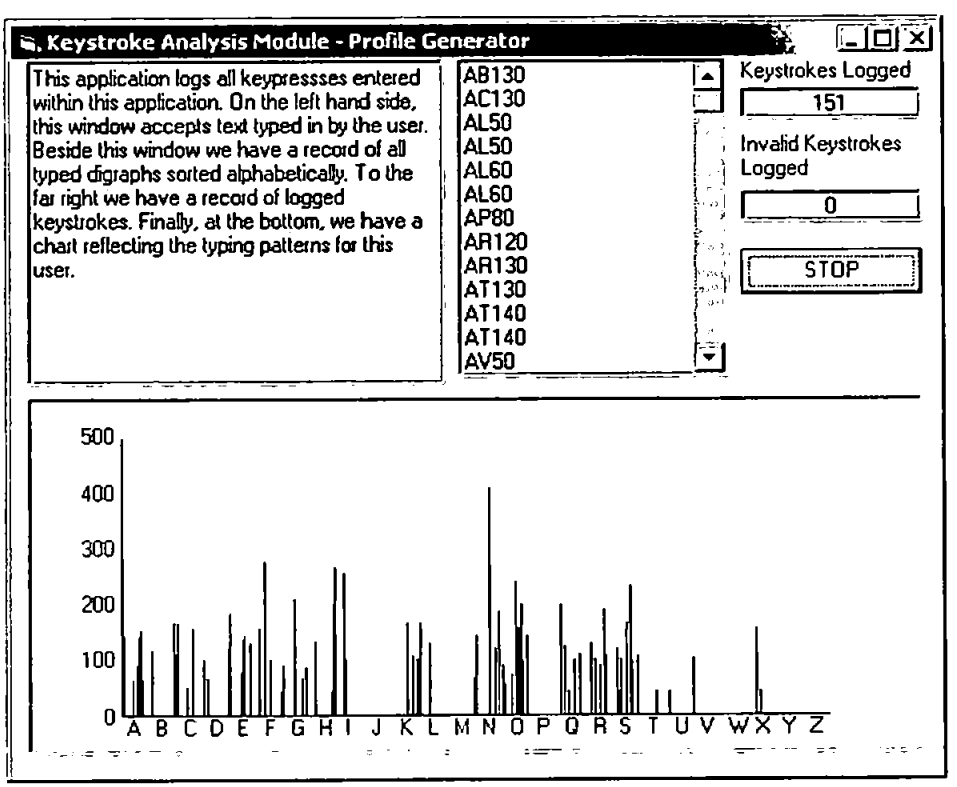


Figure 6.3 - Simple application for keystroke logging (vertical axis - time in ms)

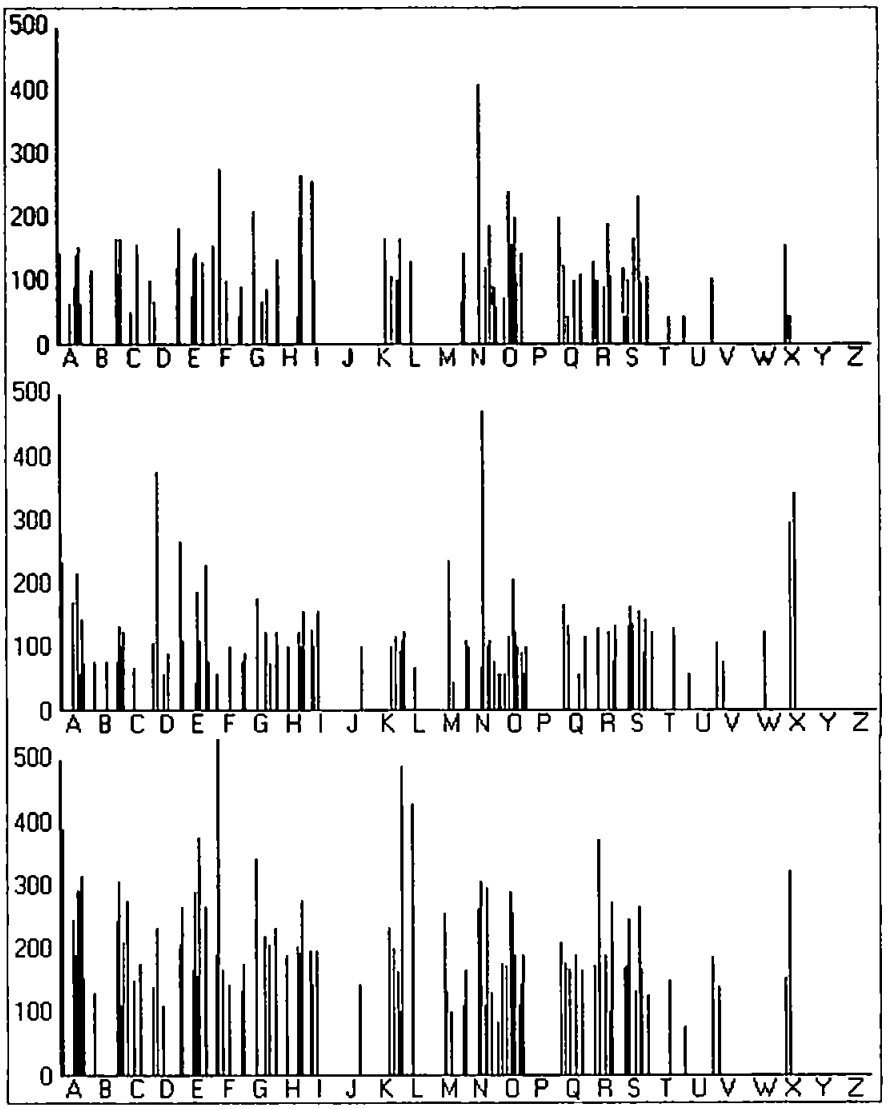


Figure 6.4 - Comparative profiles from three users (same typed text)

The digraph plots shown in Figure 6.4 show significant differences in typing patterns (particularly apparent in the third chart) – it should be noted that the actual digraph distribution varies slightly among the users as typing errors were also logged. While this application was able to successfully log all keystrokes typed within its text box, there were several limitations. The first (and obvious) limitation is that the keystrokes were only logged for the test application (i.e. all other applications were ignored). Secondly, and of more concern at this stage, was that the timing accuracy for digraph latencies were somewhat unreliable. A series of tests demonstrated that the timing resolution under

Visual Basic was in the order of 10 milli-seconds (the timings shown in Figure 6.3 show a resolution of 10ms) – far too large for keystroke latencies which are expected to be in the range 40-750ms (i.e. a 10ms resolution may not provide sufficient distinction between user samples). In order to address this, further investigations were conducted to locate a more accurate timer. The obvious alternative was to use the Windows API GetTickCount() function – while this provides a timer with a resolution of 1ms, it is somewhat unreliable as the counter used by this function uses a fixed range data type and can overflow (wraps around back to zero) there are also problems due to locking of threads that prevents the GetTickCount() function registering CPU cycles. Fortunately the Windows Application Programming Interface (API) provides another alternative, a simple solution in the form of the QueryPerformanceCounter (located in the kernel DLL).

The QueryPerformanceCounter functions provide access to a "high-resolution performance counter" (Microsoft, 2004). Using the two counter functions it is possible to obtain a timer with a resolution of approximately 1μs (Table 6.1).

Function	**Purpose**
QueryPerformanceCounter	Returns a large integer equivalent to the tick-count (i.e. number of clock ticks since boot time).
QueryPerformanceFrequency	Returns a large integer containing the performance-counter frequency, in counts per second. This can be used to determine the number of clock ticks occurring per second and hence convert the counter value into a time in seconds.

Table 6.1 – Timer functions under Windows API

The functions listed in Table 6.1 are available in all versions of Windows from Windows 95 onwards – however, not all computer systems provide access to this counter (the high-

performance counter is a feature of the system's CPU and almost all processors since the 386 support this, however, it is still hardware dependent). It is therefore important to check the return values of these functions to ensure that the counters are actually available on the system under test.

To use these timer functions for digraph latency timings, the counter needs to be stored when the first keystroke is released (time1) and at second keystroke is pressed (time2). To determine the latency in seconds it is necessary to use both functions together. Subtracting the times provides the number of counter intervals that have occurred between the keystrokes, which can then be divided by the counter frequency to give a time in seconds (the inter-keystroke latency).

```
latency = (time2 - time1) / frequency
```

Having obtained the necessary timing resolution, the problem of system-wide keystroke capture was addressed. The common approach to this problem is to implement a system-wide hook to capture all events of a specific type (keypresses in this case). This however was not possible in Visual Basic. To intercept keyboard messages it is necessary to call the SetWindowsHookEx function in the Windows API (code below). As part of this function call, a pointer is passed providing the memory address of a function to receive the intercepted messages. While Visual Basic (VB) is able to provide a function to receive the messages, it cannot be accessed by other applications (i.e. the messages must have originated within the VB application). The normal method for implementing system-wide hooks is to implement the code in a standard Windows Dynamic Link Library (DLL) file. Unfortunately VB is unable to produce standard DLL's (VB can however create OLE-DLL

files – but these are very different in structure). To resolve this problem, a DLL had to be written in Visual C++ v6 (VC) in order to create the necessary code.

```
HHOOK SetWindowsHookEx(idHook, lpfn, hMod, dwThreadId)
```

Parameters:

`idHook` – Specifies the type of hook procedure to be installed.
`lpfn` – Pointer to the hook procedure.
`hMod` – Handle to the DLL containing the hook procedure.
`dwThreadID` – identifier of the thread with which the hook procedure is to be associated

Code Sample 1 - SetWindowsHookEx() function prototype

The important parameter is the final one. The dwThreadID parameter specifies which thread (effectively which application) the hook will be installed for. While VB was able to install a hook, its use was constrained to the current thread (i.e. the application calling the function). To enable a system-wide hook across all threads (and hence all applications) it is necessary to pass the value zero – this will then ensure that the hook procedure is associated with all existing threads running in the same desktop as the calling thread.

Finally, the code to process the messages was developed. This function had to evaluate each Windows message in turn to check the nature of the message and to then pass this information back to a program that could process and log each keypress/release event. It was decided that this would be a two-stage process. Windows messages would be received by the KeyboardProc function (the keyboard system-wide hook function) and would then be filtered. Keypress events (both key-up and key-down messages) would then be sent to a VB application for further processing and logging. This sequence is represented in Figure 6.5.

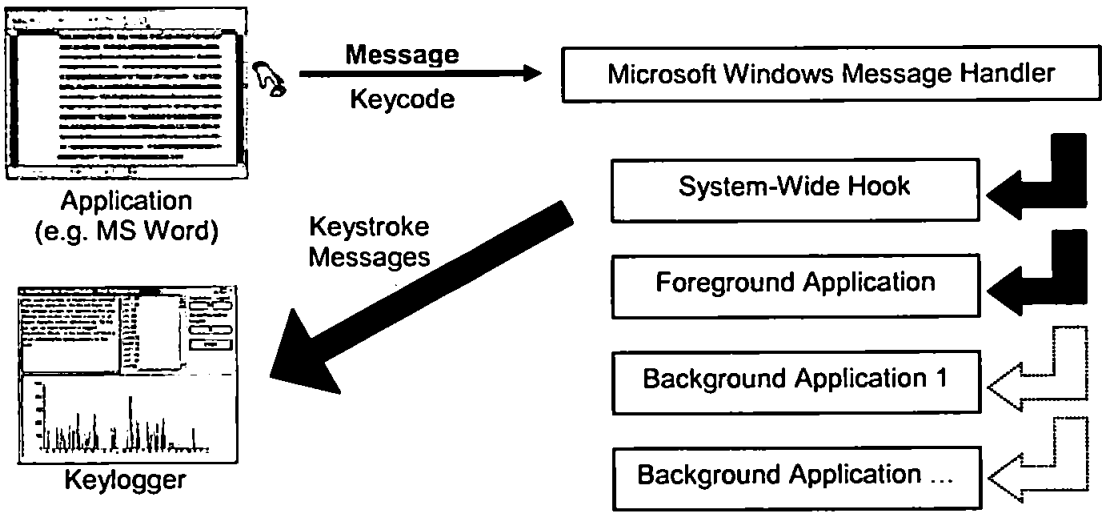


Figure 6.5 - Keylogger inserted with system-wide scope

To send the keystroke data to the keylogger, messages were directly sent to the VB application replicating the keypresses received by the hook function (and destined for the original target application). This effectively meant that the keylogger saw all system-wide keypress events as if they had occurred locally within the VB application. This was achieved by using the PostMessage API function to notify the target application (the keylogger) of a specific event. This is the same process used by Computer Based Training (CBT) packages to take remote control of an application and demonstrate normal user interaction e.g. typing and mouse movements (this is usually referred to as *injecting* messages).

The KeyboardProc function code (to receive and process keypress messages) is listed in Code Sample 2.

```
LRESULT VCKEYBOARDHOOK_API __stdcall KeyboardProc(int nCode, WPARAM wParam, LPARAM lParam)
{
    //Only examine HC_ACTION messages - other messages contain no data.
    if (nCode==HC_ACTION)
    {
       //Handle key-up action
       if (HIWORD(lParam) & KF_UP)
       {
          PostMessage(loggerWindow, WM_KEYUP, wParam, lParam);
       }

       //Handle key-down action.
       else if ((!(HIWORD(lParam) & KF_UP)) & (!(HIWORD(lParam) & KF_REPEAT)))
       {
          PostMessage(loggerWindow, WM_KEYDOWN, wParam, lParam);
       }
    }

  //
  // Once we have looked at the message and passed it to the keylogger
  //   the message must be passed to the original application
  // We will pass all messages on to CallNextHookEx.
  //
  return(CallNextHookEx(keyboardHook, nCode, wParam, lParam));
}
```

Code Sample 2 - KeyboardProc() implementation

As messages are received they are split into key-up and key-down events. Key-down events are further checked to eliminate repeat keys (i.e. when a key is held down causing multiple, repeated characters). Messages are then posted to the VB keylogger (which then sees the key events as local keypresses) and finally the original message is sent on to the original destination (e.g. to Microsoft Word). The wParam value holds the virtual key-code for the keypress event, while the lParam value holds a 32 bit value where the bit sequences indicates additional parameters, such as the nature of the keypress event (key up, down or repeat), the number of repeat occurrences of the key, scan-codes and the status of the extended keys (e.g. the ALT key). Keyboards generate scan-codes for each keypress/release – with each key generating a unique *make* and *break* scancode. Scan-codes provide low-level keyboard information – e.g. the ability to distinguish between the numerals at the top of the letter keys and those found in a dedicated numeric keypad.

An additional feature was introduced to the hook function to monitor the handle of the application in which the keypress occurred (a handle is a numeric value that uniquely identifies the application window within Windows). This would allow the keylogger to also record the title of the foreground application window (i.e. the program in which the user was actively typing). The code for this is presented below and simply monitored the foreground window handle for any change. When a change in application focus occurred, the keylogger was notified of the change of focus by a Windows message – as keypress message were already in use to notify the keylogger of keyboard activity, the keylogger was notified of a change of application focus by sending a left button mouse-click message. This would then allow the keylogger to request the title of the application window (via the API) and subsequently log this information, together with the keystroke data and the username of the currently logged in user.

```
hwndCurrent=GetForegroundWindow();
if (IsWindow(hwndCurrent))
{
   if (hwndCurrent!=hwndLast)
   {
      PostMessage(loggerWindow, WM_LBUTTONDOWN, MK_LBUTTON, 0);
   }
   hwndLast=hwndCurrent;
}
```

Code Sample 3 - Notification of application focus change

6.2.3 Keylogger implementation

The VB keylogger had a relatively simple implementation. Once started, the application loaded the hook function into memory by initiating the DLL file. The hook function was then inserted into the message event chain with system-wide scope. As part of this, the hook function was also passed a handle to a picture box located on the keylogger interface.

This control was chosen as it can receive keystroke and mouse click messages while presenting no visual interface – i.e. if a user opened the application they would not be able to interfere with the key logging process. This handle was used by the hook function to post messages directly to the VB keylogger application (Figure 6.6).

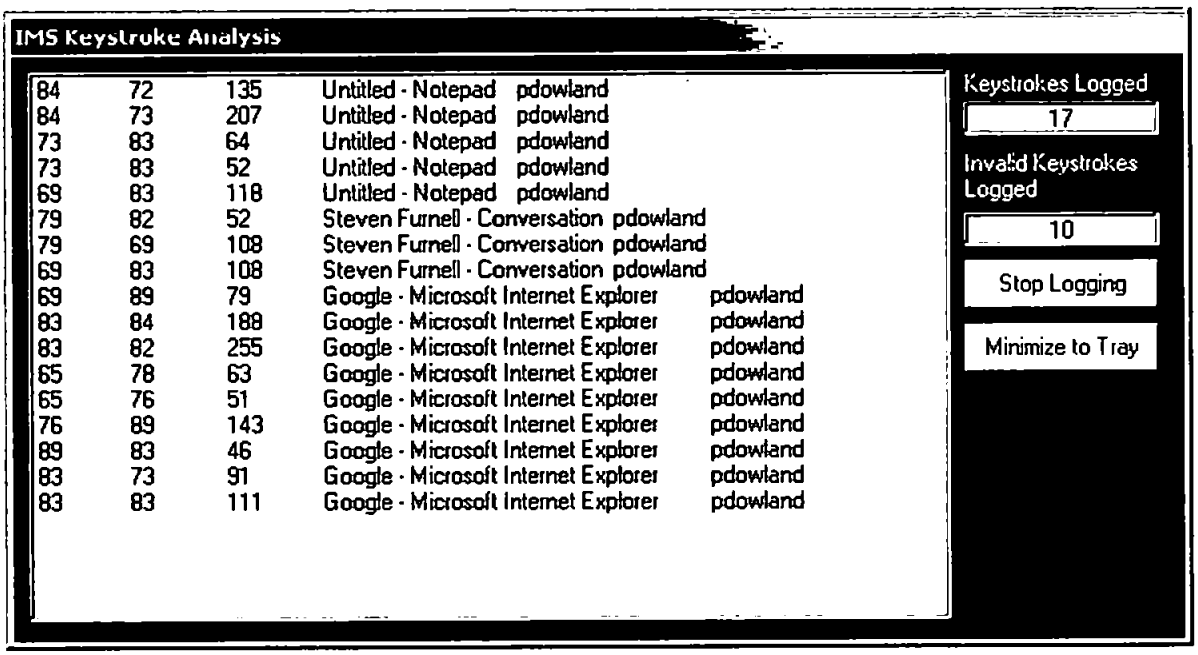


Figure 6.6 - Key logging across multiple applications

When running, the keylogger was discreetly added to the system tray to avoid any inconvenience to the user (shown as the furthest right icon in Figure 6.7).

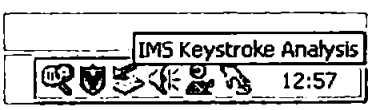


Figure 6.7 - System tray icon for keystroke analysis

For each digraph pair logged, the application stores five items of information – these being written to a text file after every 500 digraphs (Table 6.2).

Item	Data types
Left character	ASCII code representing character
Right character	ASCII code representing character
Latency	Integer representing inter-keystroke latency in milliseconds
Application	String containing the window title from the foreground application.
Username	String containing currently logged-in user

Table 6.2 - Keylogger attributes logged per digraph

6.2.4 Filtering

To eliminate extreme short/long digraph latencies that may adversely affect the distribution of digraph times, any digraph pair whose latency fell outside a nominal range was excluded from the log files. For the purpose of this experiment the range was restricted to times above 40ms and below 750ms. These thresholds were based on previous work conducted by Furnell (1995), and were designed to eliminate samples where two keys may have been accidentally struck together (thus, producing an infeasibly small latency) or, where the user may have made a pause in their typing and thus introduced an unnaturally large inter-keystroke latency.

6.3 Final implementation

The final implementation of the keylogger application is shown in Figure 6.8. This shows the messages passed from the hook (implemented in the DLL in C) and the keylogger (implemented in Visual Basic and deployed as a system tray application). The keylogger functioned completely transparently to the user, requiring no user action to start or stop the logging process. The application was automatically started when the operating system (O/S) booted (run from the Startup program group on the start menu) and shut down

automatically when the O/S closed. Gathered data was automatically saved after every 500 digraphs pairs and when the application was closed. To reassure users, an option was included to suspend logging of keystrokes. This was included due to concerns expressed by some users about monitoring of specific inputs – e.g. the typing of on-line banking login details.

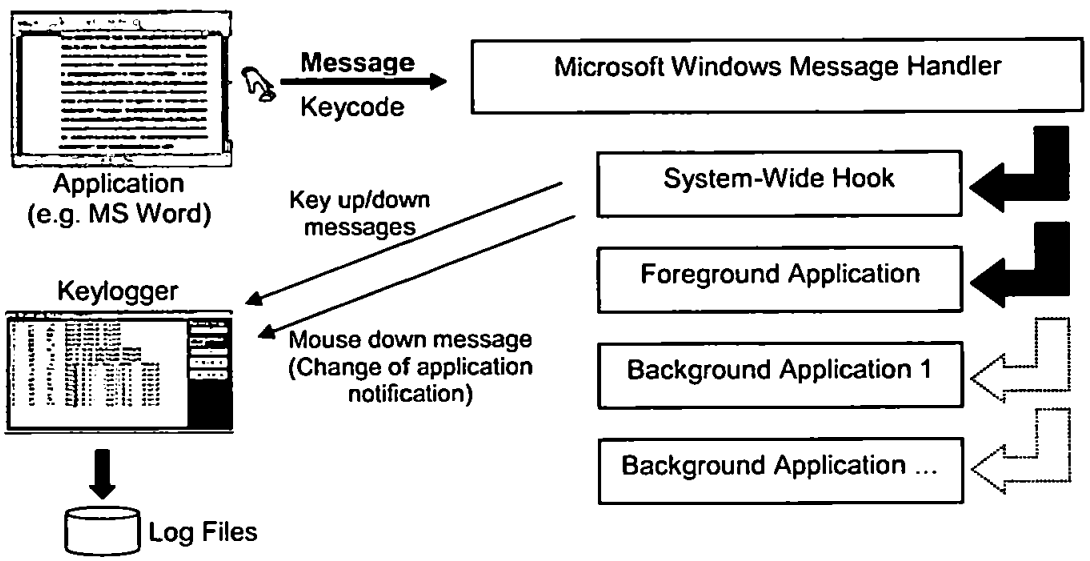


Figure 6.8 - Final implementation of keylogger

6.4 Trial participants

For this experiment a total of ten users were profiled over a period of three months. The trial participants were drawn from students and staff from the Network Research Group. As the intention was to evaluate the analysis mechanisms without implementing a large-scale trial, tests were carried out using a small set of test subjects. The main limiting factor was the need to collect data over a prolonged period (weeks rather than hours). Despite the small scale of the trial, it still proved difficult to collect sufficient data in order to provide a

valid comparison between users. Several users disabled the keylogger when entering sensitive information and consequently forgot to re-enable it. This resulted in large variations in profile size (discussed in the following section).

Due to this limited set of data, some of the discussion and analysis in the following sections focuses on the six main users who provided the largest profiled data sets in order to best illustrate the trends observed.

6.5 Analysis

Following the initial filtering described in the previous section, the experimental data for each user was processed off-line to calculate the mean and standard deviation values for each unique digraph pair. In the event that any digraph pair had a standard deviation greater than its mean value, the digraph samples were sorted and the top/bottom 10% were then removed, followed by subsequent re-calculation of the mean and standard deviation values – this was only attempted where at least ten samples were available for the digraph pair. The reason for this additional step was to remove digraph samples where the latencies would have an adverse affect on the standard deviation (i.e. the distribution of samples was tightened).

Once a profile of digraph pairs was produced (with corresponding mean/standard deviation digraph latency values), the user's profile was further constrained by filtering out digraph pairs where the sample count fell below a nominal threshold value. This initial experiment fixed this value at fifty samples; however, the software used for analysis allowed a variable threshold (Figure 6.9).

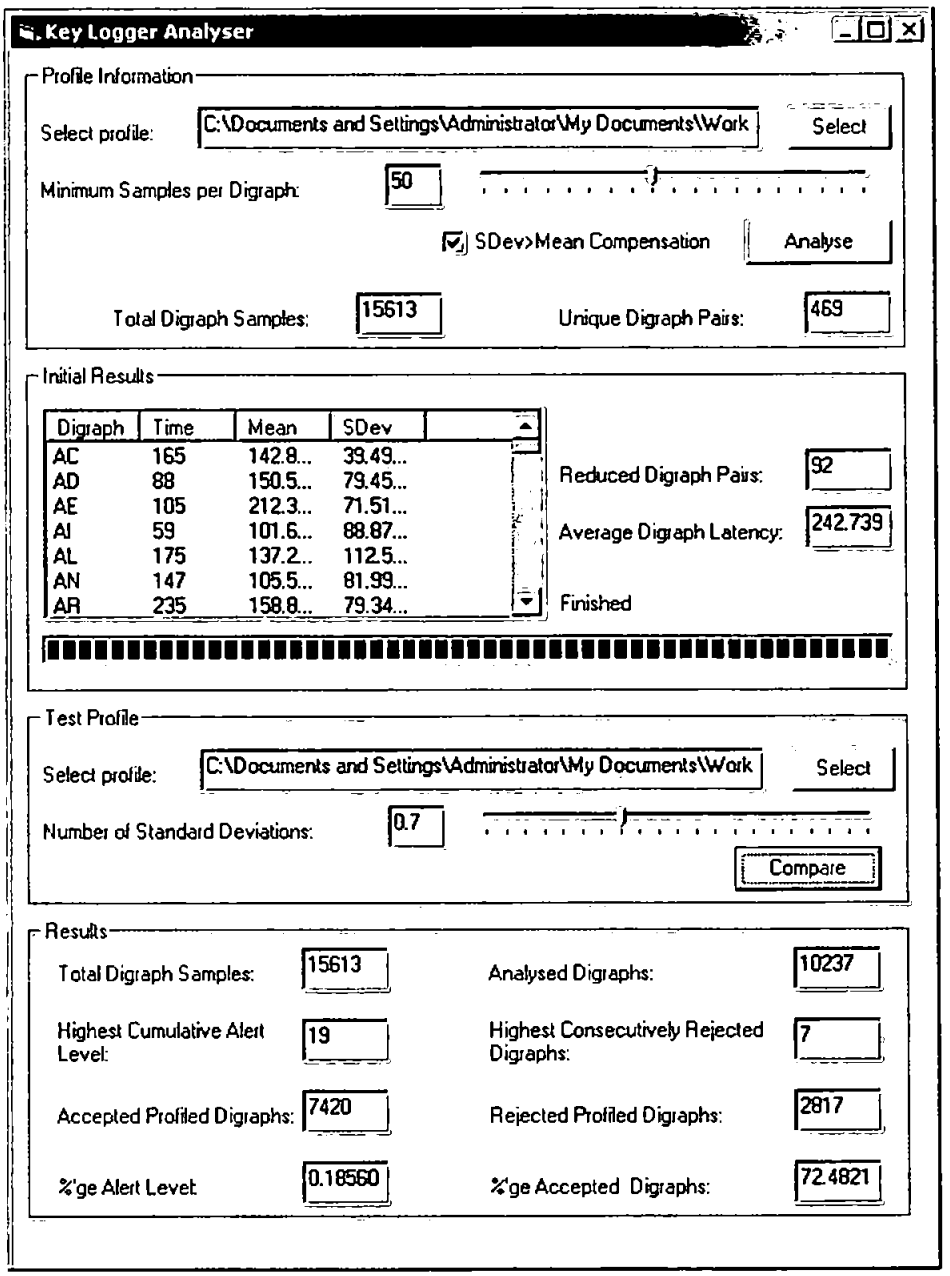


Figure 6.9 - Profile generation and testing

The profile generator and analyser is split into four separate sections with a number of options/results displayed. This is discussed in the following sections.

6.5.1 Profile settings

Figure 6.10 shows the profile selection frame within the application. This allows the selection of the original user keylogger data files (stored as a plain text file) and options to change the minimum number of samples per digraph, as well as the option to re-calculate digraph sets where the standard deviation is greater than the mean. The first setting allows a variation in the minimum number of samples required to produce a viable digraph profile. This ensures that only digraphs with a large number of samples are used in the final profile. The example in Figure 6.10 shows a profile loaded containing 15,613 individual digraph samples (a very small data set) which comprises 469 unique digraph pairs – this gives an average of only 33 samples per digraph. By applying a threshold value (set to 50 in the figure) the least frequently occurring digraphs are removed from the final profile.

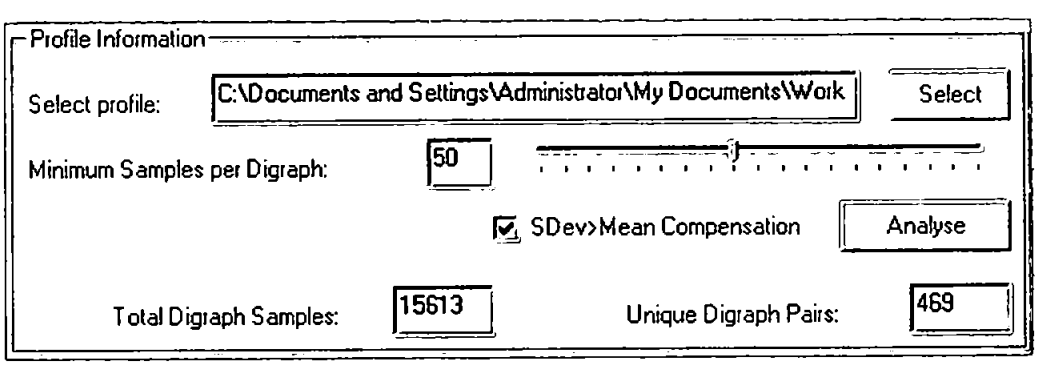


Figure 6.10 - Profile selection

The setting to select filtering of digraphs where the standard deviation is greater than the mean is used to filter digraphs with high variance. Where this condition is found true, the top/bottom 10% of samples are removed and the standard deviation and mean are recalculated with this process repeated until the variance is reduced. This will result in a digraph pair being removed from the profile altogether if the number of eligible samples is reduced below the threshold set in the first option. To give an indication of this, the test

run shown in Figure 6.9 resulted in the number of profiled digraphs being reduced from 469 to just 92.

6.5.2 Generated profile

Figure 6.11 shows the profile generated from the selected keylogger data file following the processing selected and described in the previous section. The reduced set of digraphs (92 in this example) are shown together with their associated mean and standard deviation values. Note that the digraphs are sorted alphabetically rather than by frequency of occurrence – this was by design to allow rapid searching of the profile for specific digraph pairs when comparing raw keylogger data files. The variance shown between digraphs is quite significant – the digraph pair A-C has a profile of 142.8ms ± 39.49 while the digraph pair A-L has a profile of 137.2ms ± 112.5. This variance can be adjusted in the later settings described in the next section.

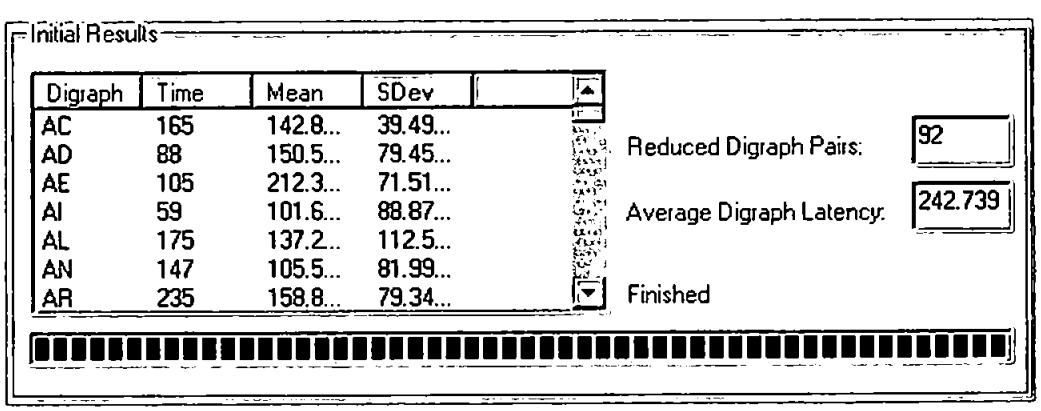


Figure 6.11 - Generated profile

6.5.3 Test profile selection and settings

Figure 6.12 shows the test profile settings frame. In this section the comparison keylogger data file is selected (the users' data that will be compared against the generated profile) and the level of deviation from the profile is determined. When the comparison is started, each digraph is compared against the reference profile and a simple analytical comparison is performed. If the test digraph is within the permitted range of the reference profile the digraph is accepted. The permitted deviation is determined by the slider control that selects the number of standard deviations from the mean.

digraph mean ± (digraph standard deviation * permitted deviation)

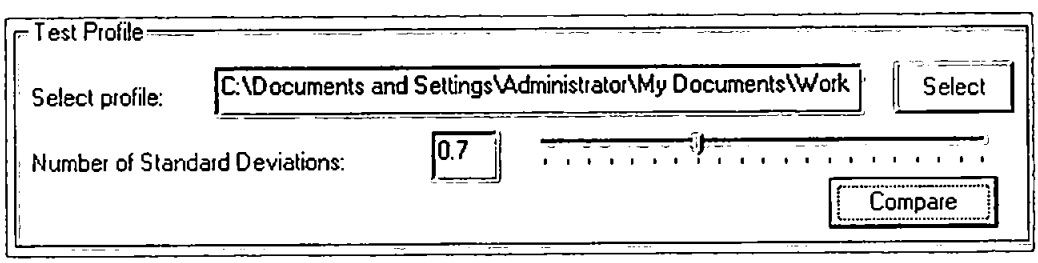


Figure 6.12 - Test profile selection and settings

The comparison described above is repeated for each captured digraph in the comparison keylogger data file and a number of statistics calculated (described in the next section).

6.5.4 Test profile results

Figure 6.13 shows the results frame. This presents the results from the analysis conducted on the keylogger data file when compared against the reference profile. In total eight values are presented, as described in Table 6.3.

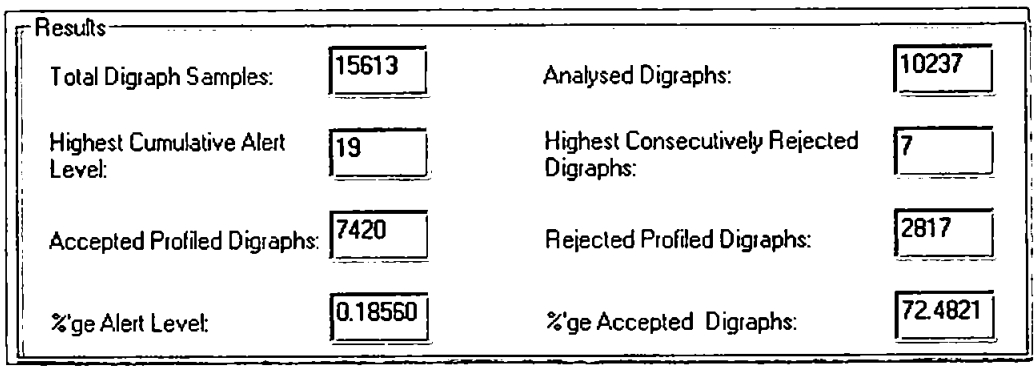


Figure 6.13 - Test profile results

Item	Description
Total Digraph Samples	Total number of digraphs in the keylogger data file.
Analysed Digraphs	Total number of digraphs that match digraphs in the comparison profile.
Highest Cumulative Alert Level	A running alert level is maintained which is incremented/decremented with each digraph decision. A live system would need a threshold value (probably individually set for each user) beyond which the user would be more explicitly challenged for authentication.
Highest Consecutively Rejected Digraphs	A count of the highest number of consecutively rejected digraphs. This may provide evidence of a poor match between reference profile and comparison keylogger file.
Accepted Profiled Digraphs	Total number of accepted digraphs (i.e. those that match the reference profile within the permitted deviation).
Rejected Profiled Digraphs	Total number of rejected digraphs (i.e. those that do not match the reference profile ± permitted deviation).
%'ge Alert Level	Alert level represented as the highest alert level divided by the number of analysed digraphs.
%'ge Accepted Digraphs	Percentage of digraphs accepted – provides an indication of the goodness of the match between reference profile and comparison keylogger data file.

Table 6.3 - Description of profile comparison results

6.6 Results

Although ten users participated in the trial run, only eight produced enough data to warrant further investigation. A summary of the user data generated in this trial is shown in Table

6.4. This shows considerable variation in the size of keylogger data files across the eight users with sample sizes in the range 7,000 to 350,000 digaphs. As such, users f and g were removed from most of the data processing as the quantity of the logged keystrokes was too small to provide a reliable profile (i.e. there were too few digraph pairs in the profile, with too few samples per profiled digraph).

User	Total Digraph Samples	Unique Digraph Pairs	Filtered Digraph Pairs	Average Typing Speed
User A	178,710	466	317	151ms
User B	59,787	405	232	145ms
User C	80,167	412	257	206ms
User D	58,987	461	224	162ms
User E	15,613	469	92	243ms
User F	8,696	391	55	285ms
User G	7,435	405	42	272ms
User H	350,567	610	369	297ms

Table 6.4 - Summary of user profile statistics

Once each user profile was generated, the reference profile was evaluated by comparison against the users' raw keylogger data. This allowed the test profile to be evaluated using the users' own data (to test the False Rejection Rate – FRR) and against other users' keystroke data (to test the False Acceptance Rate – FAR).

As there is likely to be significant variation in a users' own session data, a compensatory factor was applied to the standard deviation that could be varied in a "live" environment according to the security needs of the organisation. This factor allowed the number of standard deviations from the mean to be adjusted. For the purposes of this experiment, four weightings were considered, namely 0.5, 1, 1.5 and 2. These weightings were selected based on experimental work by Mahar *et al.* (1995) This produced an acceptable digraph range:

$$\text{digraph range} = \text{mean} \pm (\text{standard deviation} * \text{weighting factor})$$

When viewing the preliminary results (Figure 6.14), if we consider the six users A, B, C, D, E and H and follow the vertical columns of data, we can see a peak for each user's data when compared with their own profile. This is most noticeable for user D where a peak is observed (nearly 50% of all digraphs accepted) compared with 32% when user B's raw keylogger data was tested against the same profile.

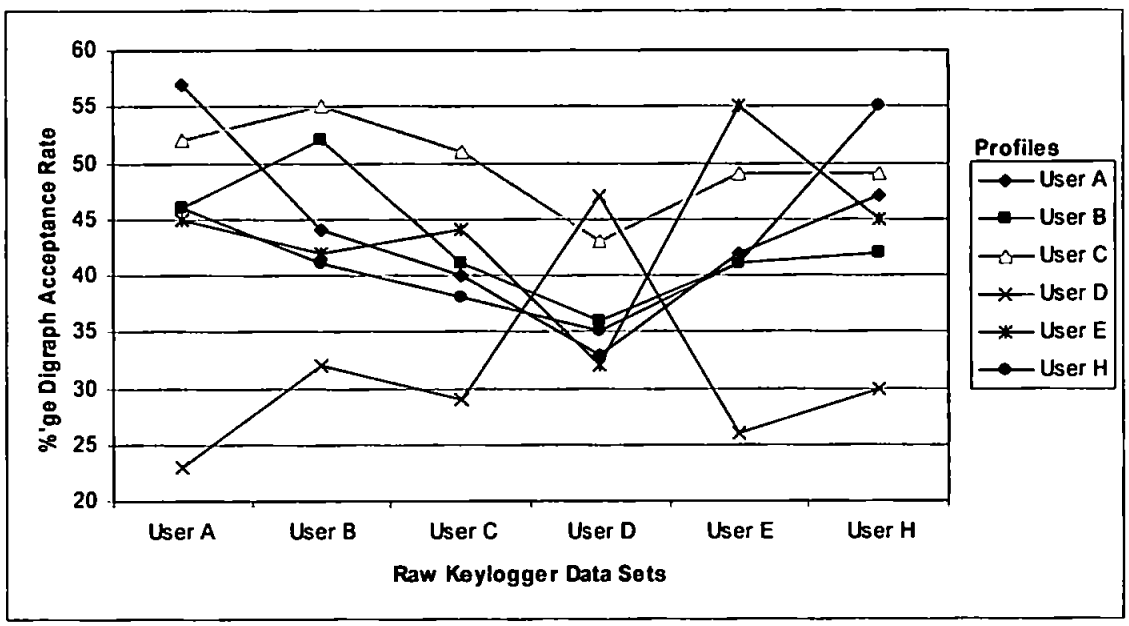


Figure 6.14 - User profile comparisons

Although there was a correlation between user D's profile and data, if we consider user A, there was a high FAR for data from users B and H (impostors) when compared with user A's profile. We can also see that in user C's profile the impostors A and B actually achieved higher acceptance rates (52% and 55% respectively) against the valid user (C) with only 51%. It is clear from these results that an additional measure of acceptance/rejection is required. To further test the FAR/FRR of the test system, the analysis software monitored the number of consecutively rejected digraph pairs – representing the highest alert level of the system (Figure 6.15 & Figure 6.16). Further

charts, presenting the comparison of all user data against all profiles can be found in Appendix B.

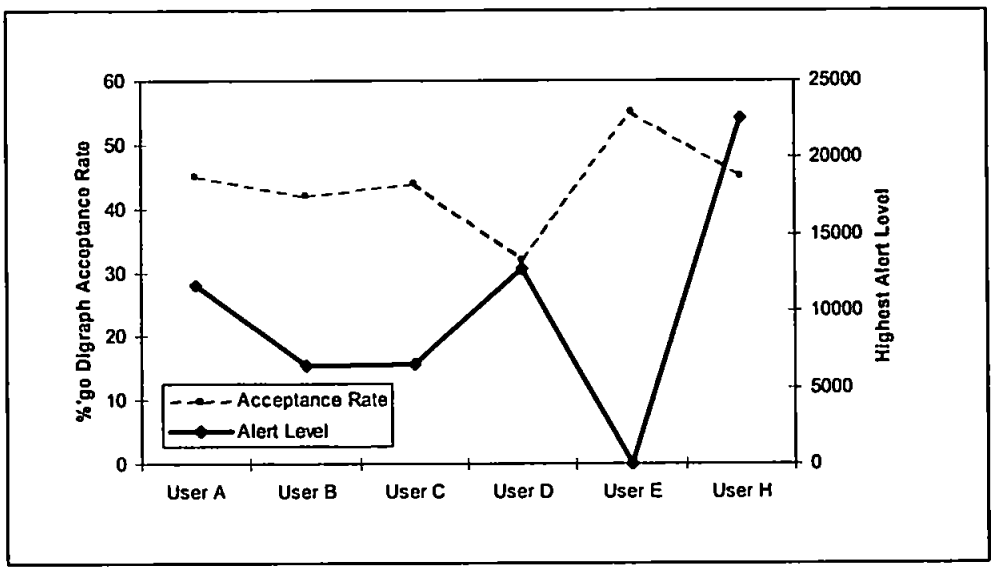


Figure 6.15 – User E profile comparison

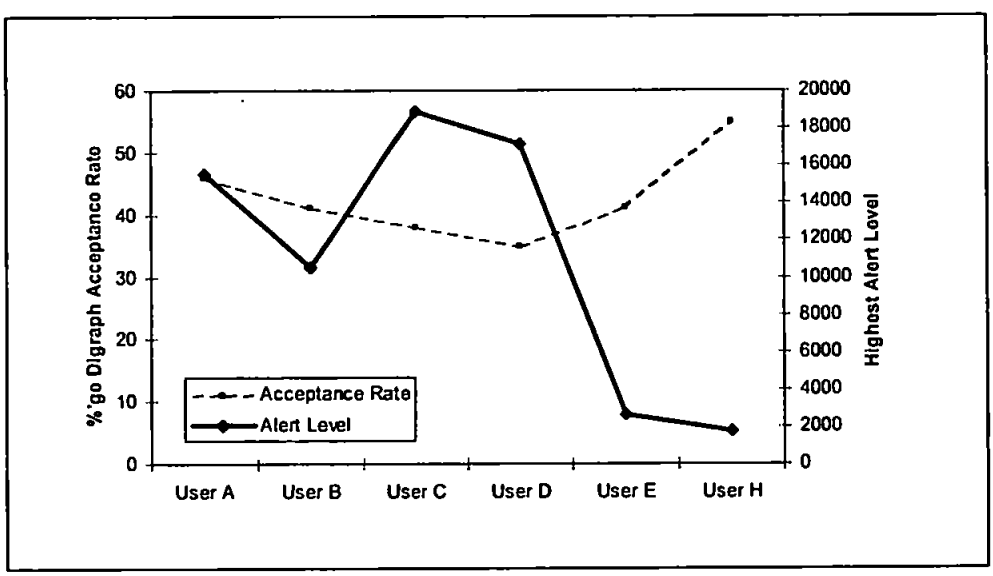


Figure 6.16 - User H profile comparison

When considering Figure 6.15 we can identify two distinct trends. Firstly, the top line plots the digraph acceptance rate for all user data sets against user E's profile. Here we can see a peak correlating to user E's own data and corresponding reductions in the acceptance

rates for the other users' data. Secondly, the lower line indicates the highest alert level detected by the analysis software. This is simply a record of the highest count of consecutively rejected digraph times (excluding non-profiled digraph pairs). Again, we can see a correlation between user E's own data when compared with their profile, and corresponding increases in the alert level as impostor data sets are compared with the target profile. Figure 6.16 further demonstrates this using user H's profile. User H also shows a significant correlation between their own profile and raw keylogger data. It is important to note that while the charts shown here (and in Appendix B) show simple comparisons between profiles, a live system would utilise thresholds for each user that would be used to determine the acceptance/rejection level for each user. For example, in Figure 6.15, user E's rejection rate is significantly smaller than even the closest impostor. It would be possible (for this user) to assign a relatively low threshold for consecutively rejected digraphs as user E had a highest count of 42 consecutively rejected digraphs whilst users B and C had counts in excess of 6000 when compared to user E's profile (i.e. as impostors).

6.7 Data mining analysis

The previous sections have considered a simple analytical approach to the problem of keystroke analysis. While this approach has shown some success both in this trial and in previous work, there are other alternatives that can be considered. One technique that was evaluated was the use of Data Mining (DM) algorithms, a previously untried approach in this field. This work was conducted in association with Harjit Singh and is covered in detail in Singh et al (2001). This part of the study will not be covered in detail here, except

to compare the FAR/FRR percentage accuracy with the approach used in the previous sections.

6.7.1 Methodology

For the DM analysis, the data sets were split into a ratio of 9:1 creating two parts; a training set and a testing set. The Intelligent Data Analysis (IDA) Data Mining Tool (Singh et al, 1999) was used to analyse the sample data sets. The IDA tool incorporates algorithms from the fields of Statistics, Machine Learning and Neural Networks, with six algorithms being selected (k-NN, COG, C4.5, CN2, OC1 and RBF). The algorithm or classifier was subjected initially with the training set, and then the classification accuracy was tested using the unseen data set or testing set. The results give an indication of the error rate (or FAR) and the overall classification accuracy of the trained algorithms.

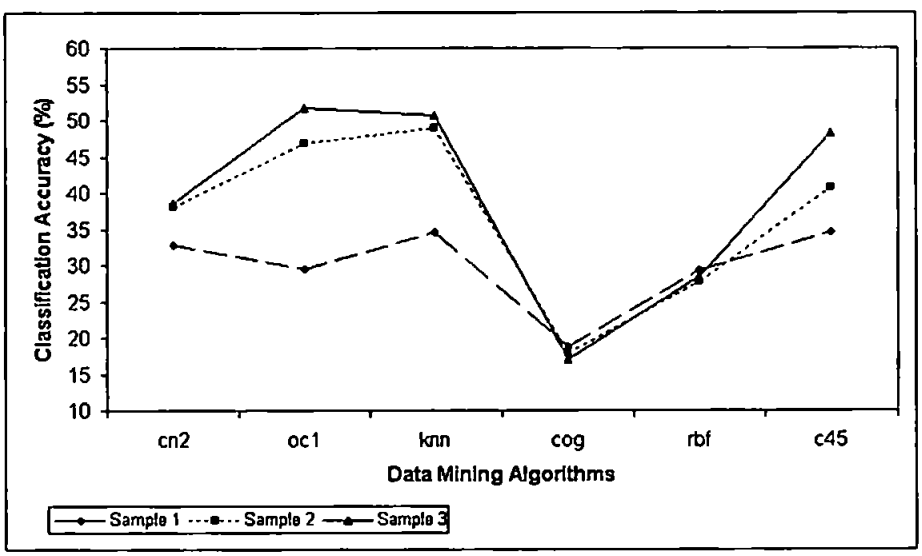


Figure 6.17 - Varying sample sizes with fixed number of classes and attributes

The percentage classification accuracy obtained is encouraging as depicted in (Figure 6.17), which shows that when the sample size is increased, the classification accuracy obtained

increases proportionally, except for COG (a statistically based algorithm) and RBF (a Neural Network based algorithm). This is important when considering the size of data being analysed.

The initial results suggest that Machine Learning (OC1 and C4.5) and Statistical (k-NN) based algorithms are suitable for these types of data sets. While these results show a classification accuracy approaching 50%, this is still far short of an acceptable level of false rejection. It is likely that with more data, and a reduced set of core (commonly occurring) digraphs, higher classification accuracies would be attainable. However, these techniques are still of marginal benefit due to the time taken to process the data sets. As such, this study proceeded with the analytical approach and the application of data mining to keystroke analysis will be revisited in chapter 9 as part of the future work proposals.

6.8 Application-specific keystroke analysis

Following the digraph-based keystroke analysis discussed earlier in this chapter, the final investigation considered the use of application-specific keystroke analysis. In this case, the analysis was conducted with a view to determining the viability of application-specific keystroke profiling (referring back to Table 6.2, the application from which the keystrokes were typed was logged in addition to other characteristics). To this end, it was necessary to identify a series of applications for profiling, with the selection criteria being those for which sufficient keystroke data had been logged during the sampling period. A review of the keystroke data revealed that the applications satisfying this requirement were Microsoft MSN Messenger, Internet Explorer, Word, Outlook and PowerPoint. While it was

considered that a numerically intensive application such as Excel would have provided an interesting candidate, insufficient keystrokes were captured to enable the creation of a profile. Additionally, of the eight users sampled during the trial (who produced sufficient data for digraph analysis), only five produced sufficient samples to analyse from all of the aforementioned applications. Although the resulting sample group was very small, it was sufficient to yield interesting results in relation to an initial assessment of application-specific profiling.

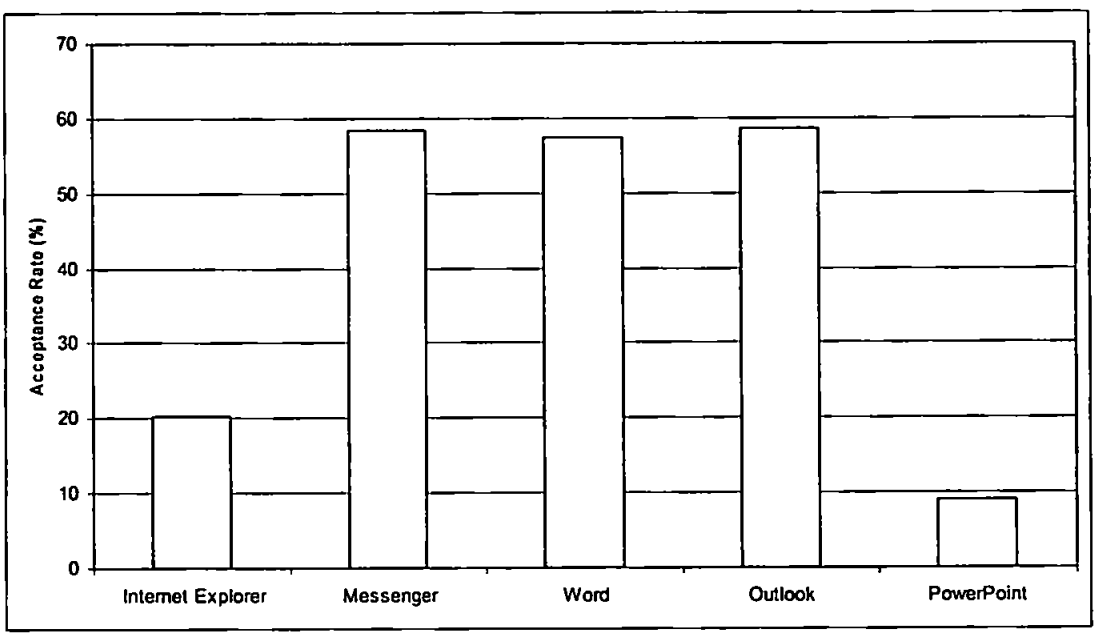


Figure 6.18 - Acceptance rate for application specific keystroke data compared against a system-wide context user profile

In Figure 6.18, a single user's application-specific keystroke data is compared against the reference profile from the same user. The reference profile was based on all keystroke data acquired from all applications. Although the figure does not show distinct differences in all cases, there is a clear distinction between PowerPoint & Internet Explorer and Messenger, Word & Outlook. This can be explained when the nature of these applications is considered. Messenger, Word and Outlook are all significantly textual in their usage, and users will typically type within them for considerable periods of time. In contrast, while

Internet Explorer and PowerPoint sessions may both involve significant elements of keyboard activity, the typing is more likely to occur in sporadic bursts. As such, any dynamic that emerges is likely to be markedly different to that which would emerge in applications where more sustained typing is the norm. Considering the information portrayed above, the creation of application specific profiles would be likely to increase the acceptance rates observed. This could be significantly more effective given more keystroke data - it may be possible (with sufficient data) to distinguish between typing-intensive applications like Word, Outlook and Messenger. For example, it may be possible to monitor the frequency of specific keys (e.g. the return/enter key) or combinations of keys (e.g. : +) to create ☺) within Messenger.

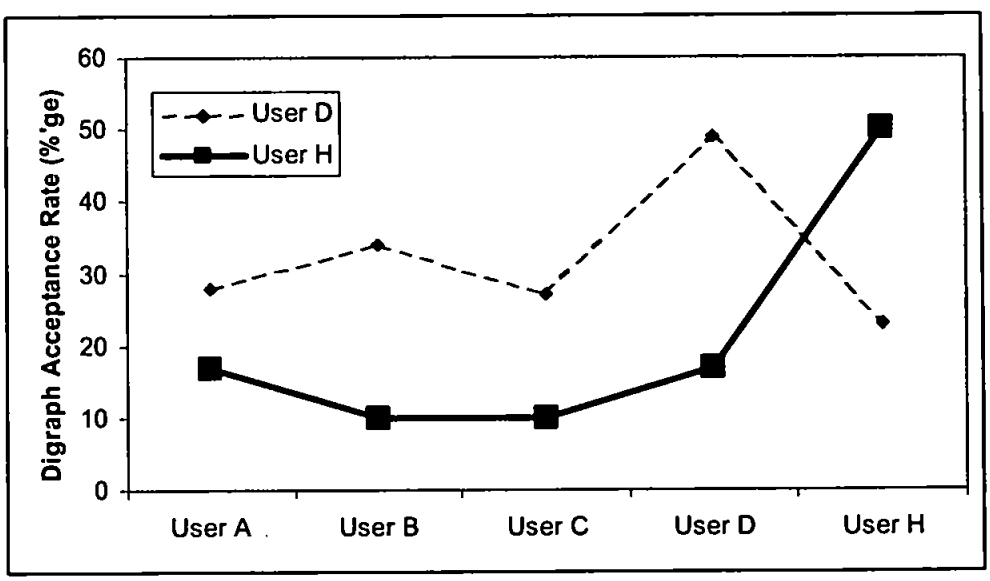


Figure 6.19 - Acceptance rate for two user profiles using Internet Explorer

In Figure 6.19, two users' profiles (users D and H when using Internet Explorer) are examined, showing there is a clear difference between other users' keystroke data (impostors) with appropriate peaks in acceptance rate for the valid users.

While the results shown do not indicate a suitably discriminative metric upon which to base a satisfactory authentication judgement, they do show a level of correlation between a user's typing patterns in an application-specific context. These preliminary results therefore suggest that further work is needed to investigate the use of application-specific keystroke analysis.

6.9 Conclusions

It is clear from the results presented in this chapter that there is potential for continuous user authentication based upon keystroke analysis. However, it is also clear that while the analytical approach provides a level of correlation between reference profile and raw data, the quantity (and range) of raw keylogger data is insufficient to draw any positive findings. The DM approach was limited due to the nature and volume of the data gathered, and is worthy of further investigation (discussed in chapter 9).

Following the findings of this trial, a more comprehensive experiment commenced in November 2003 with more users being profiled over a similar period of time. The next chapter describes the nature of this further trial, and investigates the usefulness of monitoring both trigraph keystroke combinations (timings for three consecutive keystrokes) and word-graph timings (timings for frequently occurring words).

Chapter 7

A Long-Term Trial of Keystroke Analysis

7.1 Introduction

The previous chapter described an experiment evaluating keystroke analysis based upon inter-keystroke digraph latencies under Windows. Although the trial results demonstrated the viability of this method, the results showed that reliable authentication would need user profiles to be based upon much larger sample sizes. The previous trial was also based on a limited number of users in order to quickly evaluate the viability of the technique.

This chapter presents the results of a large-scale trial that was aimed at evaluating a range of techniques using a larger number of participants.

7.2 Experiment Overview

The first trial concentrated upon the capture and subsequent analysis of digraph latencies and focussed upon inter-keystroke timings. Additionally, contextual information was also stored to allow a preliminary analysis of application-specific keystroke analysis to be conducted. This trial captured and evaluated trigraph and keyword latencies in addition to digraph timings.

7.2.1 Keylogger implementation

The method of keylogging was basically the same as used in the earlier trial with appropriate modification to allow for the logging of trigraphs and keyword latencies (illustrated in Figure 7.1).

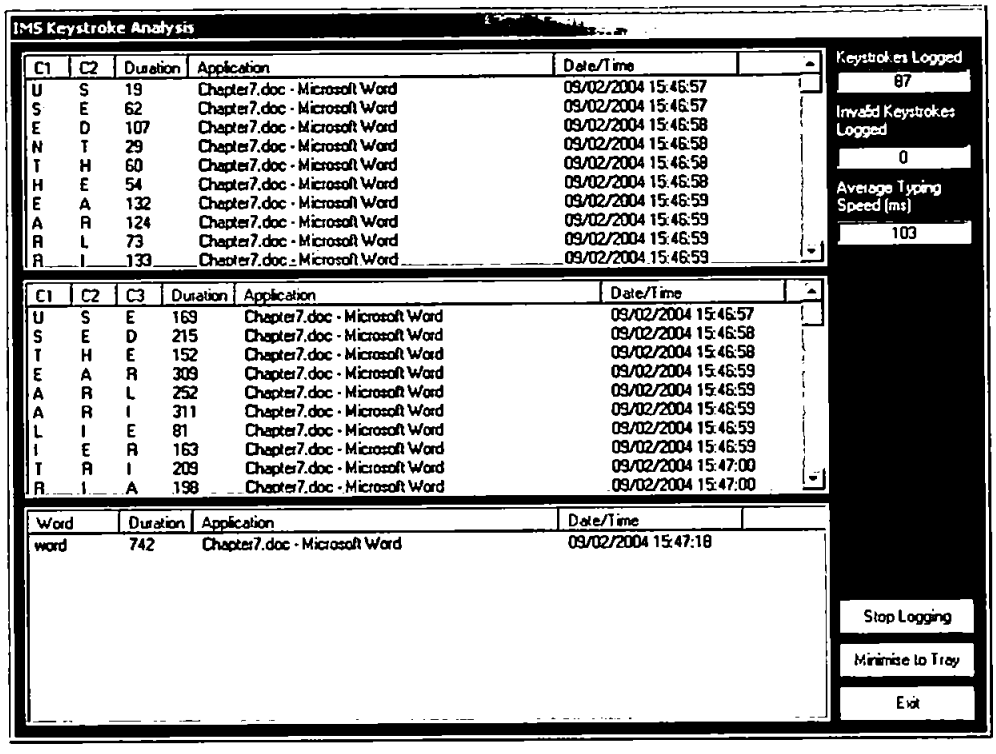


Figure 7.1 - Advanced keylogger

Due to the increased volume of information expected during this trial, the data was logged to an Access database installed as part of the key-logging software (this allowed faster, more flexible extraction of data for the analysis stage). For each digraph pair logged, the application stored six items of information – these being written to the Access database after every 500 digraphs (Table 7.1). This process was also repeated for each trigraph and keyword latency (i.e. trigraphs were stored as three consecutive characters and keywords

as a string as shown in Table 7.2). As can be seen from these examples, significant quantities of data were stored for later analysis.

Item	Data types
AutoID	Auto-incrementing record number. This is used to maintain the order of the keystrokes typed as the timestamp is only accurate to 1 second.
Left character (C1)	ASCII code representing character
Right character (C2)	ASCII code representing character
Latency	Integer representing inter-keystroke latency in milliseconds
Application	String containing the window title from the foreground application.
Timestamp	A timestamp is added to every keystroke logged for later use.

Table 7.1 - Keylogger attributes logged per digraph

Metric: Digraph

AutoID	C1	C2	Duration	Application	TimeStamp
1796	C	T	136	Google - Microsoft Internet Explorer	11/06/2003 16:12:04
1797	T	I	224	Google - Microsoft Internet Explorer	11/06/2003 16:12:05
1798	I	V	179	Google - Microsoft Internet Explorer	11/06/2003 16:12:05
1799	V	E	50	Google - Microsoft Internet Explorer	11/06/2003 16:12:05

Trigraph

AutoID	C1	C2	C3	Duration	Application	TimeStamp
25461	R	E	S	269	Document20 - Microsoft Word	26/11/2003 15:57:49
25462	E	S	E	301	Document20 - Microsoft Word	26/11/2003 15:57:49
25463	A	R	C	472	Document20 - Microsoft Word	26/11/2003 15:57:51

Keyword

AutoID	Word	Duration	Application	TimeStamp
241	that	314	RE: - Message - Microsoft Word	11/10/2003 13:13:39
242	here	457	RE: - Message - Microsoft Word	11/10/2003 13:13:43
243	need	460	RE: - Message - Microsoft Word	11/10/2003 13:14:04

Table 7.2 - Example keystroke log entries

While digraph and trigraph logging were based upon all keystrokes entered, keyword logging was based on a look up list. The top 200 commonly occurring words in the English language (based on the lexicon provided by the Oxford English Dictionary) were stored in the database file, and as each word was entered, its latency was recorded (the

complete word list can be found in Appendix C). The list used by the keylogger contained a small number of additional keywords for later investigation e.g. usernames and passwords that have been removed from the list presented in the appendix for confidentiality.

7.2.2 Filtering

As with the first trial, extreme short/long digraph latencies that could adversely affect the distribution of digraph times were excluded from the log files. In the first trial the range was restricted to 40ms – 750ms (i.e. any digraph pair whose latency fell outside a nominal range). Unfortunately, the low pass filter was responsible for substantial quantities of data being removed from the user profiles and, as such, was reduced to 10ms for the purposes of this trial. If a digraph was removed due to the filtering, this also reset the trigraph and keyword logging so no further thresholds were needed for these two measures.

7.3 Trial Participants

For this experiment a total of 35 users were profiled over a period of three months. The trial participants were drawn from students and staff in the Network Research Group, the Department of Psychology and two external companies; TMA Global and John Nichols Builders Limited. As with the previous trial, several users disabled the keylogger when entering sensitive information and consequently forgot to re-enable it. Despite this, the key-logging trial collected considerable volumes of data with nearly six million samples collected across digraphs, trigraphs and keywords (Table 7.3). There was again considerable variation in the sample sizes with the smallest digraph log file of 15,951 samples and the largest with 353,867 samples.

User	Mean Digraph Latency (ms)	Typing Skill Classification	Digraphs	Trigraphs	Words
User 1	91	Best	34352	23352	1403
User 2	156	Average (skilled)	53306	36912	2599
User 3	99	Best	156718	107107	6154
User 4	251	Average (non-skilled)	27324	18688	1310
User 5	112	Good	50822	36713	1465
User 6	154	Average (skilled)	50167	34484	1885
User 7	106	Good	78579	54959	4349
User 8	130	Good	50102	35102	2932
User 9	97	Best	37618	24755	1741
User 10	145	Average (skilled)	70337	48942	4643
User 11	147	Average (skilled)	227660	145846	10617
User 12	102	Good	20216	14142	1032
User 13	157	Average (skilled)	65312	43015	1730
User 14	141	Average (skilled)	33639	23090	1784
User 15	139	Good	15951	11159	1068
User 16	150	Average (skilled)	42839	30299	2037
User 17	106	Good	105543	68068	3173
User 18	177	Average (skilled)	89730	59292	3121
User 19	117	Good	103876	71635	4617
User 20	121	Good	78597	53495	4479
User 21	141	Average (skilled)	80626	55881	2807
User 22	110	Good	117365	79534	6557
User 23	131	Good	118805	77013	5682
User 24	89	Best	201260	131954	8517
User 25	203	Average (skilled)	38944	26655	2266
User 26	192	Average (skilled)	48469	33907	2555
User 27	125	Good	33068	23115	1679
User 28	91	Best	70217	47033	2128
User 29	104	Good	88059	55707	3815
User 30	202	Average (skilled)	40741	28789	1007
User 31	86	Best	310823	211419	19726
User 32	93	Best	353867	237274	18056
User 33	144	Average (skilled)	276669	183455	6057
User 34	143	Average (skilled)	124409	87079	953
User 35	130	Good	140044	85413	6240
Totals			**3,436,054**	**2,305,283**	**150,184**

Table 7.3 - Participant typing skill

Before considering the data from each user, the typing skill for each participant was evaluated based on the categorisations proposed by Card et al. (1980), where typists are broadly categorised into one of five categories (Table 7.4). These results are also presented in Table 7.3 and presented graphically in Figure 7.2. The results are weighted towards typists with above average skills due to the nature of the test subjects (i.e. all

subjects were regular computer users who spent prolonged periods typing). This was considered acceptable as the likely use for a fully implemented system would be in environments with semi-skilled users (i.e. relatively few unskilled typists).

Category	Average Keystroke Interval (Seconds)
Best typist (135wpm)	0.08
Good typist (90wpm)	0.12
Average skilled typist (55wpm)	0.20
Average non-skilled typist (40wpm)	0.28
Worst typist – unfamiliar with keyboard	1.20

Table 7.4 - Classification of typist skill (Card et al. 1980)

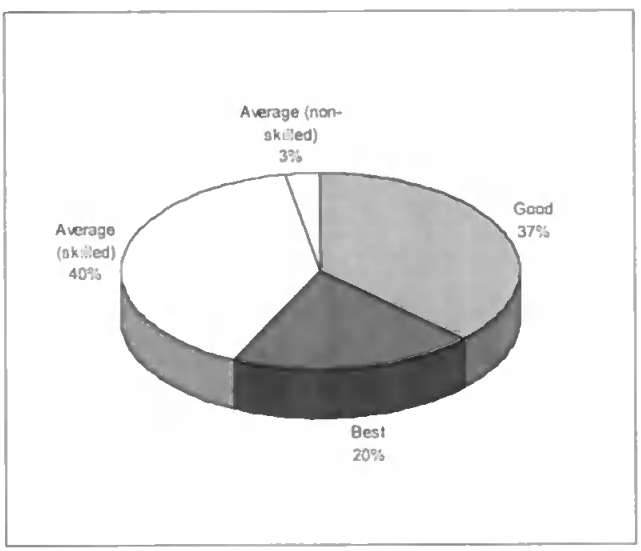


Figure 7.2 - Participant typing skills

The trial results presented in this chapter were based upon all 35 users (unlike the previous trial that had to eliminate three users due to small sample sizes). It is interesting to note that the relationship between the volume of data obtained and the False Acceptance Rate is not as might first be expected. If we consider Figure 7.3, it would be reasonable to assume that the relationship would be linear; i.e. as the number of samples increased, the FAR should decrease (as the system is able to construct better profiles). However, the figure indicates that there is no direct relationship between these two variables. This can be

assumed to indicate a stronger relationship between the user and the acceptance rate rather than the acceptance rate and the volume of data logged (i.e. simply having large volumes of data is not guaranteed to provide a lower false acceptance rate).

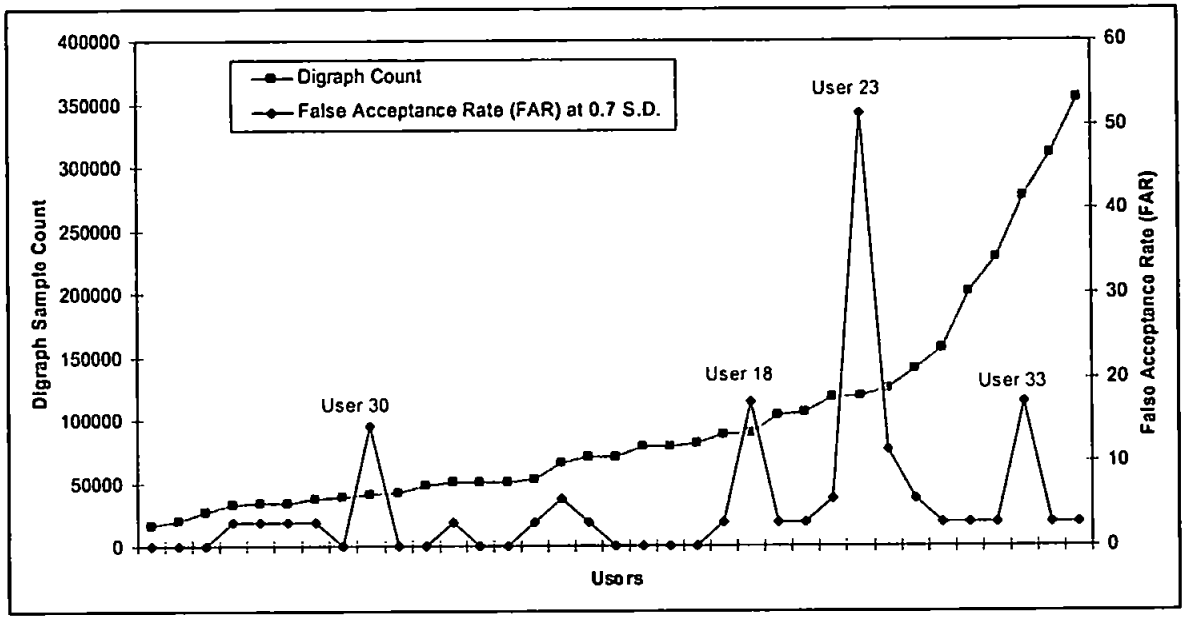


Figure 7.3 - Relationship between keylogger sample size and digraph FAR at 0.7 standard deviations

The results illustrated in Figure 7.3 show four users for whom high false acceptance rates are encountered. The results presented in this chapter (and in chapter 8) show that these users consistently produce higher FAR rates than other users – this is demonstrated later in this chapter when these users are removed and the results recalculated.

If we consider the deviation of a users' own typing, the standard deviation from the mean digraph latency is used to show the overall variance in a user's typing profile. The results from the trial participants are shown in Figure 7.4, which show significant variation across the users – effectively showing each users' consistency. The results are ordered by increasing average digraph latency with error bars indicating the standard deviation range

for each user. For reference, an ideal chart is shown in Figure 7.5 which shows a much more even distribution between users (and therefore better distinction between the classes of users). Figure 7.5 is based on the same mean latency with error bars indicating a 0.5 standard deviation from the mean.

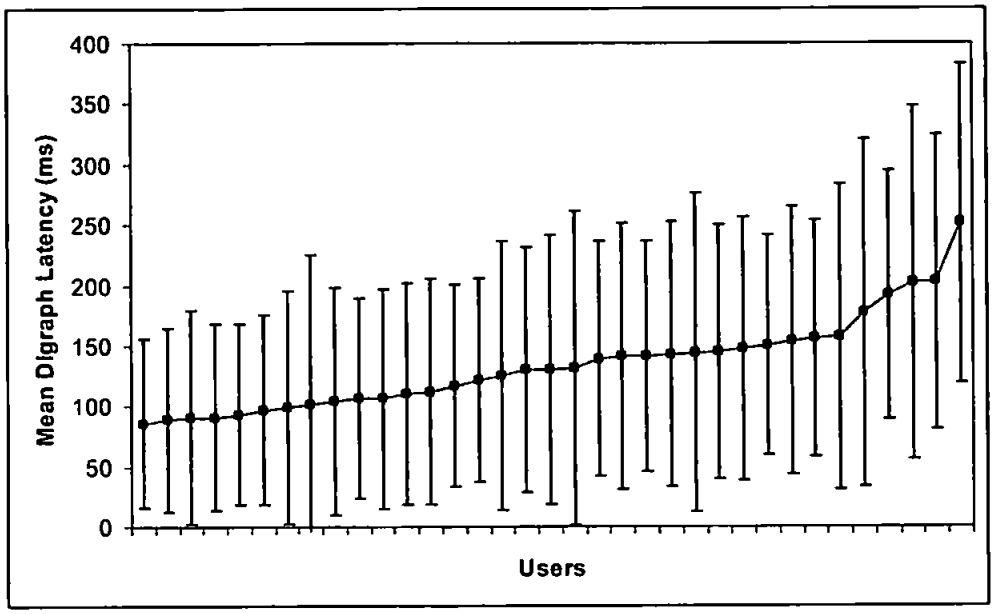


Figure 7.4 - Average digraph latency per user with standard deviation (ordered by mean latency)

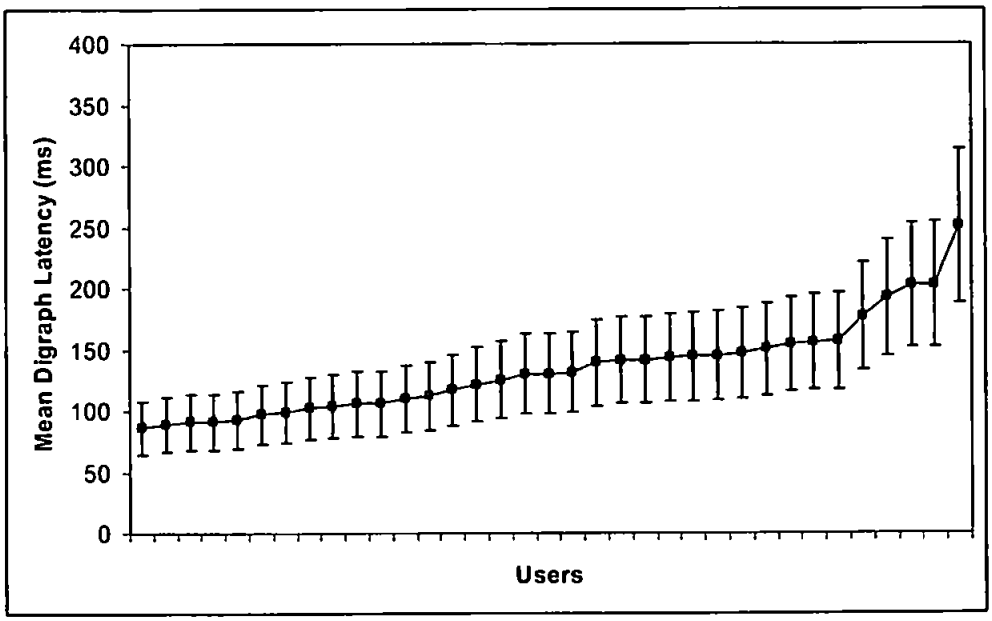


Figure 7.5 - 'Ideal' chart based on Figure 7.4

7.4 Analysis

NB due to the volume of data produced in this study, the body of the thesis, and the accompanying material in appendix C, contain only selected results. The full results can be found on the accompanying CD.

In the previous trial the experimental data for each user was processed off-line to calculate the mean and standard deviation values for each unique digraph pair. This process was again conducted on the captured data from this trial, but was repeated for each stored metric (digraphs, trigraphs and keyword latencies). Due to the volume of data, a profile generating utility was developed in Visual Basic (illustrated in Figure 7.6).

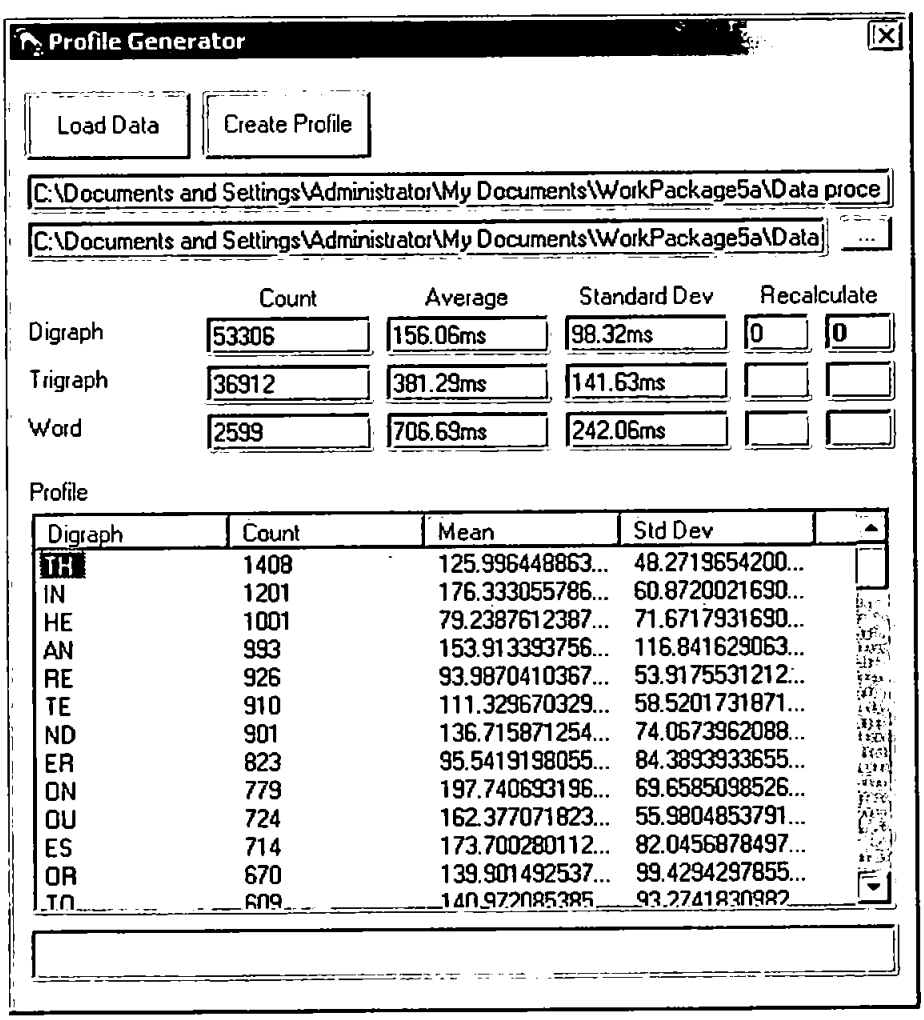


Figure 7.6 - Profile generator

The profile generator ran a series of queries against the original data (digraph, trigraph and keyword), and produced a table of results for each profiled value. For each of these, the count of samples, mean and standard deviation values were calculated and then stored to another Access database (the profile database). This resulted in a single database file containing 105 reference profiles (one each digraph, trigraph and keyword profile for each user – 3 x 35). As with the previous trial, in cases where the standard deviation was greater than the mean, the dataset was reduced by 10% to tighten the distribution. In a small number of cases (two users) this automatic adjustment did not provide sufficient change in the variance of the calculated values. In these cases, the data was recalculated manually for the problem digraphs. Where a digraph (or trigraph/word) required recalculation, this was shown as a cumulative count in the profile generator. The figure in red indicated the number of digraphs that required manual intervention (a total of eight digraph pairs from two users required this manual procedure – there were no incidents of trigraphs/words requiring manual intervention).

Completed profiles were then compared against the original keylogger files to determine the proportion of logged digraphs that were represented in the user profile. The quantity of unmatched samples (and resulting proportion represented as a percentage) are shown in Table 7.5 and graphically in Figure 7.7 (the results shown in Figure 7.7 are ordered by ascending sampled digraphs). These results show a clear relationship between the size of the raw sampled data (i.e. the number of sampled digraphs) and the proportion of matched digraphs when compared with the generated profile. This is significant as it demonstrates the importance of a substantial volume of keystroke data required to generate a usable profile.

User	Total Digraphs	Unmatched	Unmatched Percentage
User 1	34352	20869	60.8%
User 2	53306	17425	32.7%
User 3	156718	17179	11.0%
User 4	27324	15211	55.7%
User 5	50822	18636	36.7%
User 6	50167	17679	35.2%
User 7	78579	15547	19.8%
User 8	50102	18381	36.7%
User 9	37618	18794	50.0%
User 10	70337	15505	22.0%
User 11	227660	13908	6.1%
User 12	20216	12600	62.3%
User 13	65312	19509	29.9%
User 14	33639	17201	51.1%
User 15	15951	11747	73.6%
User 16	42839	15776	36.8%
User 17	105543	20026	19.0%
User 18	89730	17888	19.9%
User 19	103876	18549	17.9%
User 20	78597	16553	21.1%
User 21	80626	16647	20.6%
User 22	117365	17616	15.0%
User 23	118805	20181	17.0%
User 24	201260	16608	8.3%
User 25	38944	16997	43.6%
User 26	48469	16321	33.7%
User 27	33068	17817	53.9%
User 28	70217	21064	30.0%
User 29	88059	19817	22.5%
User 30	40741	17569	43.1%
User 31	310823	14563	4.7%
User 32	353867	14378	4.1%
User 33	276669	16721	6.0%
User 34	124409	15696	12.6%
User 35	140044	18066	12.9%

Table 7.5 - Unmatched captured digraphs

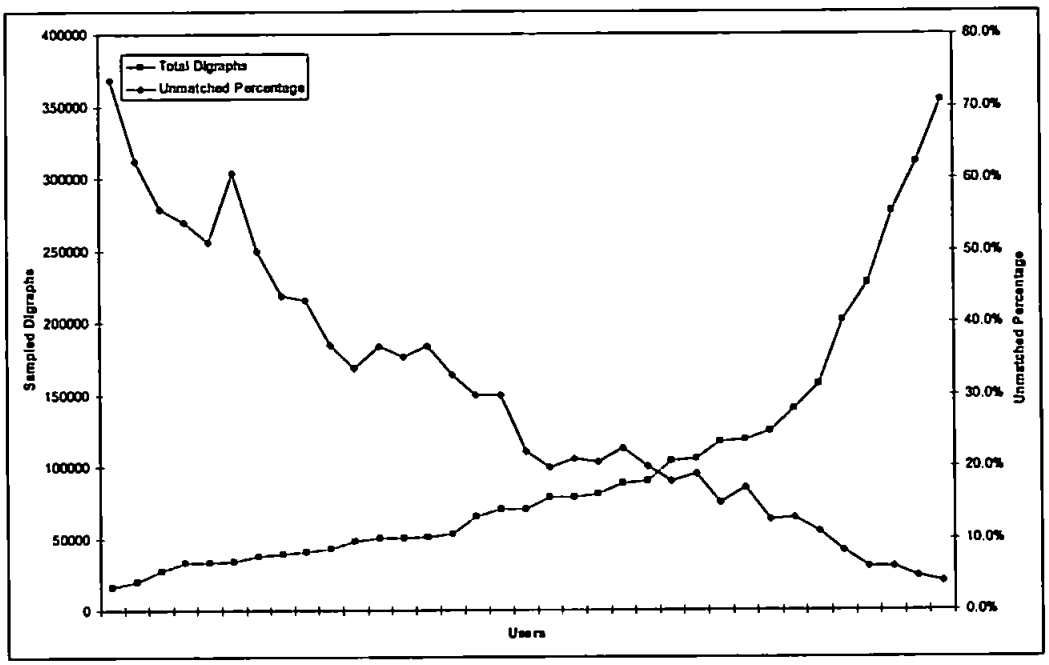


Figure 7.7 - Unmatched digraphs compared with digraph sample size

Once all the user profiles were calculated, the data comparator was used to generate tables of results for each of the methods. The data comparator (Figure 7.8) was based upon the original analyser described in section 6.5 in the previous chapter. A small number of additional features were introduced to the comparator to cater for the inclusion of trigraphs and keyword profiles. Firstly, a series of radio buttons were included to allow the selection of profile metrics (to compare based on digraphs, trigraphs or keywords). Secondly, a check box was added to allow the alert level (as described in the previous chapter) to be increased by unmatched digraphs.

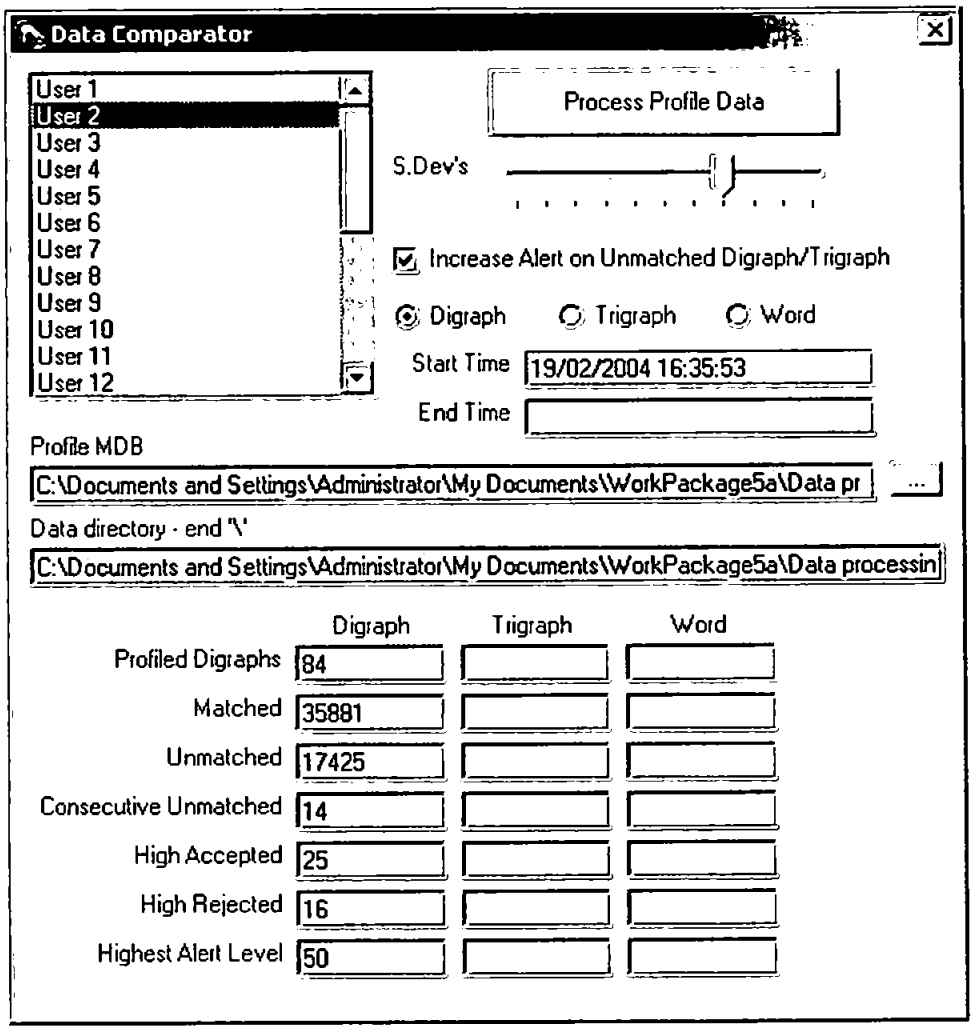


Figure 7.8 - Data comparator

In the previous trial, when a digraph was processed that did not exist in the reference profile, the alert level remained static (simply increasing the count of unmatched digraphs). This trial considered the role of unmatched digraphs as they are a potential indicator of impostor activity – i.e. if a user types a specific digraph pair infrequently (to the extent that there is insufficient data on which to base a profile), it is reasonable to assume that these occurrences are un-representative of that user's normal typing behaviour. By default, in this trial, an unmatched digraph increases the alert level by one, whilst a matched accepted/rejected digraph pair varies the alert level by two accordingly. This behaviour

can be adjusted by selecting the checkbox – once unchecked, the alert level is not affected by unmatched samples.

Before starting the full profile comparisons a trial comparison was conducted based upon a random selection of five users in order to determine the optimum settings for the deviation threshold. In the previous study the deviation settings were chosen from a range of 0.5, 1.0, 1.5 and 2.0 standard deviations with the best results obtained at 0.5. In order to determine an optimum setting, profile comparisons were made between 0.5 and 1.0 standard deviations (values below 0.5 had already been assessed). For the randomly selected users the best results were obtained at 0.7, with an increase in alert level above and below this threshold. As such, the later comparisons were performed with standard deviations settings of 0.6, 0.7 and 0.8.

Once the profile comparison was started, each users' reference profile was loaded and then compared against the raw keylogger data files for all 35 users. This produced a table of 35 sets of results for each user (i.e. 35 tables each with 35x6 result values). This process was repeated for trigraphs and keywords with the three different profile deviation settings (0.6, 0.7 and 0.8 standard deviations from the mean). It should be noted that a setting of 0.5 standard deviations was introduced to the trigraph and keyword comparisons due to poor performance at 0.6 and 0.7. This provided approximately 125,000 result values. With an average of nearly 100,000 samples per data file, each data comparison took approximately two hours, with a total of 18 comparisons conducted – six for digraph, eight for trigraph and four for keywords (see Table 7.6).

Metric	Standard Deviations (S.D.)
Digraphs	0.6, 0.7, 0.8 S.D.
Trigraphs	0.5, 0.6, 0.7, 0.8 S.D. *0.5 added due to poor performance at 0.6 and 0.7*
Keyword	0.5, 0.6, 0.7, 0.8 S.D. *0.5 added due to poor performance at 0.6 and 0.7*

Table 7.6 - Profile comparison settings

0.6 S.D.	DIGRAPHS	Reference Profile : User 3				
User	Matched	Unmatched	Consecutive unmatched	High accepted	High rejected	Highest alert level
User 1	28143	6209	65	12	18	15513
User 2	46797	6509	10	9	28	42469
User 3	139539	17179	11	17	14	72
User 4	24234	3090	15	6	40	32450
User 5	42111	8711	23	12	32	21713
User 6	43399	6768	16	14	29	36460
User 7	69491	9088	10	14	18	29808
User 8	44356	5746	9	10	28	30932
User 9	31808	5810	87	11	18	14980
User 10	62110	8227	6	8	29	49329
User 11	194687	32973	29	13	32	165327
User 12	17574	2642	6	8	32	20580
User 13	53395	11917	95	19	32	54845
User 14	29885	3754	7	10	21	22528
User 15	14037	1914	6	9	21	9774
User 16	36891	5948	7	11	38	32348
User 17	87249	18294	14	18	25	36618
User 18	78351	11379	10	11	34	65447
User 19	89323	14553	28	11	49	42925
User 20	68617	9980	10	11	48	41558
User 21	70145	10481	24	10	26	61197
User 22	100206	17159	22	14	21	45379
User 23	97646	21159	45	12	24	70889
User 24	169445	31815	16	16	22	56372
User 25	33843	5101	6	7	38	45697
User 26	42281	6188	6	6	38	54008
User 27	28297	4771	10	13	15	10809
User 28	60018	10199	19	14	16	10513
User 29	75695	12364	13	13	34	31960
User 30	34376	6365	24	13	51	37715
User 31	269989	40834	48	17	23	5486
User 32	304826	49041	29	15	23	150798
User 33	216573	60096	31	11	40	197230
User 34	100837	23572	25	20	28	94940
User 35	117220	22824	48	14	26	61550

Table 7.7 - Sample output file

Once the profile comparison was completed, the results were exported to an Excel spreadsheet (an example is shown in Table 7.7). In this table, the highlighted result line indicates the comparison between user 3's raw keylogger data and their generated reference profile with the other rows indicating the comparison of other users (i.e. impostors) compared against the reference profile (user 3). The spreadsheets contained a number of functions to derive 2-dimensional tables of data from the raw results from the comparator (Table 7.8) from which the FAR/FRR figures could be derived.

	User 1	User 2	User n	User 35
User 1	54	39186	15513	57030
User 2	16277	54	42469	112493
User n	51355	27329	72	44405
User 35	58229	91420	21713	50

Table 7.8 - Combined results showing highest alert levels

Following the basic analysis described in this section, a further modification was made to the comparator to determine how many keystrokes were needed before either the valid user was challenged or an impostor detected. The threshold for this challenge was based upon the best performance thresholds from the earlier trials and was initially set at an alert level of 70. The results from this trial using the digraph keylogger files at a threshold of 0.7 standard deviations are presented in Table 7.9. The results from this trial were somewhat variable, while some users had good results (e.g. users 7, 10 and 26), most user profiles had only moderately successful results. If we consider user 2, while 29/34 (85%) impostors were challenged in less than 100 digraphs, user 16 (when acting as an impostor against user 2's reference profile) was able to type over 40,000 digraphs before being challenged.

0.7 SD	Reference Profiles																																		
	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17	User 18	User 19	User 20	User 21	User 22	User 23	User 24	User 25	User 26	User 27	User 28	User 29	User 30	User 31	User 32	User 33	User 34	User 35
User 1	56	21	72	22	22	22	50	24	24	38	31	23	23	22	21	21	66	33	43	22	10	84	116	52	18	21	40	41	52	21	16	93	135	18	106
User 2	29	63306	23	18	26	280	65	29	28	320	63306	26	63306	67	64	60	125	63306	79	321	64	27	127	23	49	16	27	24	54	120	27	19	63306	63306	624
User 3	39	15	156718	19	26	22	20	30	20	14	28	19	17	21	21	20	127	14	28	15	14	60	98	79	15	19	23	79	29	18	49	325	22	23	25
User 4	22	68	15	300	15	37	16	23	19	55	22	27	53	17	22	32	22	178	18	19	29	18	63	15	80	66	25	16	16	120	15	18	34	53	20
User 5	25	31	34	19	73	31	31	30	35	22	22	26	29	32	27	29	145	17	20	29	21	141	50822	51	20	31	34	24	73	29	74	124	50822	30	34
User 6	18	54	26	22	70	141	43	13	23	60	347	25	156	22	23	32	44	3435	77	104	63	68	76	17	26	26	39	17	40	37	30	43	50167	203	145
User 7	43	101	372	27	145	108	78579	54	50	36	78579	55	247	42	38	39	78579	90	78579	78579	28	78579	78579	2183	16	25	68	205	408	68	78579	78579	78579	280	78578
User 8	21	30	27	26	25	39	29	225	27	30	257	25	96	34	26	24	72	63	128	24	29	30	126	24	34	26	18	77	94	22	27	107	50102	93	489
User 9	46	15	116	13	45	28	40	64	105	19	19	19	27	19	24	20	125	20	27	34	15	296	11107	110	16	16	21	252	250	17	400	316	26	34	14
User 10	23	22	20	18	21	17	25	33	18	70337	83	21	26	32	21	17	41	40	48	41	28	49	29	20	22	22	49	17	21	16	34	26	70337	44	101
User 11	20	84	25	17	19	88	15	63	16	144	227660	43	117	26	21	35	22	113	11	63	65	61	725	26	51	48	19	12	21	60	21	17	227660	145	227660
User 12	26	103	42	33	21	102	40	79	22	38	180	94	120	30	21	46	30	20216	64	69	82	46	83	21	25	18	41	28	32	200	26	49	177	103	48
User 13	24	31	24	33	27	41	23	46	15	30	265	33	107	36	27	28	52	86	28	40	51	87	182	52	20	33	43	23	25	53	15	23	65312	40	172
User 14	36	33	19	16	35	34	64	64	40	192	33639	20	49	78	32	31	217	61	81	150	81	139	1162	50	17	15	63	35	169	71	87	38	641	38	33639
User 15	17	44	12	20	23	29	23	15	17	17	15951	19	15	18	23	17	28	15951	20	46	61	18	15951	20	14	34	32	23	57	62	14	14	15951	62	15951
User 16	19	42839	17	21	18	37	17	46	26	1085	42	43	42839	39	40	60	22	42839	21	23	35	21	1111	15	144	61	42	15	19	642	15	15	42839	42839	31
User 17	33	21	85	24	51	35	89	83	53	17	56	23	49	45	23	21	105543	69	61	106	64	67	209	38	19	20	33	29	55	22	100	214	105543	38	1354
User 18	32	28	51	25	23	30	47	17	15	28	144	32	46	33	23	46	51	166	44	33	46	44	43	21	34	23	22	24	67	33	45	36	191	236	61
User 19	23	36	121	21	31	32	69	50	26	44	372	20	83	51	24	48	104	45	103876	172	33	103876	1606	61	30	22	47	56	1229	30	101	91	157	60	103876
User 20	21	67	54	20	34	64	62	85	48	841	78597	21	102	81	36	28	66	65	35	78597	73391	131	121	47	32	32	63	34	48	26	49	44	78597	132	78597
User 21	19	2555	38	22	39	206	106	129	25	149	80626	48	2850	89	22	648	84	3444	199	211	80626	131	120	61	52	69	77	38	104	370	78	105	80626	80626	80626
User 22	28	32	1160	18	29	90	29	28	24	30	343	23	27	30	19	36	23129	30	315	70	32	117365	117365	61936	19	19	30	37	3608	19	70454	190	117365	91	117365
User 23	20	31	14	20	16	20	14	21	20	29	30	28	18	30	25	16	21	42	30	24	24	19	23	24	29	24	22	13	25	30	32	30	30	31	19
User 24	45	18	602	14	34	48	45	86	33	22	39	19	52	26	28	20	90	18	29	15	18	347	201260	201260	14	16	26	135	29	20	58	201260	153	24	201260
User 25	19	14	13	24	16	18	13	17	15	22	47	22	14	26	19	17	13	15	13	13	13	13	13	13	39	41	13	14	14	49	13	13	44	19	15
User 26	16	16923	15	38	37	162	20	86	22	58	18527	84	802	31	26	489	38	48463	65	158	417	56	146	21	189	48469	30	16	56	2127	17	17	48469	48469	271
User 27	27	22	110	17	38	33	40	41	25	29	65	26	69	22	22	20	100	63	19	41	37	33068	33068	111	13	36	69	20	98	37	2079	85	33068	40	33068
User 28	27	20	70217	18	28	17	51	17	29	19	18	22	16	22	22	21	18	20	18	19	19	15	95	426	18	18	22	40	33	17	102	23	62	15	55
User 29	27	32	58	20	27	24	33	33	17	33	35	20	24	24	20	24	44	32	17	18	17	42	68	27	16	16	18	32	39	27	49	35	68	29	129
User 30	21	42	38	23	27	44	34	53	20	50	43	31	50	37	20	49	34	56	37	35	43	34	36	27	47	31	34	37	37	71	23	41	82	50	45
User 31	43	26	310823	16	61	34	66	25	39	22	14	20	26	17	22	21	310823	37	106	83	20	310823	310823	310823	17	19	38	170	3E+05	23	310823	310823	86	69	310823
User 32	23	13	52	17	36	35	53	32	25	14	273	21	23	23	25	13	112	39	69	40	13	67	404	32	18	15	32	82	267	31	32	353867	349	14	353867
User 33	39	25	39	23	26	34	32	81	20	25	27	23	26	28	20	33	76	22	22	26	26	87	122	93	17	32	23	29	76	26	74	84	276669	26	26
User 34	28	78	27	19	81	95	52	26	43	64	923	22	96	30	33	32	68	120	54	65	148	88	664	49	38	26	40	39	55	39	90	91	954	930	819
User 35	21	25	20	22	16	19	41	18	24	40	87	27	20	31	21	17	73	67	41	33	33	128	126	52	16	20	24	34	31	54	62	47	69	51	140044

Table 7.9 - Number of keypresses before a challenge

The results in Table 7.9 can also be considered in terms of the average number of keystrokes required before a challenge is issued. The results show that an average of 6,390 digraphs were accepted before an impostor was challenged compared with an average of 68,755 digraphs before the valid user was challenged. While these results seem to provide the appropriate differentiation between impostor and valid user, giving an impostor the opportunity to type over 6,000 digraphs presents a major security risk. This figure is reduced to 4,300 digraphs if the users identified in Figure 7.3 are removed (i.e. the users for whom keystroke analysis is shown to be unworkable).

For this trial the False Rejection Rate (FRR) was fixed at 0% (i.e. the valid user would not have been rejected by the system). The False Acceptance Rates (FAR) were calculated for each user at the deviation thresholds specified in Table 7.6 and are shown in Table 7.10.

When the results were calculated, the False Acceptance Rates per user were averaged across all users to provide an average FAR for each metric. The averaged results for this approach are shown in Table 7.11. It should be noted that the keyword latencies did not use the unmatched alert increase due to the use of a word list/dictionary. It was not appropriate to include an unmatched alert increase as the system is designed to monitor specific word occurrences unlike the digraph/trigraph approach that monitors all possible combinations – i.e. only words that are being monitored would actually be logged.

	Digraph FAR						Trigraph FAR								Keyword FAR			
Standard Deviation	0.6		0.7		0.8		0.5		0.6		0.7		0.8		0.5	0.6	0.7	0.8
Unmatched Alert	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	N	N	N
User 1	2.9	2.9	2.9	8.6	2.9	8.6	8.6	17.1	8.6	8.6	8.6	25.7	8.6	31.4	97.1	97.1	91.4	91.4
User 2	0.0	2.9	2.9	5.7	2.9	11.4	34.3	0.0	34.3	34.3	34.3	5.7	31.4	8.6	2.9	2.9	2.9	5.7
User 3	0.0	0.0	2.9	5.7	5.7	11.4	74.3	5.7	62.9	62.9	54.3	20.0	40.0	25.7	0.0	8.6	20.0	28.6
User 4	0.0	0.0	0.0	0.0	0.0	2.9	5.7	0.0	5.7	5.7	5.7	0.0	5.7	0.0	5.7	0.0	0.0	0.0
User 5	0.0	0.0	0.0	8.6	0.0	11.4	28.6	2.9	28.6	28.6	28.6	14.3	28.6	22.9	8.6	11.4	14.3	20.0
User 6	0.0	2.9	0.0	8.6	0.0	22.9	28.6	2.9	28.6	28.6	28.6	5.7	25.7	11.4	34.3	11.4	2.9	8.6
User 7	0.0	0.0	0.0	0.0	0.0	2.9	34.3	0.0	28.6	28.6	25.7	0.0	20.0	0.0	0.0	0.0	0.0	0.0
User 8	11.4	5.7	2.9	5.7	2.9	40.0	28.6	17.1	28.6	28.6	25.7	28.6	25.7	37.1	0.0	0.0	2.9	8.6
User 9	2.9	2.9	2.9	0.0	0.0	14.3	17.1	5.7	17.1	17.1	17.1	11.4	14.3	20.0	22.9	25.7	34.3	34.3
User 10	0.0	0.0	0.0	8.6	2.9	14.3	37.1	0.0	31.4	31.4	28.6	2.9	22.9	8.6	2.9	0.0	0.0	0.0
User 11	0.0	0.0	2.9	14.3	17.1	22.9	82.9	2.9	65.7	65.7	51.4	0.0	11.4	11.4	37.1	5.7	2.9	28.6
User 12	0.0	5.7	0.0	5.7	0.0	14.3	2.9	14.3	2.9	2.9	2.9	22.9	2.9	28.6	22.9	25.7	31.4	40.0
User 13	8.6	11.4	5.7	11.4	14.3	11.4	40.0	2.9	40.0	40.0	37.1	11.4	37.1	11.4	8.6	8.6	11.4	14.3
User 14	2.9	2.9	2.9	8.6	0.0	34.3	8.6	8.6	8.6	8.6	8.6	22.9	5.7	28.6	0.0	2.9	5.7	5.7
User 15	0.0	0.0	0.0	2.9	0.0	8.6	0.0	14.3	0.0	0.0	0.0	17.1	0.0	20.0	22.9	17.1	25.7	25.7
User 16	0.0	5.7	0.0	2.9	0.0	5.7	22.9	0.0	20.0	20.0	14.3	2.9	14.3	5.7	0.0	0.0	2.9	14.3
User 17	0.0	0.0	2.9	11.4	14.3	25.7	48.6	2.9	48.6	48.6	45.7	8.6	40.0	11.4	5.7	5.7	5.7	11.4
User 18	8.6	11.4	17.1	8.6	17.1	25.7	48.6	17.1	42.9	42.9	40.0	0.0	37.1	5.7	28.6	8.6	11.4	17.1
User 19	0.0	2.9	2.9	5.7	2.9	8.6	65.7	0.0	62.9	62.9	62.9	8.6	62.9	11.4	5.7	14.3	20.0	17.1
User 20	0.0	0.0	0.0	5.7	5.7	22.9	42.9	0.0	40.0	40.0	31.4	5.7	28.6	8.6	0.0	2.9	2.9	5.7
User 21	0.0	0.0	0.0	0.0	2.9	2.9	40.0	0.0	28.6	28.6	17.1	2.9	5.7	2.9	0.0	2.9	5.7	8.6
User 22	2.9	8.6	5.7	8.6	8.6	25.7	65.7	0.0	57.1	57.1	45.7	5.7	31.4	17.1	0.0	5.7	5.7	11.4
User 23	68.6	54.3	51.4	62.9	60.0	80.0	71.4	45.7	71.4	71.4	68.6	34.3	68.6	45.7	37.1	37.1	45.7	45.7
User 24	2.9	0.0	2.9	5.7	5.7	8.6	62.9	2.9	54.3	54.3	25.7	14.3	14.3	28.6	0.0	2.9	8.6	14.3
User 25	0.0	2.9	0.0	2.9	0.0	14.3	14.3	0.0	8.6	8.6	8.6	2.9	8.6	0.0	2.9	0.0	0.0	0.0
User 26	0.0	0.0	0.0	0.0	0.0	5.7	22.9	0.0	22.9	22.9	20.0	0.0	17.1	0.0	5.7	0.0	0.0	5.7
User 27	2.9	2.9	2.9	17.1	2.9	48.6	11.4	11.4	11.4	11.4	11.4	28.6	11.4	45.7	8.6	2.9	5.7	5.7
User 28	2.9	2.9	2.9	5.7	2.9	17.1	42.9	8.6	42.9	42.9	42.9	17.1	42.9	14.3	8.6	17.1	28.6	37.1
User 29	5.7	5.7	2.9	22.9	20.0	48.6	48.6	11.4	42.9	42.9	42.9	20.0	42.9	22.9	11.4	8.6	14.3	31.4
User 30	14.3	8.6	14.3	14.3	11.4	20.0	22.9	20.0	22.9	22.9	22.9	22.9	20.0	37.1	45.7	28.6	28.6	8.6
User 31	0.0	0.0	2.9	2.9	5.7	8.6	0.0	0.0	0.0	0.0	0.0	2.9	0.0	5.7	0.0	0.0	2.9	5.7
User 32	0.0	0.0	2.9	8.6	11.4	20.0	54.3	0.0	0.0	0.0	0.0	8.6	0.0	14.3	2.9	5.7	14.3	14.3
User 33	17.1	14.3	17.1	14.3	22.9	28.6	88.6	77.1	82.9	82.9	65.7	54.3	57.1	71.4	80.0	68.6	20.0	28.6
User 34	5.7	2.9	11.4	14.3	14.3	14.3	60.0	0.0	48.6	48.6	42.9	0.0	37.1	2.9	97.1	97.1	97.1	97.1
User 35	22.9	14.3	5.7	34.3	25.7	54.3	71.4	25.7	65.7	65.7	65.7	25.7	62.9	45.7	34.3	5.7	11.4	14.3
Average	5.2	5.0	4.9	9.8	8.1	20.5	38.2	9.1	33.3	33.3	29.5	13.0	25.2	18.9	18.3	15.2	16.5	20.2

Table 7.10 - Results from single-metric measures

Metric	S.D.	Unmatched Alert	FAR
Digraphs	0.6	No	5.2%
	0.6	Yes	5.0%
	0.7	No	4.9%
	0.7	Yes	9.8%
	0.8	No	8.1%
	0.8	Yes	20.5%
Trigraphs	0.5	No	38.2%
	0.5	Yes	9.1%
	0.6	No	33.3%
	0.6	Yes	33.3%
	0.7	No	29.5%
	0.7	Yes	13.0%
	0.8	No	25.2%
	0.8	Yes	18.9%
Words	0.5	No	18.3%
	0.6	No	15.2%
	0.7	No	16.5%
	0.8	No	20.2%

Table 7.11 - Final results

While the results shown in Table 7.11 show some encouraging FAR levels there is still significant variation with the best results obtained at 0.7 standard deviations for digraphs, 0.5 standard deviations for trigraphs (with increased alert levels for unmatched digraphs) and 0.6 for keywords. However, when the full results are considered (as shown in Table 7.10), even at the optimum settings, certain users show high FAR levels (e.g. user 23's profile returned FAR levels of 51.4%, 45.7% and 37.1% respectively for digraph, trigraph and keywords at the optimum settings). It can also be clearly observed that the results for trigraphs and keywords are significantly worse when compared with those for digraphs – this is most likely to be related to the number of underlying samples used for these techniques (i.e. the number of sampled digraphs were significantly higher than that for trigraphs and keywords, with a corresponding increase of samples per digraph). It is probable that over a longer period of time, the profiles could be refined for trigraphs and

keywords to produce a more distinct user profile with a corresponding reduction in the FAR.

These results also demonstrate that the techniques can be very effective for some users while very ineffective for others. For example, when considering digraph FAR's at 0.6 standard deviations (where 0% FAR was actually experienced for 19 out of the 35 users – 54.3%) the average FAR (5.2%) has been heavily influenced by a single user (user 23), whose 68.6% FAR dramatically increases the average. In a full implementation, it would be likely that the use of keystroke analysis would only form a part of a comprehensive user monitoring system. As such, a users' typing would only be monitored if the method was shown to be a discriminating authentication technique for that user. The removal of user 23 from the results in Table 7.10 significantly affects the average FAR's presented in Table 7.11, reducing the best digraph results from 4.9% to 3.5%, trigraph results from 9.1% to 8.0% and keywords from 15.2% to 14.5%.

Further optimisation can be achieved by removing the worst 5 participants (15%) from the trial results. This provides an improvement in the results of the technique with average FAR's as low as 1.7% for digraphs, 4.4% for trigraphs and 12.8% for keywords (Table 7.12). While the keyword FAR in particular remains unacceptably high, a reference back to Table 7.10 reveals that there were still almost a third of users for whom 0% FAR was observed at the 0.5 standard deviation threshold. This suggests a clear potential for using the technique in a subset of cases – which could also increase if additional keyword typing samples were obtained to support the profiling.

Metric	S.D.	Unmatched Alert	FAR
Digraphs	0.6	No	1.7%
	0.6	Yes	2.4%
	0.7	No	2.2%
	0.7	Yes	7.0%
	0.8	No	4.9%
	0.8	Yes	17.0%
Trigraphs	0.5	No	34.5%
	0.5	Yes	4.4%
	0.6	No	29.3%
	0.6	Yes	29.3%
	0.7	No	25.6%
	0.7	Yes	10.6%
	0.8	No	21.2%
	0.8	Yes	15.2%
Words	0.5	No	13.8%
	0.6	No	12.8%
	0.7	No	15.3%
	0.8	No	19.7%

Table 7.12 - Optimised results

7.5 Conclusions

This chapter has presented the results of a series of trials aimed at implementing and evaluating keystroke analysis with a large number of trial participants over a longer period of time than the trials outlined in the previous chapter.

It is clear from the results presented here that there is considerable potential for continuous user authentication based upon keystroke analysis. The long-term sampling of keystroke digraphs has served to reinforce the validity of the technique, while the introduction of trigraph and keyword-based monitoring has provided additional metrics that can be used as alternative (or complimentary) techniques. In particular, the use of keyword monitoring has considerable potential when used to monitor for specific, high-risk typed words (e.g. delete, format etc.).

Chapter 8

Extending Keystroke Analysis

8.1 Introduction

The previous chapter presented a range of results derived from single-metric measures. While these measures provided good overall results, there was potential for improvement. This chapter considers the use of a composite approach to keystroke analysis, combining the three previously evaluated metrics – digraphs, trigraphs and keywords. It should be noted that the results presented in this chapter represent a subset of the overall results. Due to the number of variations in thresholds and standard deviation settings for each metric it was not possible to evaluate every possible setting and threshold. As such, this chapter presents initial *optimum* settings that provide *reasonable* results (the future work in the next chapter will suggest further refinements to this approach).

In addition to the composite approach, this chapter also describes a brief trial conducted using neural networks in an attempt to improve the performance of the digraph-based keystroke analysis considered in the previous chapter. Finally, the discussion presents details of a prototype real-time keystroke analysis system developed to demonstrate the concept of continuous, non-intrusive user authentication.

8.2 A composite approach

In order to evaluate a composite approach to keystroke analysis, the three separate keylogger data files described in the previous chapter were recombined (using Access queries) to create a single composite keylogger table. This resulted in a 700MB Access database containing 35 user keylogger tables.

Once the composite keylogger database was complete, a composite profile comparator was developed. This application was again developed in Visual Basic and was based on the data comparator presented in the previous chapter. The composite comparator (Figure 8.1) had a number of additional options and settings to determine the manner in which the composite data was compared to the reference profile.

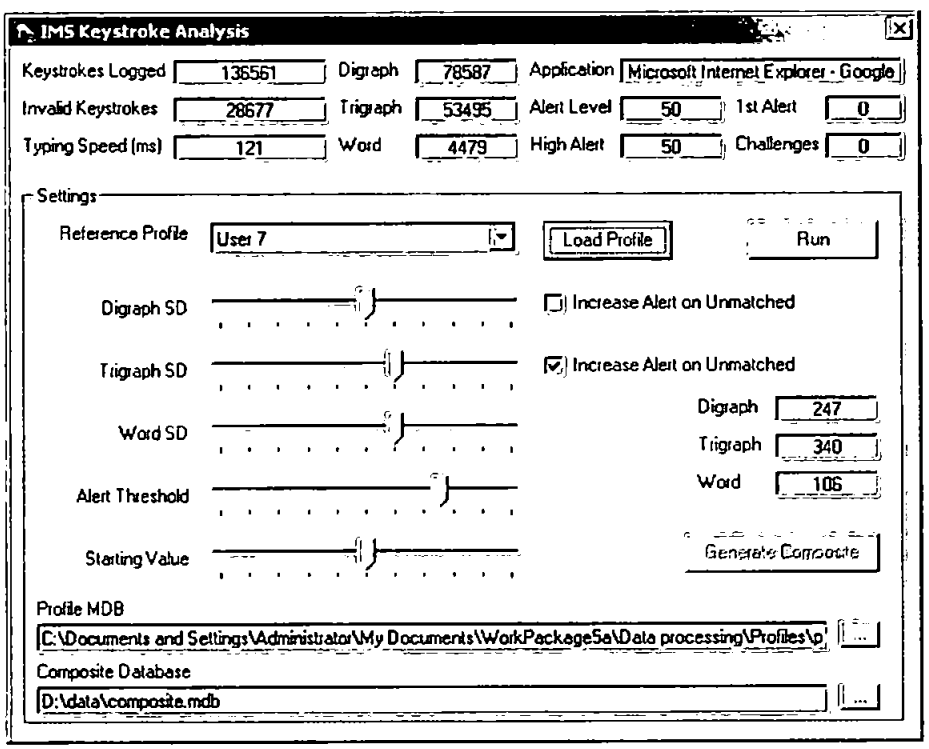


Figure 8.1 - Composite data comparator (running)

In the original comparator, each metric (digraph, trigraph and keyword) was evaluated separately, with a standard deviation setting adjustable for each profile. For the composite approach, all three settings were adjustable simultaneously together with an alert threshold and a variable starting value for the alert level (i.e. the initial confidence threshold was variable). In addition, the unmatched alert level increase was selectable for both digraphs and trigraphs. Once the profile had been selected from the drop down list and the

appropriate settings selected, the comparator was started. The comparator performed the same basic operation as the single-metric version described in the previous chapter. The users' profile was loaded and the digraph, trigraph and keyword profiles stored in memory for comparison. The alert level was initialised to a pre-determined value and then the composite data was loaded and processed one sample at a time. Each sample (either digraph, trigraph or keyword) was compared against the user profile and the alert level adjusted dependent on the match/unmatched status and the deviation from the mean with a set tolerance (according to the standard deviation setting).

sample mean ± (sample standard deviation * permitted deviation)

In the previous experiments the main consideration was to evaluate the alert level and to determine the FAR rate while fixing the FRR rate at 0%. The aim of this constraint was to ensure minimum inconvenience to the legitimate user (i.e. the monitoring software was optimised to never challenge the valid user). In this trial, a level of user inconvenience was considered acceptable – on the assumption that the user would pass any explicit challenge issued by the system (the nature of these challenges could be based upon basic passwords, or the authentication methods considered in chapter 4). While a number of user profiles offered 0% FAR in the previous trial, the difference between an impostor and the genuine user was frequently small (e.g. in one case the genuine user reached a highest alert level of 50 compared with the nearest impostor who reached 66). By fixing the FRR at 0%, it is possible that the nearest matching impostor (even when detected) were not significantly different to that of the genuine user.

	Digraph SD	Trigraph SD	Word SD	Unmatched Alert	
				Digraph	Trigraph
User 1	0.6	0.5	0.7	No	No
User 2	0.6	0.5	0.5	No	No
User 3	0.6	0.5	0.5	No	No
User 4	0.7	0.6	0.6	No	No
User 5	0.6	0.5	0.6	No	No
User 6	0.7	0.6	0.6	No	No
User 7	0.7	0.6	0.6	No	Yes
User 8	0.7	0.6	0.5	No	Yes
User 9	0.7	0.6	0.6	No	No
User 10	0.7	0.6	0.6	No	No
User 11	0.7	0.6	0.6	No	Yes
User 12	0.6	0.5	0.6	No	No
User 13	0.7	0.6	0.6	No	Yes
User 14	0.6	0.5	0.6	No	No
User 15	0.7	0.6	0.6	No	No
User 16	0.6	0.5	0.6	No	No
User 17	0.6	0.5	0.6	No	No
User 19	0.7	0.6	0.6	No	Yes
User 20	0.7	0.6	0.6	No	Yes
User 21	0.7	0.6	0.6	No	Yes
User 22	0.7	0.6	0.6	No	No
User 24	0.7	0.6	0.6	No	Yes
User 25	0.7	0.6	0.6	No	No
User 26	0.7	0.6	0.6	No	Yes
User 27	0.6	0.5	0.6	No	No
User 28	0.7	0.6	0.6	No	No
User 29	0.6	0.5	0.6	No	No
User 31	0.7	0.6	0.6	No	Yes
User 32	0.7	0.6	0.6	No	Yes
User 34	0.6	0.5	0.6	No	No

Table 8.1 - Composite profile settings

In order to address this issue (and to potentially improve the FAR rates in the previous chapter), this trial concentrated upon the number of keystrokes accepted for each user before a challenge was issued. This effectively fixed the alert level for all users at 70, the starting confidence/alert level at 50 and then varied the standard deviation and unmatched alert increase settings for each user profile comparison. The application settings were varied for each profile comparison in order to achieve the best distinction between the

genuine user and the impostors. The initial settings for each user were based upon the best performing settings from the earlier single-metric trials, and were then adjusted to obtain the best results (i.e. the optimisation was considered complete when the genuine user was able to type significantly more than an impostor before a challenge would have been issued by the system). The settings used for each user are shown in Table 8.1.

The results presented in Table 8.2 aimed to only challenge the genuine user after a significant number of samples had been processed (i.e. a genuine user would be infrequently challenged compared with the impostors). These results also show that an average of 4,759 digraphs were accepted before an impostor was challenged compared with an average of 76,228 digraphs before the valid user was challenged. While allowing an impostor to type large amounts of text is a considerable security risk, the average figure (4,759 samples) is significantly affected by a small number of users. When looking at Table 8.2, it is clear that the majority of impostors are detected in less than 200 samples (comprising digraphs, trigraphs and keywords). For most users, 200 samples equates to approximately 2 minutes of typing and as such is sufficient time to detect and challenge an impostor who is not intent on conducting serious damage. It is however possible that 200 samples would be more than enough to allow an impostor to issue sufficient commands to cause serious damage. For example, an impostor typing a high risk command like "format" would only generate 5 digraphs, 4 trigraphs and one keyword sample – insufficient to cause a challenge using the techniques described in this section. As such, it would be necessary to monitor for specific strings requiring a more definitive authentication prior to allowing the execution of the command.

Reference Profiles	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17	User 18	User 20	User 21	User 22	User 24	User 25	User 26	User 27	User 28	User 29	User 31	User 22	User 34
User 1	34352	61	49	53	73	44	50	16	872	38	17	94	36	101	673	34	115	31	26	17	104	64	27	28	74	170	101	44	29	25
User 2	60	83306	27	46	50	161	55	83	34	489	202	75	159	87	83306	390	93	61	94	85	112	40	130	62	361	37	323	24	22	63306
User 3	649	96	15879	90	220	49	52	29	519	53	29	207	29	172	519	76	127	79	39	20	270	70	62	39	99	1706	76	43	32	76
User 4	72	60	16	27324	39	34	37	16	47	28	17	181	22	71	167	47	27	14	17	26	22	29	42	72	53	21	21	16	16	52
User 5	181	39	117	29	50822	54	103	17	71	229	35	298	22	1780	367	39	224	35	35	34	64922	67	30	27	505	110	106	35	26	37
User 6	61	37	20	31	61	320	34	20	21	42	103	176	37	41	67	35	331	42	50	32	17	23	43	42	910	17	223	113	16	48
User 7	744	67	102	52	17450	66	76579	66	71	23	47	265	33	47	415	2801	410	50	94	47	29	63	43	45	78579	272	223	295	297	29
User 8	45	30	25	24	63	61	68	71	23	47	30	44	47	47	415	410	50	34	47	29	50102	36	24	36	75	35	60	22	26	63
User 9	44	21	31	35	29	35	41	12	37698	34	32	40	16	69	77	29	41	33	19	21	44	102	23	29	56	64	410	47	19	21
User 10	48	69	12	63	32	38	29	28	52	70337	35	79	27	86	194	33	16	14	18	18	35	22	101	34	345	16	16	11	12	63
User 11	79	254	19	36	67	64	51	70	36	522	1387	59	64	447	345	59	72	49	64	54	720	53	38	42	94	34	40	49	35	317
User 12	142	133	61	137	63	72	49	39	71	121	56	20216	44	122	518	133	50	37	41	27	72	29	218	71	119	87	98	29	27	133
User 13	114	87	15	61	82	96	41	18	52	92	28	138	1043	92	216	27	87	18	22	22	215	29	124	46	49	16	24	12	19	88
User 14	137	27	64	48	156	53	87	25	100	46	361	91	44	33639	115	85	39	49	45	48	678	48	77	39	444	50	171	41	35	96
User 15	137	137	21	97	142	93	104	53	65	12351	29	29	78	473	15951	36	36	53	53	53	15951	56	21	27	489	63	142	53	54	50
User 16	66	42839	37	63	100	89	60	41	44	42839	994	100	79	666	42839	42839	63	29	29	82	1059	49	104	64	962	63	58	26	28	42839
User 17	293	44	47	63	51	39	85	17	1599	23	23	83	33	105	153	44	105543	17	29	37	58	68	48	41	58	232	49	44	17	61
User 19	311	94	94	62	87	72	63	49	134	63	43	82	66	153	300	102	220	11273	43	48	1039876	97	44	38	6323	103	350	96	113	120
User 20	71	69	53	63	57	46	63	25	57	78597	7619	113	25	78597	703	51	28	41	78597	257	78597	38	30	41	78597	52	78	31	38	126
User 21	111	80626	12	66	27	98	44	14	47	148	114	93	13	290	328	330	15	35	308	80626	115	32	47	61	114	30	74	18	12	187
User 22	369	63	190	79	175	63	71	25	256	106	155	119	55	262	735	57	147	55	57	28	117365	152	53	53	528	201	349	34	55	72
User 24	2138	39	25	23	293	50	68	24	261	23	26	273	27	85	73	80	311	23	24	23	135	40	27	201250	21	27	284	192	149	22
User 25	87	128	16	232	41	47	44	21	31	241	47	273	27	72	111	33	281	23	35	44	40	27	38344	50	64	15	32	9	7	53
User 26	73	48469	11	211	44	144	31	14	63	53	53	48469	142	121	173	424	26	22	44	149	37	21	48469	48469	71	21	22	14	14	48469
User 27	651	137	17	44	66	82	78	58	107	32	26	269	93	606	586	265	40	60	47	44	30058	27	99	63	33068	39	60	10	18	269
User 28	505	30	354	79	169	66	49	49	152	34	21	81	27	305	594	44	309	43	32	23	70217	148	37	41	157	70217	179	53	76	29
User 29	474	34	30	62	72	52	62	23	75	52	38	66	20	68	1446	44	309	23	23	19	88053	74	49	35	61	3013	88053	31	23	35
User 31	310823	37	310823	38	38	54	153	26	310823	14	17	53	22	213	38	310823	38	27	28	27	310823	310823	37	37	73	310823	310823	310823	409	57
User 32	1073	63	111	76	131	63	73	37	86	43	21	141	37	114	101	70	61	53	31	21	128	371	64	37	603	3113	85	30	352857	38
User 34	48	73	223	123	241	77	34	18	38	84	63	396	45	73	2490	73	209	22	30	34	82	34	101	42	85	26	48	20	24	124409

Table 8.2 - Number of keypresses before a challenge (composite)

When considering the inconvenience to genuine users the results indicate that a user would only be challenged after approximately 5 hours of work – with some users only being challenged after 50 days of typing. Given that most users will authenticate at least once a day, an authentication challenge after the equivalent of 50 days of typing is unlikely to ever occur in normal operation.

8.3 Neural Network Approach

Following the analysis conducted in the previous chapter, a brief final trial was conducted using a neural network approach. There has been relatively little research conducted using neural networks to authenticate keystrokes with the two published works listed in chapter 5 (Brown & Rogers, 1993 and Furnell et al., 1996) both considering the use of neural networks on static keystroke data. The work presented in this section considers the use of neural network techniques on dynamic keystroke data (i.e. continuous monitoring).

Due to the volume of data and the complexity of the required networks, it was necessary to create a reduced subset of data upon which to base the neural network analysis. The first stage was to extract a number of common samples across all user profiles. For the purposes of this experiment, the top ten commonly occurring digraphs (across all profiles) were selected. It should be noted that while lists are available for common digraph pairs in the English language; these did not match the data logged by these experiments with the top ten digraphs showing variation across all users. Table 8.3 presents the minimum and maximum occurrence counts for each digraph pair across all user profiles (this reflects the range of sample sizes originally presented in the previous chapter).

Digraph	Min	Max
TH	319	8845
IN	214	8643
HE	212	7901
TE	352	7503
ES	135	6298
ON	271	5824
AN	135	5239
AT	193	5164
SE	167	4553
ST	92	3530

Table 8.3 - Profile counts of common digraphs

As it was necessary to have a common sample size for each input chosen for the neural network, the minimum sample count for each digraph pair determined the profile sample size (i.e. each user's data needed to share a common subset of digraphs with a common number of samples per digraph). In order to optimise the accuracy of the network; a number of user profiles were removed in order to increase the fixed sample size for the neural network (a higher number of samples is likely to increase the classification accuracy). Having removed the worst performing profiles (the five users identified in earlier experiments), the resultant digraph counts are shown in Table 8.4.

Digraph	Min	Max
TH	407	8845
IN	512	8643
HE	426	7901
TE	333	7503
ES	356	6298
ON	328	5824
AN	446	5239
AT	206	5164
SE	258	4553
ST	227	3530

Table 8.4 - Reduced profile counts of common digraphs

The reduced profile counts shown in Table 8.4 provide a maximum data set size of 200 samples per digraph. This effectively provided 2000 samples per user (200 samples x 10 digraphs). This data was randomly extracted from the original keylogger files (i.e. for each

digraph, 200 random samples were selected from the raw keylogger file and then saved as text files in preparation for the neural network analysis).

The neural network was configured as a ten input, multi-layer perceptron network with each of the ten inputs representing one of the digraphs (Figure 8.2). The data was then fed into the neural network for each user, with the resultant overall FAR/FRR chart shown in Figure 8.3.

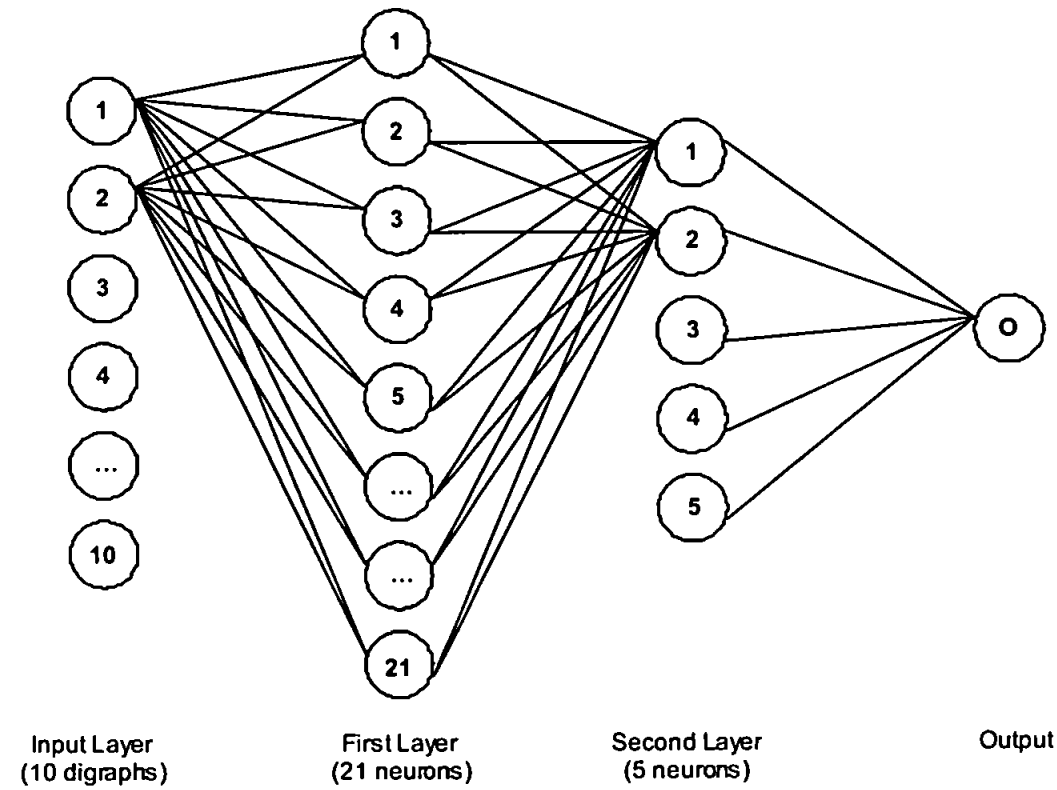


Figure 8.2 - Neural network configuration (21:5:1)

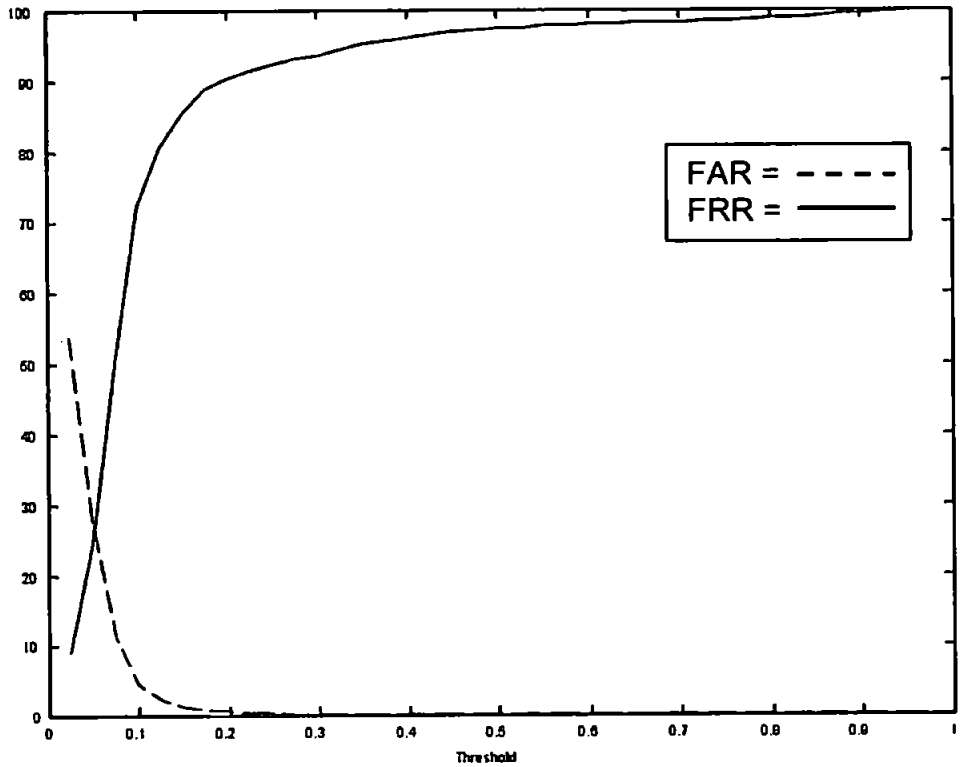


Figure 8.3 - Overall FAR/FRR rates for all users

Figure 8.3 shows the relationship between the average FAR and FRR with an Equal Error Rate (EER) of approximately 24.9%. It should be noted that the neural network approach utilised only a small proportion of the overall data set and as such is only an indication of the potential performance that this technique may offer. For several users, the common ten digraphs were not representative of the ten highest occurring digraphs for their profile – it is likely that by increasing the number of inputs to the network a higher degree of authentication could be achieved (i.e. increasing the number of digraphs on which users are classified).

The results from the neural network approach are presented here as an indication of the applicability of the approach to the problem of keystroke analysis, and it is hoped that further work would improve the classification of users and hence reduce the EER by optimising the training and size of the neural network.

8.4 A prototype demonstrator for comprehensive user authentication

In addition to the comparator illustrated in Figure 8.1 (that processed keystroke data effectively off-line), a real-time keystroke analysis demonstrator was developed. This program implements the keylogging and profile comparison features into a single application. The screen-shot shown in Figure 8.4 illustrates the dynamic alert level as a continuous moving line in white, with alerts/challenges shown in red. In addition, the digraph alert level is shown in green with corresponding lines for trigraphs (yellow) and keywords (blue). In this example, a challenge would have been issued at the mid-point of the plot where the alert level reached 75 (the threshold for a challenge was set at 70 and is indicated in the figure by a broken cyan line). The screen-shot also indicates the number of challenges in the current session together with the number of keystrokes before the first challenge was issued. It should be noted that the example chart in Figure 8.4 shows no further activity in the latter part of the plot – i.e. the horizontal lines indicate no keyboard activity. It can also be observed that there is no deviation in the keyword (blue line) alert level as this profile had relatively few profiled keywords.

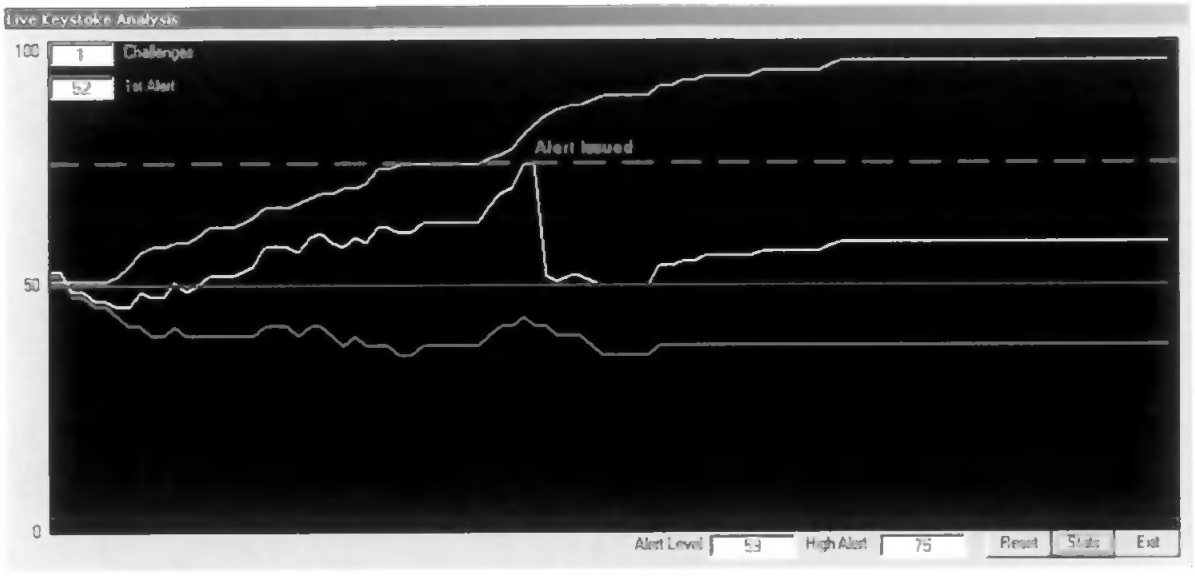


Figure 8.4 - Live keystroke analysis demonstrator – impostor (detected)

Figure 8.5 and Figure 8.6 illustrate two further typing sessions. Both figures are based upon a genuine user typing with different profile settings. Figure 8.5 was based on the default settings (i.e. not optimised for the individual user) and shows considerable difference between the digraphs (shown in green – high acceptance) and the trigraphs (shown in yellow – low acceptance) thus resulting in a moderate composite acceptance rate (shown in white). When considering Figure 8.6, the same user has a much better (and more consistent) overall acceptance rate and hence a lower alert level for digraphs and trigraphs.

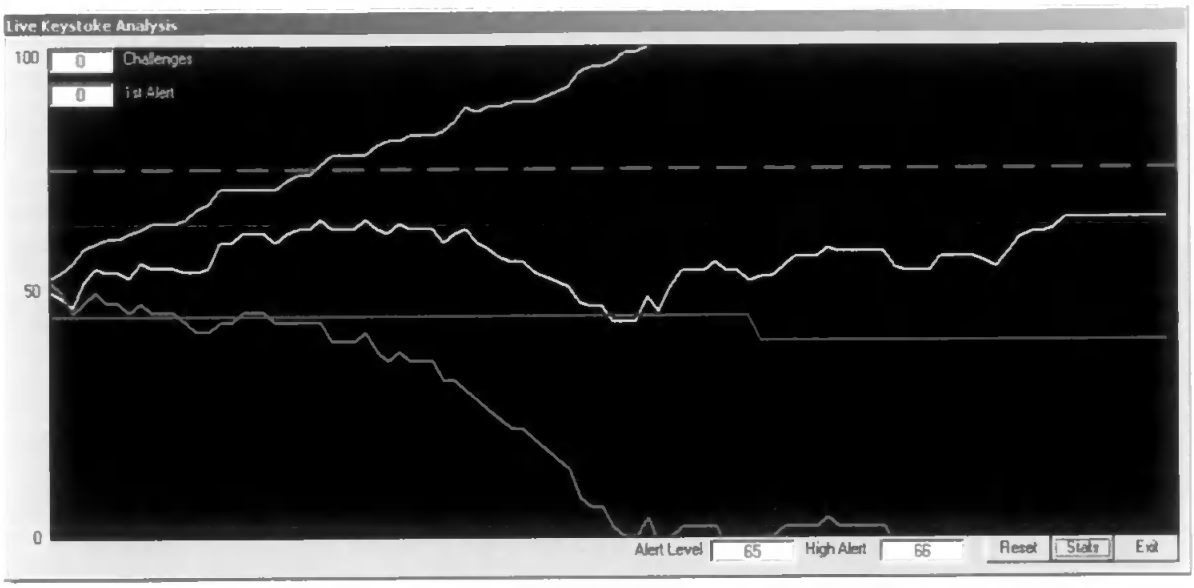


Figure 8.5 - Live keystroke analysis demonstrator – genuine user (default settings)

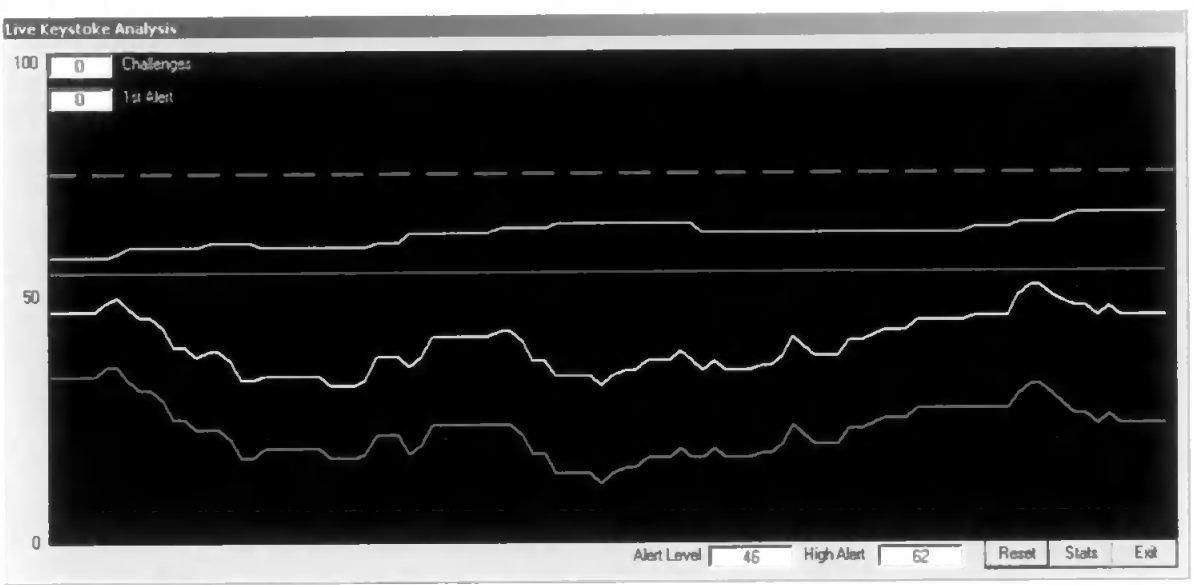


Figure 8.6 - Live keystroke analysis demonstrator – genuine user (optimised settings)

The live keystroke analysis application also has a similar options window to that of the off-line comparator shown in Figure 8.1, which allows a range of options to be selected to determine the comparisons made against the reference profile (illustrated in Figure 8.7).

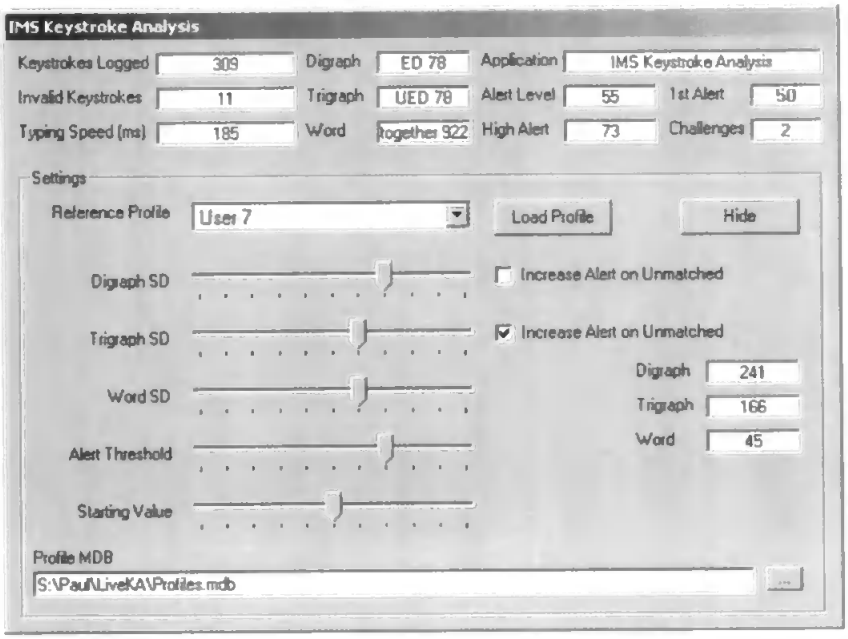


Figure 8.7 - Live keystroke analysis options

8.5 Conclusions

This chapter has further extended the novel approach to keystroke analysis. The combination of all three metrics into a single composite metric has shown significant potential, with genuine users able to work unhindered for hours (up to days) while impostors (with few exceptions) were only able to type for two minutes prior to an authentication challenge being triggered.

In addition to the results for the composite techniques, this chapter has also presented an overview of a demonstrator system capable of composite keystroke analysis coupled with a selection of challenge techniques to respond to alerts raised by the system.

While the results show keystroke analysis is a viable authentication technique, it is also clear that the analytical approach does not provide sufficient distinction for all users and a live implementation would have to consider which metric (if any) is most appropriate for each user. Over time, the profiles would also need to be updated to reflect variations in typing style and the possibility of improvement (or degradation) in a user's typing abilities. It is envisaged that keystroke analysis would become only one of a number of monitoring techniques used by a more comprehensive system together with other authentication and supervision techniques. Such a system would be able to select appropriate authentication and supervision techniques for the current user working with the resources available on the computer.

Chapter 9

Conclusions

9.1 Achievements of the research programme

The research programme has met all of the objectives outlined in the first chapter. The main achievements of this research programme were.

- Identifying user attitudes and opinions to login and continuous user authentication. This justified the selection and development of the alternative authentication techniques.

- Implementation and evaluation of novel continuous user authentication approaches based upon digraphs, trigraphs and keywords. The use of trigraph and keyword authentication represents novel work in the authentication domain.

- Design of novel methods for implementing and evaluating composite continuous user authentication. The use of composite user authentication represents entirely new work in this domain.

- Development of a proof-of-concept prototype composite authentication and response system. This demonstrated the concept of continuous non-intrusive user authentication using keystroke analysis.

The thesis began by describing current authentication techniques and proposing alternatives before evaluating user attitudes and opinions to the current authentication methods as well as their opinions on a range of alternative authentication and continuous

supervision techniques. Having identified a preference for secret-based approaches, the thesis presented the results of a series of trials conducted to evaluate the use of password, ImagePIN and cognitive/associative questions. These approaches demonstrated that while users still prefer secret-based methods, their frequent inability to use them both reliably and properly (in accordance with guidelines on correct use) indicated that alternative techniques are needed. While other methods were available, a key requirement concerning the selection of an alternative approach was that an ideal technique must be transparent to the user (i.e. it must have little/no impact on the users' normal activity). In order to achieve this objective, non-intrusive authentication mechanisms were considered. Having eliminated a number of alternatives (due to their intrusive nature or the cost of dedicated hardware), keystroke analysis was selected for a number of experimental trials.

To evaluate keystroke analysis, a series of experiments aimed at implementing and evaluating alternative user authentication techniques were conducted and are described in detail in chapters 6, 7 and 8. These experiments evaluated keystroke analysis in an unconstrained manner monitoring keystroke activity across all applications in a users' session. Users were authenticated on digraph, trigraph and keyword metrics with further trials conducted to evaluate the performance of composite metrics.

In addition to the simple analysis approach used, a number of alternatives have been considered. In chapters 6 and 7, the use of data mining and neural network approaches were evaluated and while both of these techniques were able to provide a level of user classification, there was insufficient time to fully investigate their applicability and as such they will be discussed in section 9.3 as part of the future work proposals. In addition to

these two techniques, section 9.3 also outlined the potential for the application-specific keystroke analysis method described in chapter 6.

Several papers relating to the research programme have been presented at refereed conferences and in appropriate international journals (a complete list is presented in Appendix D together with a subset of relevant papers). As such, it is believed that the research has made valid and useful contributions to the information system security field.

9.2 Limitations of the research

Although the research programme met the objectives outlined in chapter 1, there were a number of limitations to the research.

1. The survey conducted in chapter 2 assessed a limited range of participants and as such is not representative of the wider population. User attitudes towards alternative forms of user authentication have, as a consequence, been assessed by dominantly IT-literate individuals (mainly drawn from academic staff and students). Had the survey been conducted across a wider cross-section of the public, it is probable that acceptance for continuous user authentication would have been ranked less favourably due to the *big-brother* concerns of the general public. However, this does not unduly affect the consequent selection of non-intrusive continuous monitoring techniques since the most likely use for these methods is within an organisational context.

2. The alternative techniques described in chapter 4 were assessed using a limited number of users – all of whom were drawn from staff and students within the University. To give a more rounded viewpoint, it would have been ideal to have had a wider population participating in the trial. However, there were technical limitations that required a supervised installation and guidance for the users before participating.

3. The experiments conducted in chapters 6 and 7 were both limited in terms of the duration of the trial and the number of participants. The first experiment was limited to one month's duration with 10 participants to provide an initial set of results to inform the implementation of more comprehensive follow-on tests. The later experiment's duration was constrained to three months with 34 participants in order to meet deadlines for the entire research programme. Obtaining more participants was a significant problem in both experiments due to reluctance on the part of some users to be monitored and the lack of cooperation from management when approached for permission to install the software for members of administrative staff. Despite these limitations, the experiments were able to proceed, albeit with fewer participants than hoped for.

4. There were also limitations in the analysis techniques used – in particular, the data mining and neural network approaches could have been expanded further, as described in the next section.

9.3 Future work

Throughout the research programme, a number of extensions to the project have been considered. While some of these ideas have been discussed in the body of the thesis, this section presents a range of potential ideas for future work to progress the research beyond the PhD programme.

9.3.1 Data mining approach

In chapter 6, the use of data mining was considered as an alternative approach. This technique did not prove to be a reliable method – only achieving a classification accuracy of 50%. Using larger data sets and a reduced set of digraphs (in the same manner as the neural network approach discussed in chapter 8), it may be possible to improve the classification accuracy quoted. However, the benefits of this approach may not be sufficient to warrant the time needed to analyse and classify user samples. It may, however, be of use when enrolling users and developing the initial authentication profile, as the data mining approach may detect more subtle distinguishing characteristics than the current methods.

9.3.2 Neural network approach

The results presented in chapter 8 showed an overall Equal Error Rate (EER) of 26% for the neural network approach. While this error rate is far too high for a usable system, it does indicate a correlation between the genuine user and their profile; and a system able to distinguish between genuine users and impostors. Further research may be able to improve

the acceptance rates through optimisation of the training methods used coupled with changes in the neural network construction and size.

9.3.3 Optimised composite metric comparison parameters

In chapter 8, the use of composite keystroke analysis was evaluated using a range of settings in the data comparator (e.g. varying the deviation thresholds etc.). While the results from these trials were considered acceptable, further refinement could improve the FAR/FRR rates to detect impostors quickly while avoiding any unnecessary interruption to the genuine user. Due to the range of settings that could be applied (e.g. 0.5-0.8 standard deviations for the metrics, the inclusion of unmatched alert increases and the alert threshold) there are at least six variables that can be adjusted independently of each other resulting in several hundred potential combinations per profile.

In order to improve the error rates encountered in chapter 8, it would be necessary to optimise the settings used for each profile. This would require an iterative approach to home-in on a suitable group of settings.

9.3.4 Application-specific profiles

Chapter 6 proposed the use of application specific profiles in order to differentiate between a users' typing behaviour in individual application contexts. The results from this section demonstrated differences in typing styles between classes of applications, but further work is required to (a) determine if these differences remain once sufficient data is available to produce a profile, and (b) determine if certain users can be authenticated more accurately

using application specific profiles or by using the application-independent approaches already demonstrated.

The data gathered in chapter 7 has not, at this stage, been processed for application profiles, as such; this will be one of the first elements of the post-thesis research. It is likely that, given the increased size of keylogger files, there will be an improvement over the results presented in chapter 6.

9.3.5 Application-specific monitoring

Following the previous section, it may be advantageous to have specific monitoring enabled for individual applications. For example, when a user is interacting with a command interpreter (e.g. the DOS shell), they are unlikely to be typing sufficient keystrokes to authenticate the user. However, the misuse-potential for such applications is high (e.g. the user could reformat the hard disk or delete key files). As such, it may be beneficial to monitor specific applications more rigorously than others in order to detect an impostor in a shorter period of time (e.g. using keyword-based monitoring).

9.3.6 Keyword Latencies

The keystroke analysis presented in this thesis has considered the latencies of digraphs, trigraphs and keywords. However, there may also be scope for monitoring of the individual digraph (or trigraph) elements of the monitored keywords. For example, when monitoring the keyword 'format', rather than just monitoring the overall latency, the digraphs FO, OR, RM, MA and AT could be evaluated (equally the trigraphs FOR, ORM,

RMA and MAT could be used). As such, keyword profiles would require a series of timings, rather than the single timing value currently used.

9.3.7 Larger-Scale Trial

Having demonstrated the reliability of continuous supervision using keystroke analysis, it would be necessary to implement a larger trial using a greater number of trial participants over a longer period of time. It would also be necessary to ensure the participants represent a less homogenous range of users (i.e. ensuring a wider range of typing abilities and IT literacy are represented by the trial user group).

9.3.8 Statistical Analysis

In addition to extending the trial to include a wider range of participants, further work should be conducted to investigate the use of formal statistical analysis on the data generated using the techniques described in this thesis.

9.3.9 Response System

In addition to the user monitoring proposed in this thesis, it is also necessary to respond to the challenges issued by the monitoring system. The ImagePin and cognitive question methods from chapter 4 are potential candidates for this role, as they would allow a valid user to quickly authenticate and continue to work while preventing an impostor from proceeding. The issue of response is currently being investigated in another research programme (Papadaki et al., 2002)

9.3.10 Implementation issues

To implement the proposed monitoring system, it would also be necessary to consider the wider framework in which the system would be placed. It is envisaged that a monitoring system would be activated upon satisfactory authentication by the operating system. Following this, the monitoring system would automatically monitor user activity and respond with a challenge in the event that the alert level increased beyond a pre-defined threshold. To realise a system of this nature under Windows would require considerable low-level access to the operating system in order to guarantee that a user could be denied access to their computer in the event of suspicious activity. Investigating the wider implementation of a comprehensive monitoring system forms a part of a number of Intrusion Monitoring System (IMS) related projects building on the original work of Furnell (1995).

9.4 Conclusions

The role of information technology in both the work and home environments has increased beyond any imaginable boundary. As part of the increasing pervasiveness of IT, we have to accept that our computer systems will be increasingly vulnerable to attack and misuse. The simple fact that computer users are unable to apply even the simplest form of protection (the humble password) in a secure and appropriate manner demonstrates that alternative forms of securing IT systems are required.

This research has emphasised the role of security in modern computer systems and has focussed upon addressing the need for reliable and dependable user authentication and

supervision. The application of a single authentication mechanism or monitoring technique is in itself insufficient to guarantee system security. Instead a comprehensive security system is required to initially authenticate and subsequently monitor and challenge users as appropriate. Such a system would have to adapt to the differences between users, and apply appropriate authentication and supervision techniques for each one, as well as having the ability to learn from the user over time and to adapt to discrete changes in natural behaviour. In addition to the ability to authenticate and monitor, the ideal system would also have to respond to perceived impostor activity with an appropriate challenge in order to more confidently authenticate the user. All of this needs to occur as transparently as possible in order to achieve the requirement of non-intrusiveness so often required by users.

In conclusion, the research presented in this thesis has addressed a number of authentication issues. A mechanism has been developed to monitor a users' session in a non-intrusive manner by evaluating normal typing patterns. This has demonstrated that a user can be authenticated in a transparent manner without requiring any explicit action. However, the reliability with which this can be done is dependent on a number of factors and, as such, cannot be used as the sole method of authentication. Instead, the techniques described in this thesis can evaluate user characteristics that could be monitored as part of a much wider authentication and supervision framework.

Authentication will remain a crucial factor in protecting access to the applications and services we depend upon in our day-to-day lives. With greater access to information and the introduction of new services like e-Voting, e-Health etc., the need for effective and reliable authentication techniques will only increase.

References

Reference List

1. @Stake. 2004. LC4 Password Auditing and Recovery, @Stake security web site, http://www.atstake.com/products/lc/

2. 2600. 2000. 2600 Web Site – Hacked site archive, May 2000, http://www.2600.com/hacked_pages

3. ACI. 2004. Access Control International Web Site, EyeDentify retinal reader, http://www.acisecurity.com/products_eyedentify.htm

4. Anderson J.P. 1980. *Computer Security Threat Monitoring and Surveillance*, James P. Anderson Co., Fort Washington, PA (Apr.).

5. Audit Commission. 1990. *Survey of Computer Fraud and Abuse*, Audit Commission Publications, UK.

6. Audit Commission. 1994. *Opportunity Makes a Thief*, Audit Commission Publications, UK, ISBN 0-11-886137-9.

7. Audit Commission. 1998. *Ghost in the Machine – An Analysis of IT Fraud and Abuse*, Audit Commission Publications, UK, ISBN 1-86240-056-3.

8. Audit Commission. 2001. *yourbusiness@risk - An update on IT Abuse 2001*, Audit Commission Publications, UK, ISBN 1-86240-289-2

9. Bensinger D. 1998. "*Human Memory and the Graphical Password*", http://www.passlogix.com/media/pdfs/bensinger.pdf

10. Bergadano F., Gunetti D. and Picardi C. 2002. "*User authentication through keystroke dynamics*", ACM Transactions on Information and System Security (TISSEC), vol. 5, issue 4, pp367-397.

11. BioAPI. 2004. BioAPI Consortium, 2004, http://www.bioapi.com

12. BioPassword. 2004. BioPassword Web Site, http://www.biopassword.com

13. Bleha S., Slivinsky C. and Hussein B. 1990. "*Computer-access security systems using keystroke dynamics*", Actions on pattern analysis and machine intelligence, vol. 12, no. 12, pp1217-1222.

14. Blonder G. 1996. *Graphical Password*, United States Patent 5559961.

15. Brown M. and Rogers S.J. 1993. "*User identification via keystroke characteristics of typed names using neural networks*", International Journal of Man-Machine Studies, vol. 39, pp999-1014.

16. Card S.K., Moran T.P. and Newell A. 1980. "*The keystroke level model for user performance time with interactive systems*", Communications of the ACM, vol. 23, issue 7, pp396-410.

17. Cherry A., Henderson M.W., Nickless W.K., Olson R. and Rackow G. 1992. "*Pass or Fail: A New Test for Password Legitimacy*", Mathematics and Computer Science Division, Argonne National Laboratory, MCS-P328-1092, September 25th 1992.

18. Clarke N.L., Furnell S.M., Lines B. and Reynolds P.L. 2003. "*Keystroke Dynamics on a Mobile Handset: A Feasibility Study*", Information Management and Computer Security, vol. 11, no. 4, pp161-166.

19. Cope B.J.B. 1990. "*Biometric Systems of Access Control*", Electrotechnology, April/May, pp71-74.

20. CSI. 2003. *Eighth Annual CSI/FBI Computer Crime and Security Survey*, CSI, USA, 2003.

21. Deane F., Barrelle K., Henderson R., and Mahar D. 1995. "*Perceived acceptability of biometric security systems*", Computers and Security, vol. 14, no. 3: 225-231.

22. Dhamija R. and Perrig A. 2000. "*Déjà Vu: A User Study Using Images for Authentication*", SIMS/ CS, University of California Berkeley.

23. Dinnie G. 1999. "*The second annual global information security survey*", Information Management and Computer Security, Vol. 7, No. 3, pp112-120.

24. Dowland P.S., Furnell S.M., Illingworth H.M. and Reynolds P.L. 1999. "*Computer crime and abuse: A survey of public attitudes and awareness*", Computers and Security, Vol. 18, No. 8, pp715-726.

25. Dowland P.S., Singh H. and Furnell S.M. 2001. "*A Preliminary Investigation of User Authentication Using Continuous Keystroke Analysis*", Proceedings of the IFIP 8th Annual Working Conference on Information Security Management and Small Systems Security, Las Vegas, 27-28 September

26. DTI. 2002. *Information Security Breaches*, DTI Technical Report, April 2002, URN 02/318

27. F-Secure. 2004. *Mydoom worm is now the worst email worm incident in virus history*, F-Secure Web Site News Release, http://www.f-secure.com/news/items/news_2004012800.shtml

28. Furnell S.M., Morrissey J.P., Sanders P.W. and Stockel C.T. 1996. "*Applications of keystroke analysis for improved login security and continuous user authentication*", Proceedings of the 12th International Conference on Information Security (IFIP SEC '96), Island of Samos, Greece, 22-24 May, pp283-294.

29. Furnell S.M. 1995. "*Data Security in European Healthcare Information Systems*", PhD Thesis, University of Plymouth, UK.

30. Furnell S.M., Dowland P.S. and Sanders P.W. 1999. "*Dissecting the 'Hacker Manifesto'*", Information Management and Computer Security, Vol. 7, No. 2, pp69-75.

31. Gaines R., Lisowski W., Press S. and Shapiro N. 1980. *Authentication by Keystroke Timing: some preliminary results*, Rand Report R-256-NSF. Rand Corporation.

32. Guven A. and Sogukpinar I. 2003. "*Understanding users' keystroke patterns for computer access security*", Computers and Security, vol. 22, no. 8, pp 695-706

33. Haga W.J. and Zviran M. 1991. "*Question-and-Answer Passwords: An Empirical Evaluation*", Information systems, vol. 16, no. 3, pp335-343.

34. Heskett B. 1998. *A new windows password cracker*, CNet News.com, 13th February 1998, http://news.cnet.com/news/0-1003-200-326537.html

35. HMSO. 2000. *The Telecommunications (Lawful Business Practice) (Interception of Communications) Regulations 2000*, Statutory Instrument 2000 No. 2699, ISBN 0 11 099984 3

36. Home Office. 2003. *David Blunkett: National ID Card Scheme to be Introduced*, Home Office Government Press release, Reference: 307/2003 - Date: 11 Nov 2003 12:30, http://www.homeoffice.gov.uk/n_story.asp?item_id=675

37. IBG. 1999. *International Biometric Group Comparative Finger and Face Scan Study*, IBG Web Site, February 1999 http://www.biometricgroup.com/a_shared/finger_and_face.htm

38. IBG. 2002. *Zephyr Charts*, International Biometric Group, 2002, http://www.biometricgroup.com/zephyr_2002.pdf

39. ICM. 1999. *ICM Poll – The Internet – January 1999*, ICM Research. http://www.icmresearch.co.uk/reviews/1999/internet-99-jan.htm

40. Identix. 2004. *Success Stories*, Identix Incorporated web site, 2004, http://www.identix.com

41. ISBS. 2000. *Information Security Breaches Survey 2000*, Department of Trade and Industry, 2000. http://www.dti.gov.uk/cii/dtiblue/dti_site_site/

42. IT Week. 2003. *Security needs the right people*, IT Week web site, http://www.itweek.co.uk/Analysis/1140673

43. ITSEC, 1991, *Information Technology Security Evaluation Criteria, Provisional Harmonised Criteria*, Commission of European Communities.

44. Jermyn I., Mayer A., Monrose F., Reiter M.K., and Rubin A.D, 1999, "The Design and Analysis of Graphical Passwords", *Proceedings of the 8th USENIX Security Symposium*, August

45. Jobusch D.L. and Oldehoeft A.E. 1989. "*A Survey of Password Mechanisms : Part I*", Computers & Security, Vol. 8, No. 7, pp587-604.

46. Joyce R. and Gupta G. 1990. "*Identity authentication based on keystroke latencies*", Communications of the ACM, vol. 33, no. 2, pp168-176.

47. Khew P. 2002. "*Iris Recognition Technology for Improved Authentication*", Sans Security Essentials, Sans Institute, http://www.sans.org/rr/papers/6/132.pdf

48. Klein D. 1990. "*A survey of, and improvements to, password security*", Proceedings of the USENIX Second Security Workshop, Portland, Oregon, pp. 5-14, August 1990.

49. KPMG. 1998. *Information Security Survey 1998*, KPMG Information Risk Management, UK, http://www.kpmg.co.uk.

50. Legett J. and Williams G. 1988. "*Verifying user identity via keystroke characteristics*", International Journal of Man-Machine Studies, vol. 28, pp67-76.

51. Legett J., Williams G., Usnick M. and Longnecker M. 1991. "*Dynamic identity verification via keystroke characteristics*", International Journal of Man-machine Studies, vol. 35, pp859-870.

52. Mahar D., Napier R., Wagner M., Laverty W., Henderson R.D. and Hiron M. 1995. "*Optimizing digraph-latency based biometric typist verification systems: inter and intra typist differences in digraph latency distributions*", International Journal of Human-Computer Studies, vol. 43, pp579-592.

53. McCullagh D. and Arent L. 2000. *A Frenzy of Hacking Attacks*, Wired News, 9 February 2000, http://www.wired.com/news/print/0,1294,34234,00.html.

54. Microsoft. 2000. *Microsoft and I/O Software Strengthen Industry Adoption of Biometrics*, Microsoft Corporation, May 2000 http://www.microsoft.com/PressPass/press/2000/May00/BiometicsPR.asp

55. Microsoft. 2004. *QueryPerformanceCounter documentation – Platform SDK: Windows System Information*, Microsoft MSDN web site, http://msdn.microsoft.com/library/default.asp?url=/library/en-us/sysinfo/base/queryperformancecounter.asp

56. Morris R. and Thompson K. 1979. "*Password Security: A Case History*", Communications of the ACM, Vol. 22, No. 11, pp594-577, November 1979.

57. Napier R., Laverty W., Mahar D., Henderson R., Hiron M. and Wagner M. 1995. "*Keyboard user verification: towards an accurate, efficient, and ecologically valid algorithm*", International Journal of Human-Computer Studies, vol. 43, pp213-222.

58. Observer. 2002. *Who do we think we are?*, Observer newspaper, January 13th, http://www.guardian.co.uk/Archive/Article/0,4273,4334189,00.html

59. Ofcom (formerly Oftel). 2003. *Consumer's use of Internet, Oftel residential survey*, Q14 August 2003, http://www.ofcom.org.uk/static/archive/oftel/publications/research/2003/q14intres1003.pdf

60. Ord T. and Furnell S.M. 2000. "*User authentication for keypad-based devices using keystroke analysis*", Proceedings of the Second International Network Conference (INC 2000), Plymouth, UK, 3-6 July, pp263-272.

61. Panasonic. 2004. *Authenticam Iris Recognition Camera*, Panasonic web site, http://www.panasonic.com/cctv/products/bmet100us.asp

62. Papadaki M., Furnell S.M., Lee S.J., Lines B.L. and Reynolds P.L., 2002, "Enhancing Response in Intrusion Detection Systems", *Journal of Information Warfare*, vol. 2, no. 1, pp90-102

63. Recogsys. 2003. *IR Recognition Systems Biometric Hand Readers Increase Security at University of Central Florida Sororities*, Recogsys Web Site, 10th December 2003 http://www.recogsys.com/news/pressreleases/2003/031210.htm

64. Sasse M.A., Brostoff S. and Weirich D. 2001. "*Transforming the 'weakest link'- a human/ computer interaction approach to usable and effective security*", BT Technology Journal, vol. 19, no 3, pp122-131.

65. Schultz E. 2003a. "*Massive Distributed Denial of Service Attack*", Security Views, Computers and Security, vol. 22, no. 6, p465

66. Schultz E. 2003b. "*Researchers develop powerful Windows password cracking method*", Security Views, Computers and Security, vol. 22, no. 6, pp471-472

67. Secure Computing. 1995. "*Body Check*", Secure Computing, July 1995.

68. Sherman R.L. 1992. "*Biometrics Futures*", Computers and Security, Vol. 11, No. 2, pp128-133.

69. Singh H., Burn Thornton K.E. and Bull P.D. 1999. "*Classification of Network State Using Data Mining*", Proceedings of 4th IEEE international MICC and ISCE conference-1999, Vol. 1, pp183-187

70. Singh H., Furnell S.M., Lines B. and Dowland P.S. 2001. "*Investigating and Evaluating Behavioural Profiling and Intrusion Detection Using Data Mining*", Proceedings of International Workshop on Mathematical Methods, Models and Architectures for Computer Networks Security, St. Petersburg, Russia, 21-23 May

71. Song D., Venable, P. and Perrig A. 1997. *User Recognition by Keystroke Latency Pattern Analysis*, Project report, http://www.ece.cmu.edu/~adrian/projects/keystroke

72. Spafford E.H. 1992. "*Opus: Preventing Weak Password Choices*", *Computers and Security*, Vol. 11, No. 3, pp273-278, May 1992.

73. Umphress D. and Williams G. 1985. "*Identity verification through keyboard characteristics*", International Journal of Man-Machine Studies, vol. 23, pp263-273.

74. USDoS. 2003. *U.S. Visitor and Immigrant Status Indicator Technology Program*, United States Department of State, 28th October, http://usinfo.state.gov/topical/pol/terror/texts/03102801.htm

75. Wood H. 1977. "*The use of passwords for controlled access to computer resources*", NBS Special Publications, U.S. Dept. of Commerce/NBS, pp. 500-509, May 1977.

76. Woodward J.D., Horn C., Gatune J. and Thomas A. 2003. "*Biometrics - A Look at Facial Recognition*", RAND Public Safety and Justice briefing, ISBN: 0-8330-3302-6

77. ZDNet. 2002. *Frequent fliers: The biometric guinea pigs*, ZDNet web site, March 27th, http://news.zdnet.co.uk/business/0,39020645,2107450,00.htm

78. ZoneH. 2003. *What happened yesterday?*, ZoneH web site News,

http://www.zone-h.com/en/news/read/id=3024/

Appendix A

Survey Form and Results

Additional Survey Material

This appendix contains additional material to support chapter 3. The first part of this appendix includes a printed copy of the questionnaire form that was distributed to potential respondents, while the second part presents the collated results from the relevant sections of the survey. In addition to the raw results, a number of summary tables are presented, together with additional charts not included in the body of the thesis.

Computer Crime And Abuse Survey

A Survey of Attitudes and Awareness

Paul Dowland

I am currently undertaking a Ph.D. in the field of network security. This survey aims to establish people's awareness of computer crime and abuse and their attitudes towards it. In addition, attitudes towards new security techniques will be assessed.

The first section of the survey asks about your general details, including use of computers at home and at work. The second section looks at the area of computer crime and abuse, your awareness of it and your opinions on how it is treated by the media. The final section assesses your opinions about the acceptability of various security techniques and their benefits to different business sectors.

Most of the questions require tick-box responses although a few require a short written answer.

Please feel free to add any additional comments wherever you feel it is necessary.

The survey is anonymous and the information will be treated confidentially.

The survey will take approximately 20 minutes to complete.
Please ensure you have enough time to complete the survey.

An on-line version of this survey can be found at

http://jack.see.plym.ac.uk/survey

Please feel free to distribute this URL.

Thank you for you co-operation.

Contact address

Network Research Group
School of Electronic, Communication and Electrical Engineering
University of Plymouth
Drake Circus
Plymouth
PL4 8AA

Section One - General Details

1. Sex
   Male ☐ Female ☐

2. Age Group
   16 -24 ☐ 25 -34 ☐ 35 -49 ☐ 50 -64 ☐ 65+ ☐

3. Which of the following categories best describes your current (or intended) occupation?
   - ☐ Government
   - ☐ Defence
   - ☐ Education
   - ☐ Health
   - ☐ Social and Public Services
   - ☐ Banking, Finance and Insurance
   - ☐ Utility Services (Water, Gas, Electricity etc.)
   - ☐ Manufacturing
   - ☐ Engineering
   - ☐ Travel Industries and Tourism
   - ☐ Property Services
   - ☐ Communications Industry
   - ☐ Computing Industry

   Other, please specify_____

4. Educational qualifications
   None ☐ GCSEs ☐ O-levels ☐ A-levels ☐ Degree ☐
   Other, please specify_____

5. Do you read a newspaper on a regular basis? If yes, is it a broadsheet or tabloid newspaper?
   Broadsheet ☐ Tabloid ☐

6. Have you ever used a computer?
   Yes ☐ No ☐ *If no, please go to Section 2.*

7. For how many years have you used a computer?
   Less than 1 year ☐ 1 - 5 years ☐ 6 - 10 years ☐
   11 - 15 years ☐ 16 - 20 years ☐ 20+ years ☐

Use of computer at work

8. Do you use a computer at work?
   Yes ☐ No ☐ *If no, please go to Question 15.*

9. Approximately how many hours per day do you use it?
   Less than 1 hour ☐ 1 - 4 hours ☐
   4 - 8 hours ☐ More than 8 hours ☐

10. Do you have access to Internet e-mail and the World Wide Web at work?
    Yes ☐ No ☐

11. What applications do you use it for? Please indicate how often you use each application.

	Often	Occasionally	Never
Word-processing	☐	☐	☐
Spreadsheet	☐	☐	☐
Database	☐	☐	☐
Desktop Publishing	☐	☐	☐
Software Development	☐	☐	☐
Graphical Design	☐	☐	☐
Engineering/Scientific	☐	☐	☐
Web-browsing	☐	☐	☐
E-mail	☐	☐	☐
Games	☐	☐	☐
Computer Aided Learning	☐	☐	☐

	Yes	No	Don't know
12. Have you ever used unlicensed software at work?	☐	☐	☐
13. Have you ever used the office computer for activities that are not related to your work ?	☐	☐	
14. Have you ever used someone else's password or account without their consent or knowledge?	☐	☐	

Use of computer at home

15. Do you have a computer at home?
Yes ☐ No ☐ ***If no, please go to Section 2.***

16. Approximately how many hours per week do you use it?
Never ☐ 1 - 5 hours ☐
6 - 15 hours ☐ 15+ hours ☐

17. Do you have access to the Internet at home?
Yes ☐ No ☐

18. What applications do you use your home computer for? Please indicate how often you use each application.

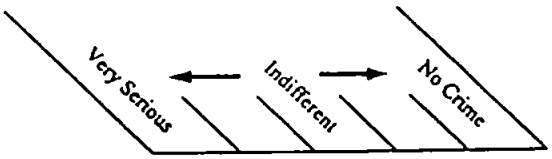
	Often	Occasionally	Never
Word-processing	☐	☐	☐
Spreadsheet	☐	☐	☐
Database	☐	☐	☐
Desktop Publishing	☐	☐	☐
Software Development	☐	☐	☐
Graphical Design	☐	☐	☐
Engineering/Scientific	☐	☐	☐
Web-browsing	☐	☐	☐
E-mail	☐	☐	☐
Games	☐	☐	☐
Computer Aided Learning	☐	☐	☐

19. Have you ever used unlicenced software at home?
Yes ☐ No ☐

Section Two - Computer Crime and Abuse

Computer systems and networks are increasingly becoming a part of everyday life. Many institutions, including banks and hospitals, use them to hold personal records. Many businesses would no longer be able to operate without them. As their use increases so the move towards a new society, an information society, comes a step closer. It is argued that this information society will be, and is, subject to many of the problems that exist in society. This section of the survey is dedicated to the problem of computer crime and abuse and people's awareness of the problem, as well as the part that the media plays in portraying such news.

20. Please rate the following acts of computer crime and abuse. Rate on a scale from very serious to no crime at all.



Viruses.	☐	☐	☐	☐	☐
Viewing someone else's data.	☐	☐	☐	☐	☐
Altering someone else's data.	☐	☐	☐	☐	☐
Theft of computer equipment.	☐	☐	☐	☐	☐
Copying software.	☐	☐	☐	☐	☐
Copying data.	☐	☐	☐	☐	☐
Computer fraud.	☐	☐	☐	☐	☐
Sabotage.	☐	☐	☐	☐	☐

21. Do you consider computer crime and abuse to be:-

	Agree	Disagree	Neutral
Not a problem?	☐	☐	☐
A technical problem?	☐	☐	☐
An ethical problem?	☐	☐	☐

22. Are you aware of the following UK legislation: -

	Yes	No
Data Protection Act?	☐	☐
Computer Misuse Act?	☐	☐

23. Please describe in a few words your image of a hacker.

_____

_____

24. Do you consider 'hacking' to be:-

	Yes	No	Don't know
Acceptable?	☐	☐	☐
An invasion of privacy?	☐	☐	☐
Theft?	☐	☐	☐

25. Do you think that people hack:-

	Yes	No	Don't know
Out of curiosity?	☐	☐	☐
To make money?	☐	☐	☐
For the thrill of it?	☐	☐	☐
To "beat the system"?	☐	☐	☐
For malicious reasons?	☐	☐	☐

Please give any other reasons. _____

	Yes	No	Don't know
26. Regarding confessed or convicted hackers:-			
Should they be allowed to work in the computing field?	☐	☐	☐
Should they be allowed to have a computer at home?	☐	☐	☐
27. Do you think that acting via a computer makes hackers feel less responsible for their actions?	☐	☐	☐
28. Do you think that a user would act differently if their activity was being monitored and recorded?	☐	☐	☐
29. Is it acceptable to read information on someone else's computer, without their permisssion, as long as you do not alter or damage it?	☐	☐	☐
30. Have you read about or heard of any computer crime cases in the news?	☐	☐	☐

31. Are you aware of any of the following:-

	Yes	No
'Chaos Computer Club'?	☐	☐
'Kevin Mitnick'?	☐	☐
'Michelangelo' virus?	☐	☐
'Zebedee virus'?	☐	☐
'Friday the 13th' virus?	☐	☐
'War Games'?	☐	☐

32. Other than the above, do you remember any other computer crime and abuse cases or keywords that you have read or heard about?

_____

_____

33. In your opinion, how does the media treat the issue of computer crime and abuse?

_____

_____

34. In your opinion, how do films treat the issue of computer crime and abuse?

_____

_____

Section Three - User Authentication and Supervision

As computers become more widely used, so the need for stronger security techniques becomes more important. Current research and evolving technologies are making new security techniques available. These techniques ensure that users are who they claim to be - this is known as *authentication* of identity. Some can also continuously monitor the authenticity of the user and the legitimacy of their activity - this is known as *supervision*. Some of these techniques are able to verify a user's identity using their physical characteristics whilst others analyse their behaviour. Physical characteristics that can be recognised include a user's face, eye (iris scanning), hand and fingerprint. Behavioural characteristics include a user's typing style (keystroke analysis), their use of the mouse (mouse dynamics), the way they sign their name (signature analysis) or their voice characteristics (voice verification).

35. Are any of the following security monitoring methods currently used in your workplace?

	Yes	No	Unsure
Password	☐	☐	☐
Restricted number of password attempts	☐	☐	☐
Audit trails	☐	☐	☐
Virus scanners	☐	☐	☐
Access restrictions	☐	☐	☐

Other, please specify _____

If passwords are used, please answer questions 36 - 43 inclusive.

36. How many different systems or applications do you use that require a password? Please indicate how many.

_____

37. Do you use a different password for each one?
Yes ☐ No ☐

38. How often do you change your password?
- ☐ Weekly
- ☐ Fortnightly
- ☐ Monthly
- ☐ Six-monthly
- ☐ Less frequently
- ☐ Never

	Yes	No	Unsure
39. Are you forced to change your password after a certain length of time?	☐	☐	☐
40. Is a minimum password length enforced?	☐	☐	☐
41. Do you write your password(s) down?	☐	☐	☐
42. Does anyone else know your password(s)?	☐	☐	☐
43. Do you think that your password could be easily guessed e.g. is it part of your address, name, partner's name?	☐	☐	☐

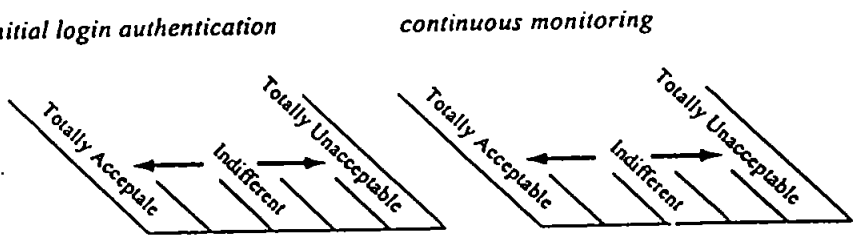
~~If virus scanners are used, please answer questions 44 and 45.~~

44. Is the virus scanner run at login?
    Yes ☐ No ☐ Unsure ☐

45. How often is the virus scanner run?
    Daily ☐ Weekly ☐ Monthly ☐ Less frequently ☐ Unsure ☐

Security techniques

46. How acceptable to you would the following techniques be when used for:-
    a) initial login authentication (where user identification is verified when they begin using the system)?
    b) continuous monitoring (where user identification is verified periodically whilst the user is logged in)?



Password *The most common method in current systems.*	☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐
Keystroke analysis *Research has shown that users have different typing styles and that they can be identified by measuring the times between keystrokes.*	☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐
Face recognition *A snapshot of the user, taken by a camera positioned on the monitor, is compared with a previously stored 'faceprint'.*	☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐
Mouse dynamics *Similar to keystroke analysis, users can be identified by the way in which they use the mouse.*	☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐
Voice verification *A user's voice, when speaking a word or phrase into the computer's microphone, is compared with a previously stored 'voiceprint'.*	☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐
Signature analysis *A user signs their name using a special pen and pad, the signature is digitised and compared with a previously stored version.*	☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐
Iris scanning *A snapshot of the user's iris, taken by a camera, is compared with a previously stored image.*	☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐
Hand geometry *This technique measures the physical dimensions of the hand using a small camera and compares these with previously stored values.*	☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐
Fingerprint analysis *An automated version of the fingerprint identification system similar to that traditionally used in criminology.*	☐ ☐ ☐ ☐ ☐	☐ ☐ ☐ ☐ ☐

	Yes	No	Unsure
47. Do you consider continuous monitoring, in general, to be acceptable?	☐	☐	☐

48. Most of these techniques require a user-profile to be set up. This is a stored record of the user's characteristics which gives the computer system a reference for comparisons.
How long would you be prepared to spend in order to set up your user-profile?

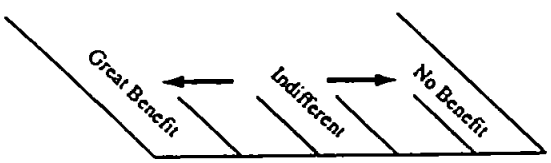
- No time ☐
- Up to 5 minutes ☐
- Up to 15 minutes ☐
- Up to 30 minutes ☐
- Up to 1 hour ☐
- More than 1 hour ☐

49. Such approaches are not 100% accurate - how frequently would you be prepared for the system to reject you and then demand that you login once again?

- Hourly ☐
- Daily ☐
- Weekly ☐
- Never ☐

	Yes	No	Unsure
50. Would you regard supervision of your activities by the computer as an invasion of your personal privacy?	☐	☐	☐
51. Would you trust your organisation to only use the monitoring for security purposes?	☐	☐	☐
52. Should users be aware that they are being monitored?	☐	☐	☐

53. Which fields/sectors do you consider would benefit from supervision of users by the computer? Please rate from a great benefit to no benefit at all.



Government	☐	☐	☐	☐	☐
Defence	☐	☐	☐	☐	☐
Education	☐	☐	☐	☐	☐
Health	☐	☐	☐	☐	☐
Social and Public Services	☐	☐	☐	☐	☐
Banking, Finance and Insurance	☐	☐	☐	☐	☐
Utility Services (Water, Gas, Electricity etc.)	☐	☐	☐	☐	☐
Manufacturing	☐	☐	☐	☐	☐
Engineering	☐	☐	☐	☐	☐
Travel Industries and Tourism	☐	☐	☐	☐	☐
Property Services	☐	☐	☐	☐	☐
Communications Industry	☐	☐	☐	☐	☐
Computing Industry	☐	☐	☐	☐	☐

Please check that you have answered all the questions.

Thank you for your cooperation.

Collated survey results

Survey Method
Online
Paper

Count of Method	
Method	Total
Online	27
Paper	148
(blank)	
Grand Total	175

Q1 Sex
Male
Female

Count of Q1	
Q1	Total
0	140
1	35
(blank)	
Grand Total	175

Q2 Age group
16-24
25-34
35-49
50-64

Count of Q2	
	Total
0	73
1	56
2	34
3	12
(blank)	
Grand Total	175

Q3 Occupation
Banking, Finance & Insurance
Computing Industry
Communications Industry
Defence
Education
Engineering
Government
Health
Manufacturing
Social and Public Services
Travel Industries and Tourism
Utility Services (Water, Gas, Electricity etc.)
Other

Count of Q3	
Q3	Total
0	1
1	25
2	37
3	3
4	27
5	41
6	3
7	6
8	1
10	4
11	5
12	4
13	18
(blank)	
Grand Total	175

Q4 Qualifications

None
GCSE
O-Level
A-Level
Degree
Other

Count of Q4	
Q4	Total
0	6
1	10
2	16
3	46
4	77
5	20
(blank)	
Grand Total	175

Q5 Newspapers

Broadsheet
Tabloid
None

Count of Q5	
Q5	Total
0	65
1	48
2	62
(blank)	
Grand Total	175

Q6 Used a computer

Yes
No

Count of Q6	
Q6	Total
0	170
1	5
(blank)	
Grand Total	175

Q7 Years of computer use

<1 year
1-5 years
6-10 years
11-15 years
16-20 years
20+years

Count of Q7	
Q7	Total
0	3
1	33
2	53
3	43
4	31
5	7
(blank)	
Grand Total	170

Q8 Use computer at work

Yes
No

Count of Q8	
Q8	Total
0	151
1	20
(blank)	
Grand Total	171

Q9 Hours of use

<1 hour
1-4 hours
4-8 hours
>8 hours

Count of Q9	
Q9	Total
0	10
1	65
2	63
3	14
(blank)	
Grand Total	152

Q10 WWW/Email access
Yes
No

Count of Q10	
Q10	Total
0	129
1	18
(blank)	
Grand Total	147

Q11a Word Processing
Often
Occasionally
Never

Count of Q11a	
Q11a	Total
0	115
1	28
2	6
(blank)	
Grand Total	149

Q11b Spreadsheet
Often
Occasionally
Never

Count of Q11b	
Q11b	Total
0	50
1	67
2	25
(blank)	
Grand Total	142

Q11c Database
Often
Occasionally
Never

Count of Q11c	
Q11c	Total
0	29
1	72
2	37
(blank)	
Grand Total	138

Q11d Desktop Publishing
Often
Occasionally
Never

Count of Q11d	
Q11d	Total
0	17
1	49
2	71
(blank)	
Grand Total	137

Q11e Software Development
Often
Occasionally
Never

Count of Q11e	
Q11e	Total
0	57
1	37
2	44
(blank)	
Grand Total	138

Q11f Graphical Design
Often
Occasionally
Never

Count of Q11f	
Q11f	Total
0	18
1	49
2	67
(blank)	
Grand Total	134

Q11g Engineering/Scientific
Often
Occasionally
Never

Count of Q11g	
Q11g	Total
0	44
1	47
2	45
(blank)	
Grand Total	136

Q11h Web-browsing
Often
Occasionally
Never

Count of Q11h	
Q11h	Total
0	88
1	39
2	18
(blank)	
Grand Total	145

Q11i E-Mail
Often
Occasionally
Never

Count of Q11i	
Q11i	Total
0	113
1	20
2	14
(blank)	
Grand Total	147

Q11j Games
Often
Occasionally
Never

Count of Q11j	
Q11j	Total
0	21
1	64
2	58
(blank)	
Grand Total	143

Q11k Computer aided learning
Often
Occasionally
Never

Count of Q11k	
Q11k	Total
0	5
1	63
2	68
(blank)	
Grand Total	136

Q12 Use of unlicensed software
Yes
No
Don't Know

Count of Q12	
Q12	Total
0	53
1	66
2	35
(blank)	
Grand Total	154

Q13 Use of office computer for non-work activities
Yes
No

Count of Q13	
Q13	Total
0	116
1	38
(blank)	
Grand Total	154

Q14 Use of other's password/account
Yes
No

Count of Q14	
Q14	Total
0	31
1	122
(blank)	
Grand Total	153

Q15 Computer at home
Yes
No

Count of Q15	
Q15	Total
0	144
1	27
(blank)	
Grand Total	171

Q16 Hours of use
Never
1 to 4
6 to 15
15 +

Count of Q16	
Q16	Total
0	3
1	60
2	47
3	33
(blank)	
Grand Total	143

Q17 WWW/Email access
Yes
No

Count of Q17	
Q17	Total
0	69
1	74
(blank)	
Grand Total	143

Q18a Word Processing
Often
Occasionally
Never

Count of Q18a	
Q18a	Total
0	112
1	27
2	2
(blank)	
Grand Total	141

Q18b Spreadsheet
Often
Occasionally
Never

Count of Q18b	
Q18b	Total
0	42
1	67
2	25
(blank)	
Grand Total	134

Q18c Database
Often
Occasionally
Never

Count of Q18c	
Q18c	Total
0	17
1	56
2	52
(blank)	
Grand Total	125

Q18d Desktop Publishing
Often
Occasionally
Never

Count of Q18d	
Q18d	Total
0	22
1	47
2	57
(blank)	
Grand Total	126

Q18e Software Development
Often
Occasionally
Never

Count of Q18e	
Q18e	Total
0	37
1	52
2	39
(blank)	
Grand Total	128

Q18f Graphical Design
Often
Occasionally
Never

Count of Q18f	
Q18f	Total
0	26
1	46
2	57
(blank)	
Grand Total	129

Q18g Engineering/Scientific
Often
Occasionally
Never

Count of Q18g	
Q18g	Total
0	35
1	48
2	45
(blank)	
Grand Total	128

Q18h Web-browsing
Often
Occasionally
Never

Count of Q18h	
Q18h	Total
0	49
1	20
2	60
(blank)	
Grand Total	129

Q18i E-Mail
Often
Occasionally
Never

Count of Q18i	
Q18i	Total
0	51
1	21
2	58
(blank)	
Grand Total	130

Q18j Games
Often
Occasionally
Never

Count of Q18j	
Q18j	Total
0	56
1	54
2	27
(blank)	
Grand Total	137

Q18k Computer aided learning
Often
Occasionally
Never

Count of Q18k	
Q18k	Total
0	6
1	40
2	81
(blank)	
Grand Total	127

Q19 Use of unlicensed software
Yes
No

Count of Q19	
Q19	Total
0	106
1	37
(blank)	
Grand Total	143

Questions 20 to 34 considered issues relating to computer crime and abuse. These results have not been included here as they are not relevant to the PhD research. The full results can be found on the CD at the rear of the thesis.

Use of monitoring methods

Q35a Password

Yes
No
Unsure

Count of Q35a	
Q35a	Total
0	160
1	9
2	2
(blank)	
Grand Total	171

Q35b Restricted password attempts

Yes
No
Unsure

Count of Q35b	
Q35b	Total
0	91
1	44
2	32
(blank)	
Grand Total	167

Q35c Audit trails

Yes
No
Unsure

Count of Q35c	
Q35c	Total
0	47
1	49
2	66
(blank)	
Grand Total	162

Q35d Virus scanners

Yes
No
Unsure

Count of Q35d	
Q35d	Total
0	145
1	12
2	11
(blank)	
Grand Total	168

Q35e Access restrictions

Yes
No
Unsure

Count of Q35e	
Q35e	Total
0	139
1	13
2	13
(blank)	
Grand Total	165

Q36 Number of systems using passwords

Count of Q36	
Q36	Total
1	31
2	32
3	24
4	16
5	13
6	6
7	4
8	2
9	1
10	11
12	2
20	3
25	1
100	1
150	1
(blank)	
Grand Total	148

Q37 Different password for systems

Yes

No

Count of Q37	
Q37	Total
0	81
1	70
(blank)	
Grand Total	151

Q38 Frequency of password changes

Weekly

Fortnightly

Monthly

Six-Monthly

Less Frequently

Never

Count of Q38	
Q38	Total
0	3
1	2
2	40
3	28
4	32
5	53
(blank)	
Grand Total	158

Q39 Forced to change password

Yes

No

Unsure

Count of Q39	
Q39	Total
0	56
1	89
2	11
(blank)	
Grand Total	156

Q40 Minimum password length
Yes
No
Unsure

Count of Q40	
Q40	Total
0	111
1	26
2	21
(blank)	
Grand Total	158

Q41 Write down passwords
Yes
No

Count of Q41	
Q41	Total
0	27
1	131
(blank)	
Grand Total	158

Q42 Anyone else know password
Yes
No
Unsure

Count of Q42	
Q42	Total
0	51
1	94
2	13
(blank)	
Grand Total	158

Q43 Easily guessed password
Yes
No
Unsure

Count of Q43	
Q43	Total
0	24
1	125
2	9
(blank)	
Grand Total	158

Q44 Virus scanner run at login
Yes
No
Unsure

Count of Q44	
Q44	Total
0	106
1	31
2	17
(blank)	
Grand Total	154

Q45 How often is virus scanner run
Daily
Weekly
Monthly
Less Frequently
Unsure

Count of Q45	
Q45	Total
0	97
1	16
2	6
3	7
4	28
(blank)	
Grand Total	154

Question 46 considered the acceptability of a range of authentication techniques, the left pivot tables present the rating of the approach specified for initial login authentication while the right pivot tables present the rating for continuous monitoring.

Q46a Password
0=Totally Acceptable
1=Acceptable
2=Indifferent
3=Unacceptable
4=Totally Unacceptable

Count of Q46aa	
Q46aa	Total
0	148
1	10
2	5
4	1
(blank)	
Grand Total	164

Count of Q46ab	
Q46ab	Total
0	40
1	13
2	35
3	30
4	39
(blank)	
Grand Total	157

Q46b Keystroke Analysis

Count of Q46ba	
Q46ba	Total
0	61
1	21
2	45
3	15
4	19
(blank)	
Grand Total	161

Count of Q46bb	
Q46bb	Total
0	43
1	35
2	41
3	10
4	28
(blank)	
Grand Total	157

Q46c Face Recognition

Count of Q46ca	
Q46ca	Total
0	76
1	30
2	28
3	15
4	12
(blank)	
Grand Total	161

Count of Q46cb	
Q46cb	Total
0	39
1	22
2	39
3	25
4	31
(blank)	
Grand Total	156

Q46d Mouse Dynamics

Count of Q46da	
Q46da	Total
0	59
1	16
2	44
3	16
4	25
(blank)	
Grand Total	160

Count of Q46db	
Q46db	Total
0	46
1	29
2	40
3	13
4	28
(blank)	
Grand Total	156

Q46e Voice Verification

Count of Q46ea	
Q46ea	Total
0	83
1	26
2	29
3	11
4	12
(blank)	
Grand Total	161

Count of Q46eb	
Q46eb	Total
0	32
1	27
2	36
3	25
4	35
(blank)	
Grand Total	155

Q46f Signature Analysis

Count of Q46fa	
Q46fa	Total
0	72
1	25
2	33
3	14
4	18
(blank)	
Grand Total	162

Count of Q46fb	
Q46fb	Total
0	24
1	10
2	32
3	33
4	57
(blank)	
Grand Total	156

Q46g Iris Scanning

Count of Q46ga	
Q46ga	Total
0	78
1	28
2	25
3	10
4	20
(blank)	
Grand Total	161

Count of Q46gb	
Q46gb	Total
0	34
1	16
2	29
3	33
4	43
(blank)	
Grand Total	155

Q46h Hand Geometry

Count of Q46ha	
Q46ha	Total
0	74
1	28
2	30
3	12
4	18
(blank)	
Grand Total	162

Count of Q46hb	
Q46hb	Total
0	29
1	18
2	31
3	39
4	39
(blank)	
Grand Total	156

Q46i Fingerprint Analysis

Count of Q46ia	
Q46ia	Total
0	85
1	24
2	23
3	10
4	20
(blank)	
Grand Total	162

Count of Q46ib	
Q46ib	Total
0	36
1	15
2	29
3	35
4	41
(blank)	
Grand Total	156

Q47 Monitoring acceptable
Yes
No
Unsure

Count of Q47	
Q47	Total
0	75
1	49
2	40
(blank)	
Grand Total	164

Q48 User-profile set-up time
No time
Upto 5 mins
Upto 15 mins
Upto 30 mins
Upto 1hr
> 1hr

Count of Q48	
Q48	Total
0	18
1	59
2	39
3	22
4	19
5	8
(blank)	
Grand Total	165

Q49 Frequency of system rejection
Hourly
Daily
Weekly
Never

Count of Q49	
Q49	Total
0	12
1	45
2	60
3	48
(blank)	
Grand Total	165

Q50 Invasion of privacy
Yes
No
Unsure

Count of Q50	
Q50	Total
0	70
1	73
2	25
(blank)	
Grand Total	168

Q51 Monitoring for security only
Yes
No
Unsure

Count of Q51	
Q51	Total
0	50
1	79
2	39
(blank)	
Grand Total	168

Q52 Awareness of monitoring
Yes
No
Unsure

Count of Q52	
Q52	Total
0	149
1	12
2	7
(blank)	
Grand Total	168

The final question (53) presented a list of sectors and asked for comments on the benefit to each sector.

Q53a Government

Great benefit

Indifferent

No benefit

Count of Q53a	
Q53a	Total
0	128
1	27
2	10
3	1
4	3
(blank)	
Grand Total	169

Q53b Defence

Great benefit

Indifferent

No benefit

Count of Q53b	
Q53b	Total
0	152
1	10
2	4
3	1
4	2
(blank)	
Grand Total	169

Q53c Education

Great benefit

Indifferent

No benefit

Count of Q53c	
Q53c	Total
0	40
1	42
2	62
3	11
4	13
(blank)	
Grand Total	168

Q53d Health

Great benefit

Indifferent

No benefit

Count of Q53d	
Q53d	Total
0	85
1	46
2	27
3	5
4	6
(blank)	
Grand Total	169

Q53e Social & Public Services
Great benefit

Indifferent

No benefit

Count of Q53e	
Q53e	Total
0	68
1	47
2	33
3	11
4	10
(blank)	
Grand Total	169

Q53f Banking, Finance & Insurance
Great benefit

Indifferent
No benefit

Count of Q53f	
Q53f	Total
0	129
1	27
2	9
4	3
(blank)	
Grand Total	168

Q53g Utility Services
Great benefit

Indifferent

No benefit

Count of Q53g	
Q53g	Total
0	45
1	38
2	63
3	11
4	11
(blank)	
Grand Total	168

Q53h Manufacturing
Great benefit

Indifferent

No benefit

Count of Q53h	
Q53h	Total
0	26
1	41
2	76
3	18
4	8
(blank)	
Grand Total	169

Q53i Engineering
Great benefit

Indifferent

No benefit

Count of Q53i	
Q53i	Total
0	34
1	43
2	63
3	19
4	10
(blank)	
Grand Total	169

Q53j Travel Industries and Tourism

Great benefit

Indifferent

No benefit

Count of Q53j	
Q53j	Total
0	17
1	31
2	80
3	25
4	16
(blank)	
Grand Total	169

Q53k Property Services

Great benefit

Indifferent

No benefit

Count of Q53k	
Q53k	Total
0	27
1	33
2	79
3	15
4	15
(blank)	
Grand Total	169

Q53l Communications Industry

Great benefit

Indifferent

No benefit

Count of Q53l	
Q53l	Total
0	67
1	43
2	38
3	12
4	9
(blank)	
Grand Total	169

Q53m Computing Industry

Great benefit

Indifferent

No benefit

Count of Q53m	
Q53m	Total
0	83
1	39
2	28
3	11
4	9
(blank)	
Grand Total	170

Summarised Data

Question 1/2

	Male		Female		Total
16-24	64	37%	8	5%	72
25-34	42	24%	14	8%	56
35-49	26	15%	8	5%	34
50-64	8	5%	4	2%	12
Total	140	80%	34	20%	174

Question 11

	Often	Occasionally	Never
Word Processing	115	28	6
Spreadsheet	50	67	25
Database	29	72	37
Desktop Publishing	17	49	71
Software Development	57	37	44
Graphical Design	18	49	67
Engineering/Scientific	44	47	45
Web-browsing	88	39	18
E-Mail	113	20	14
Games	21	64	58
Computer aided learning	5	63	68

Question 18

	Often	Occasionally	Never
Word Processing	112	27	2
Spreadsheet	42	67	25
Database	17	56	52
Desktop Publishing	22	47	57
Software Development	37	52	39
Graphical Design	26	46	57
Engineering/Scientific	35	48	45
Web-browsing	49	20	60
E-Mail	51	21	58
Games	56	54	27
Computer aided learning	6	40	81

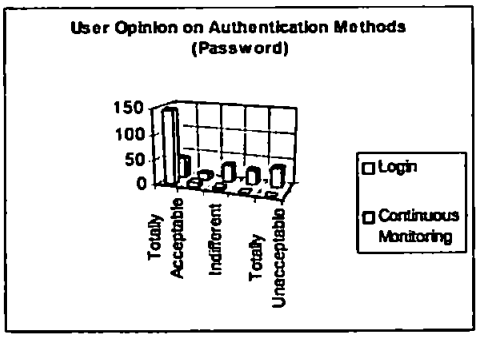
Question 46

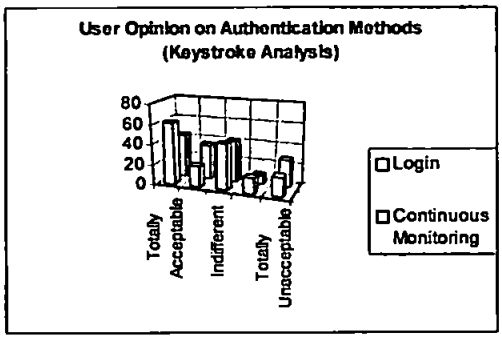
Login	Password	Keystroke Analysis	Face Recognition	Mouse Dynamics	Voice Verification	Signature Analysis	Iris Scanning	Hand Geometry	Fingerprint Analysis
Totally Acceptable	148	61	76	59	83	72	78	74	85
Acceptable	10	21	30	16	26	25	28	28	24
Indifferent	5	45	28	44	29	33	25	30	23
Unacceptable	1	15	15	16	11	14	10	12	10
Totally Unacceptable	0	19	12	25	12	18	20	18	20
Total	164	161	161	160	161	162	161	162	162
Continuous									
Totally Acceptable	40	43	39	46	32	24	34	29	36
Acceptable	13	35	22	29	27	10	16	18	15
Indifferent	35	41	39	40	36	32	29	31	29
Unacceptable	30	10	25	13	25	33	33	39	35
Totally Unacceptable	39	28	31	28	35	57	43	39	41
Total	157	157	156	156	155	156	155	156	156
Login									
Acceptable	158	82	106	75	109	97	106	102	109
Unacceptable	1	34	27	41	23	32	30	30	30
Login	157	48	79	34	86	65	76	72	79
Continuous									
Acceptable	53	78	61	75	59	34	50	47	51
Unacceptable	69	38	56	41	60	90	76	78	76
Continuous	-16	40	5	34	-1	-56	-26	-31	-25

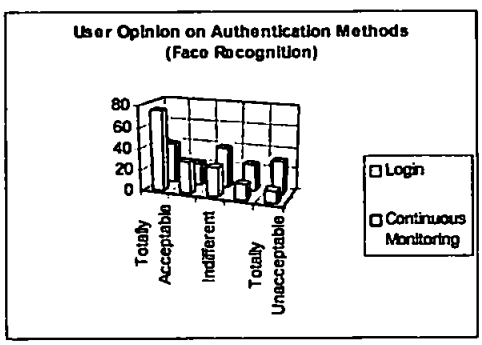
Question 53

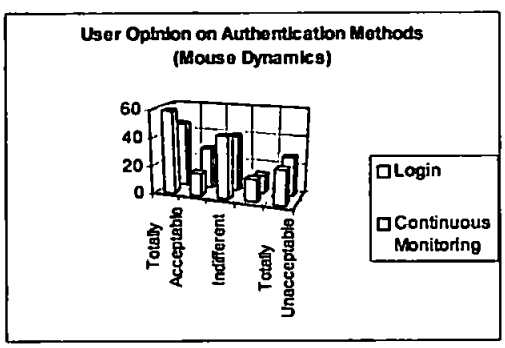
	Government	Defence	Education	Health	Social & Public Services	Banking, Finance & Insurance	Utility Services	Manufacturing	Engineering	Travel Industries and Tourism	Property Services	Communications Industry	Computing Industry
Great Benefit	128	152	40	85	68	129	45	26	34	17	27	67	83
-	27	10	42	46	47	27	38	41	43	31	33	43	39
Indifferent	10	4	62	27	33	9	63	76	63	80	79	38	28
-	1	1	11	5	11	3	11	18	19	25	15	12	11
No Benefit	3	2	13	6	10	0	11	8	10	16	15	9	9
Benefit	155	162	82	131	115	156	83	67	77	48	60	110	122
No Benefit	4	3	24	11	21	3	22	26	29	41	30	21	20
Rank	151	159	58	120	94	153	61	41	48	7	30	89	102

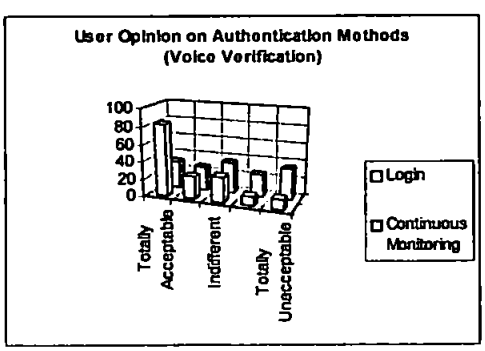
Additional Charts

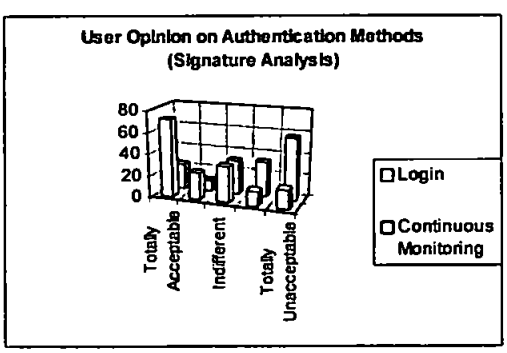


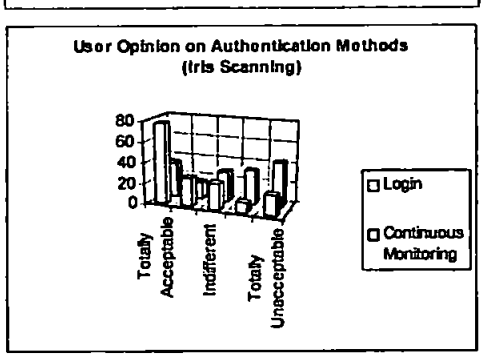


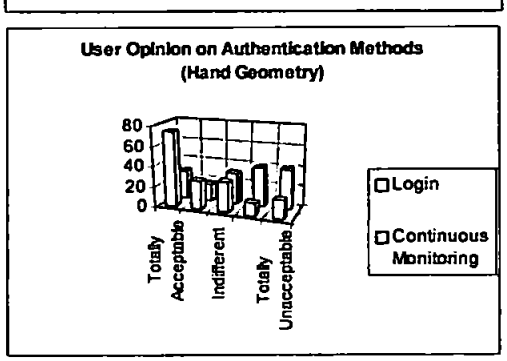


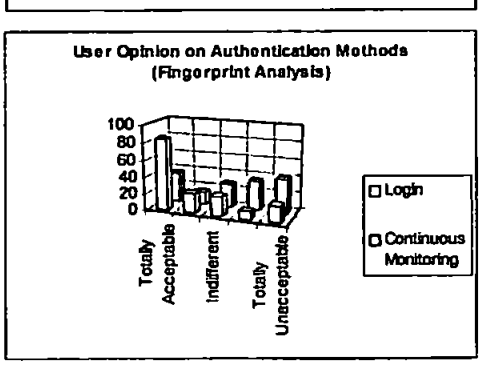


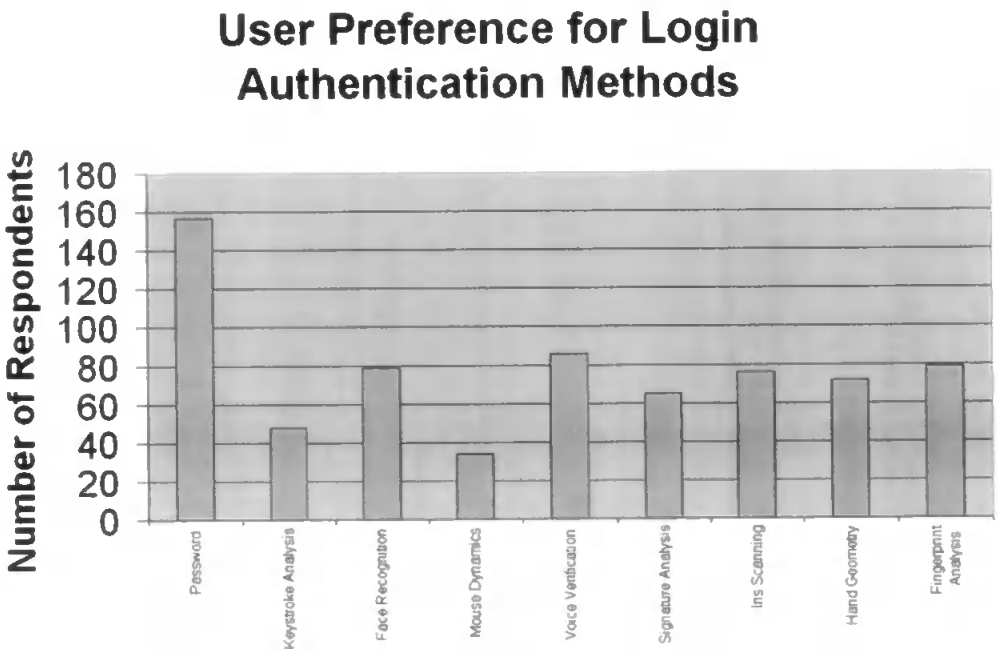


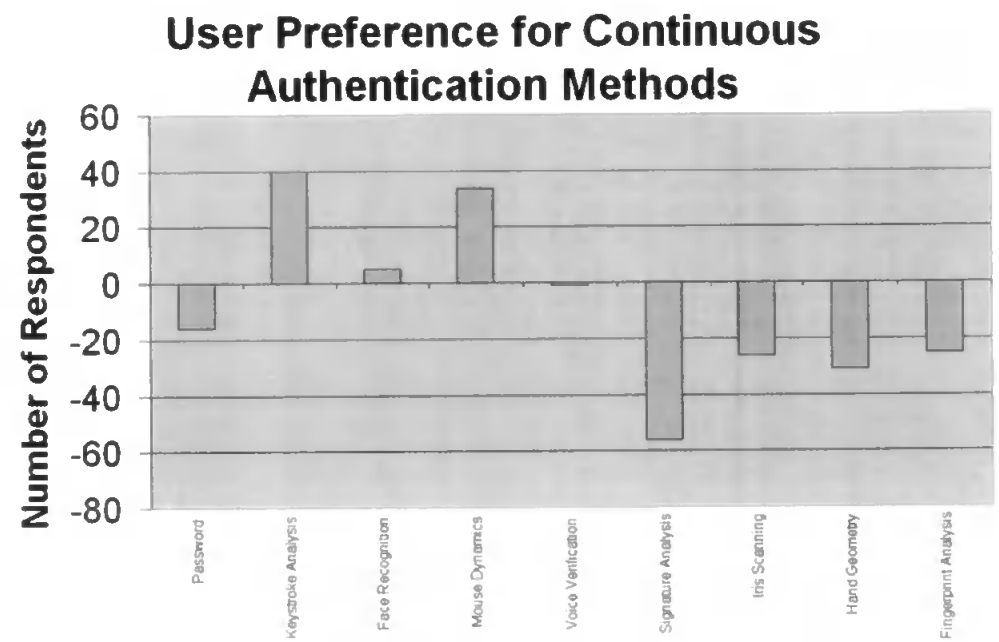









User Preference for Login Authentication Methods
Number of Respondents
180
160
140
120
100
80
60
40
20
0
Password
Keystroke Analysis
Face Recognition
Mouse Dynamics
Voice Verification
Signature Analysis
Iris Scanning
Hand Geometry
Fingerprint Analysis


User Preference for Continuous Authentication Methods
Number of Respondents
60
40
20
0
-20
-40
-60
-80
Password
Keystroke Analysis
Face Recognition
Mouse Dynamics
Voice Verification
Signature Analysis
Iris Scanning
Hand Geometry
Fingerprint Analysis

Appendix B

Initial Trial Results

Additional Trial Results

This appendix contains additional material to support chapter 6. The first part of this appendix presents the summarised results for the initial trial, together with additional charts not included in the body of the thesis. The second part presents summarised results for the application-specific profiling, toget-her with a selection of charts based upon these results.

Trial Results

Comparison of reference profile against other user raw keylogger data

Acceptance Rate

	User A	User B	User C	User D	User E	User H
User A	57	44	40	33	42	47
User B	46	**52**	41	36	41	42
User C	52	55	**51**	43	49	49
User D	23	32	29	**47**	26	30
User E	45	42	44	32	**55**	45
User H	46	41	38	35	41	**55**

Cumulative Alert Level

	User A	User B	User C	User D	User E	User H
User A	**332**	6749	15376	18986	2355	17409
User B	11996	**182**	12485	14636	2441	53941
User C	507	67	**182**	7709	360	8790
User D	82163	19976	29543	**3069**	6369	125025
User E	11669	6402	6486	12713	**42**	22572
User H	15574	10501	18905	17125	2667	**1750**

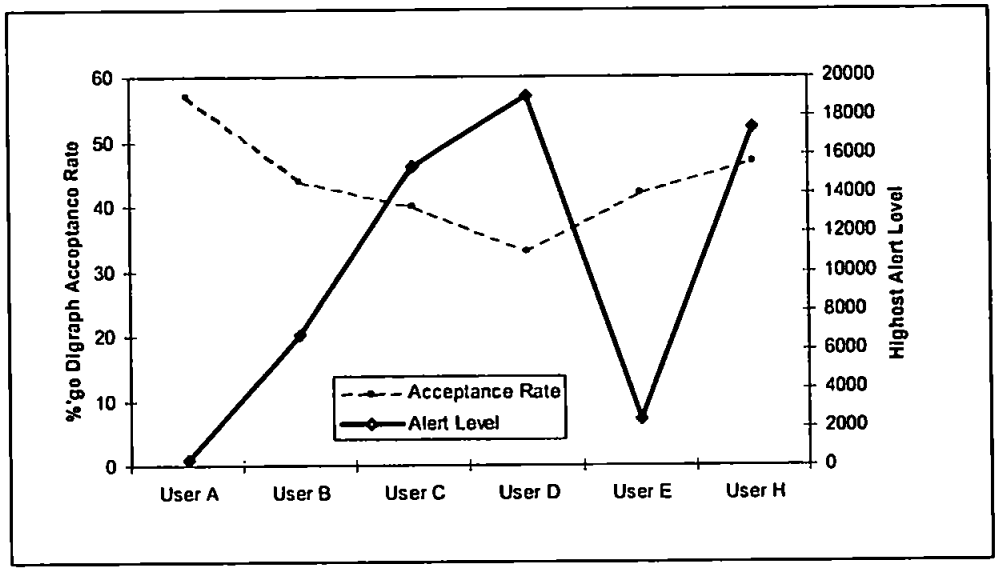
Highest Alert Level

	User A	User B	User C	User D	User E	User H
User A	**25**	26	28	26	23	49
User B	43	**16**	28	34	19	119
User C	26	11	**20**	22	14	100
User D	60	28	38	**19**	23	36
User E	26	19	34	35	**10**	129
User H	29	21	34	25	15	**64**

%'ge Cumulative Alert Level

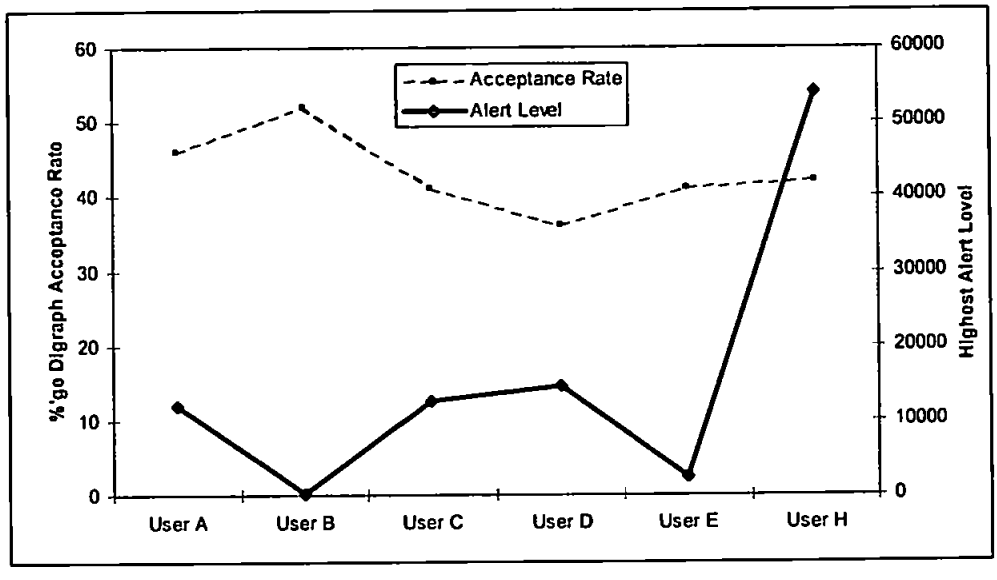
	User A	User B	User C	User D	User E	User H
User A	**0.19**	11.86	20.27	33.99	16.09	5.17
User B	7.88	**0.33**	17.36	27.86	17.8	17.1
User C	0.31	0.12	**0.24**	14.11	2.55	2.68
User D	54.32	36.91	41.36	**5.54**	47.24	39.67
User E	10.88	16.65	12.63	35.51	**0.41**	10.1
User H	8.94	17.87	24.12	29.62	17.62	**0.51**

Reference Profile : User A



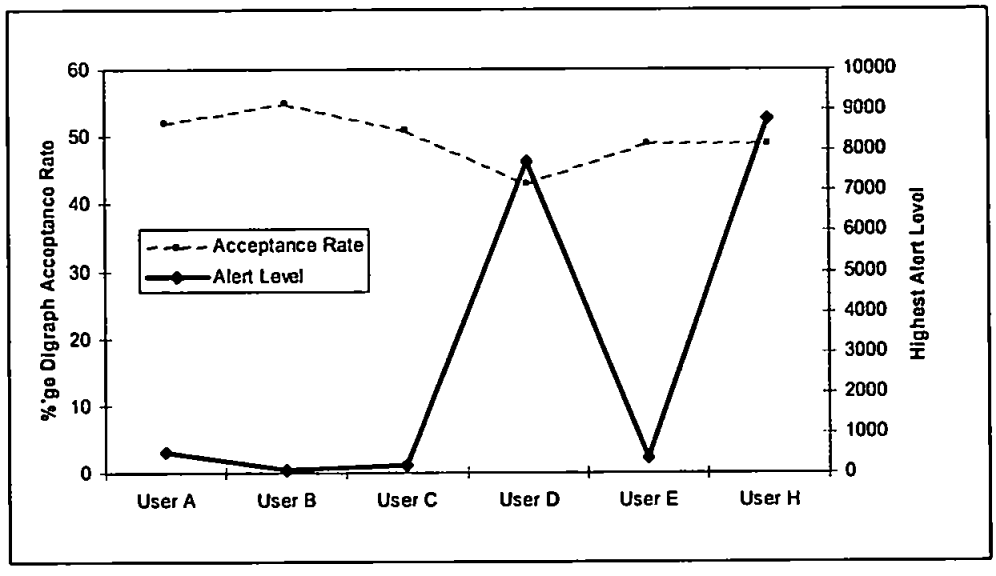
User A – 0% FRR, 0% FAR

Reference Profile : User B



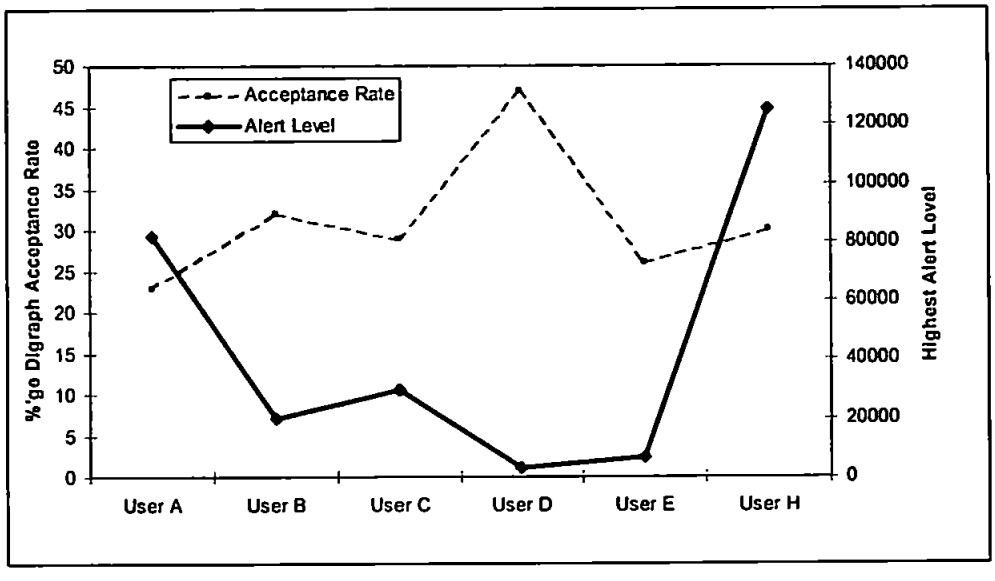
User B – 0% FRR, 0% FAR

Reference Profile : User C



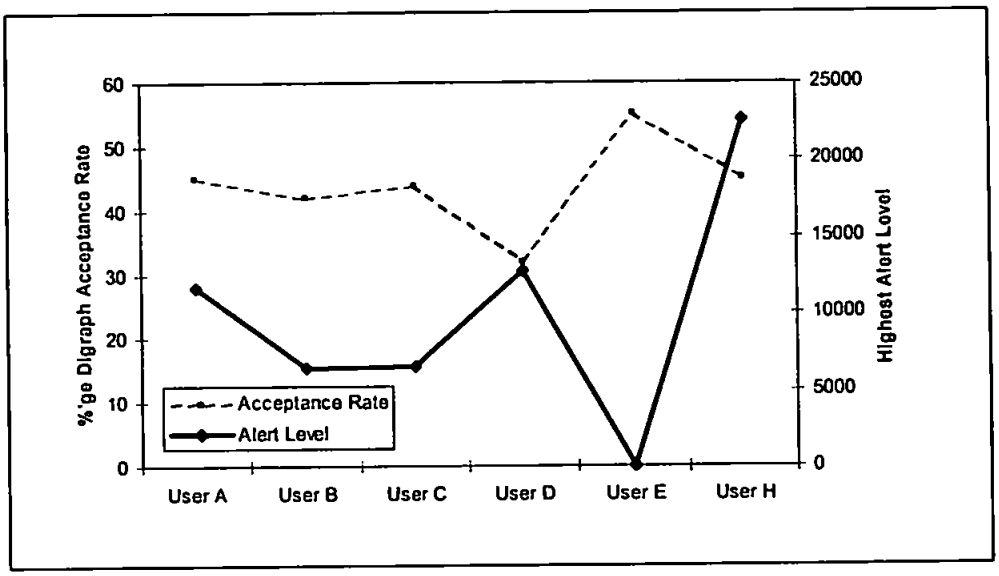
User C – 0% FRR, 3.3% FAR

Reference Profile : User D



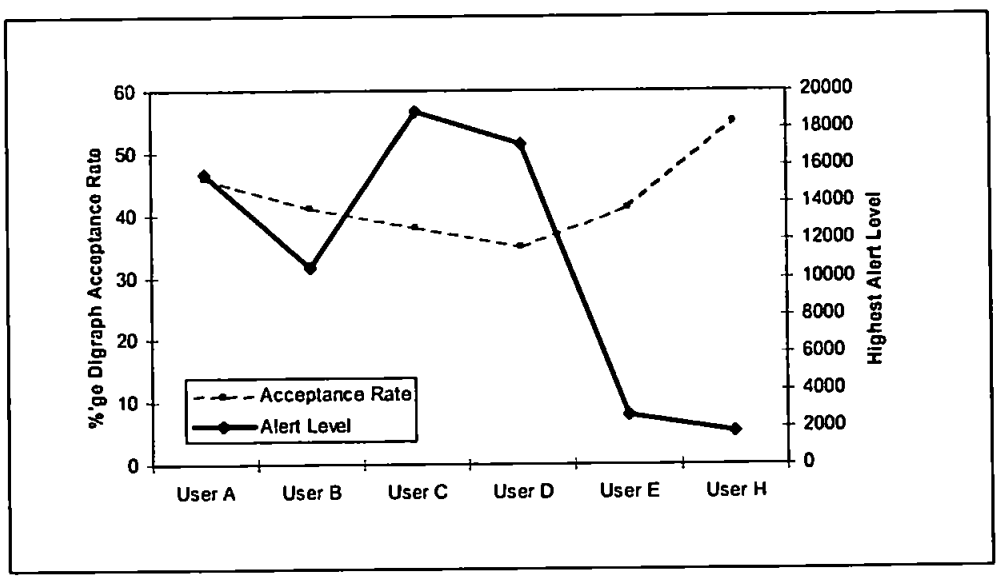
User D – 0% FRR, 0% FAR

Reference Profile : User E



User E – 0% FRR, 0% FAR

Reference Profile : User H



User H – 0% FRR, 0% FAR

Application Specific Results

Internet Explorer

Acceptance Rate (%)

	User A	User B	User C	User D	User H
User A	**56**	42	38	27	22
User B	41	**50**	38	29	13
User C	56	57	**53**	44	18
User D	28	34	27	**49**	23
User H	17	10	10	17	**50**

Outlook

Acceptance Rate (%)

	User B	User C	User D	User E	User H
User B	**54**	40	37	39	43
User C	56	**53**	43	51	52
User D	33	30	**48**	27	33
User E	41	42	28	**53**	47
User H	39	38	36	44	**61**

Word

Acceptance Rate (%)

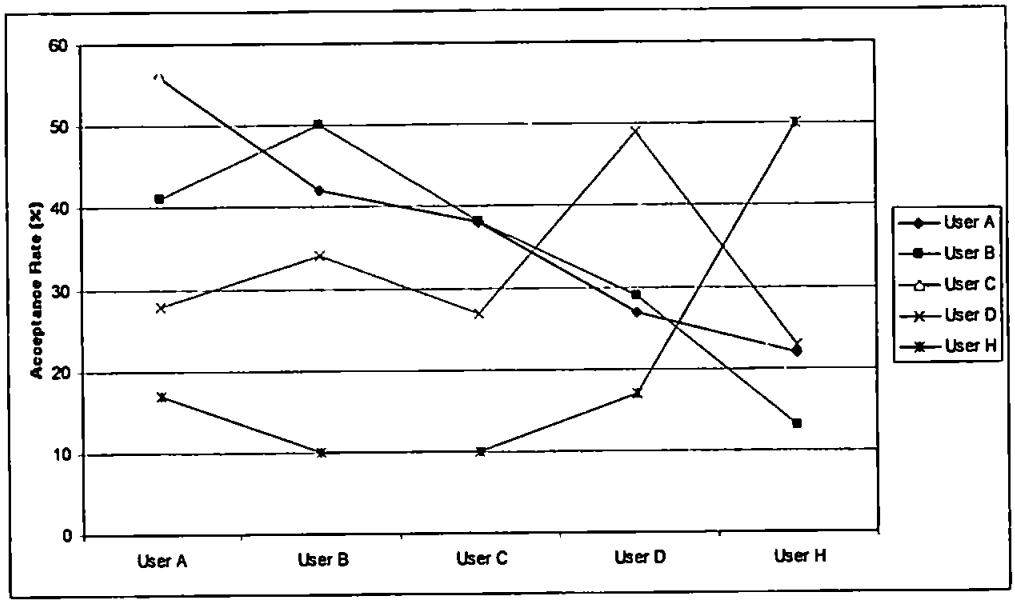
	User A	User B	User C	User D
User A	**56**	41	56	28
User B	42	**50**	57	34
User C	38	38	**53**	27
User D	27	29	44	**49**

Summary

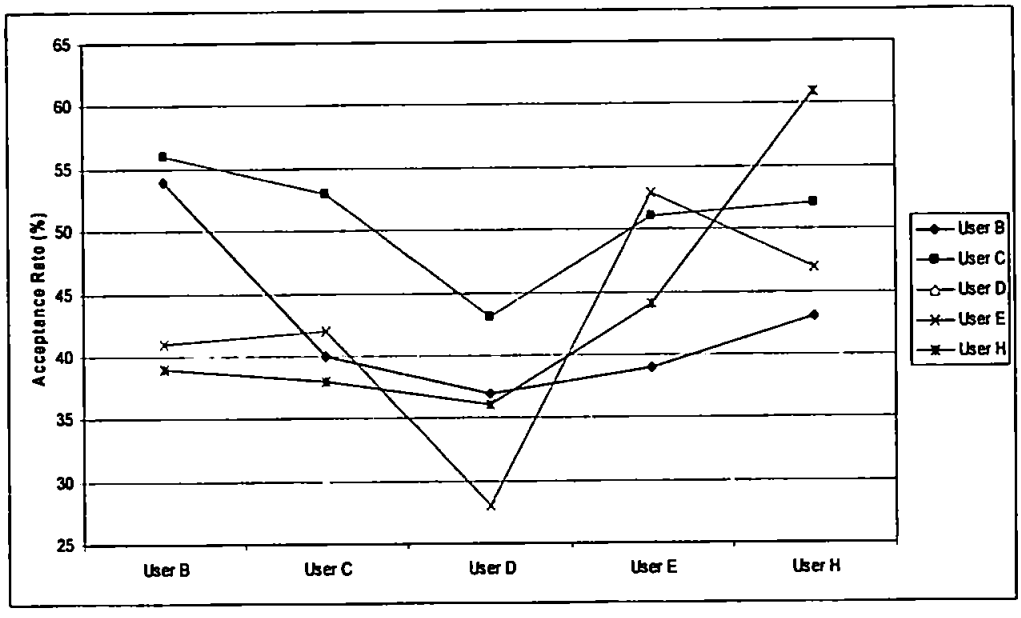
Acceptance Rate (%)

	User A	User B	User C	User D	User E	User F	User G	User H
ALL	**57**	**52**	**51**	**47**	**55**	**52**	**60**	**55**
IE	55.4	51.1	50.6	49.1	50.8	52.5	60.3	20.3
Messenger	55.9	52.3						58.4
Outlook		52.7	51.3	47.5	55.5	50.1	55.6	58.5
Word	59.1	52.9	52.6	46.1	56.8	51.3	61.2	57.4

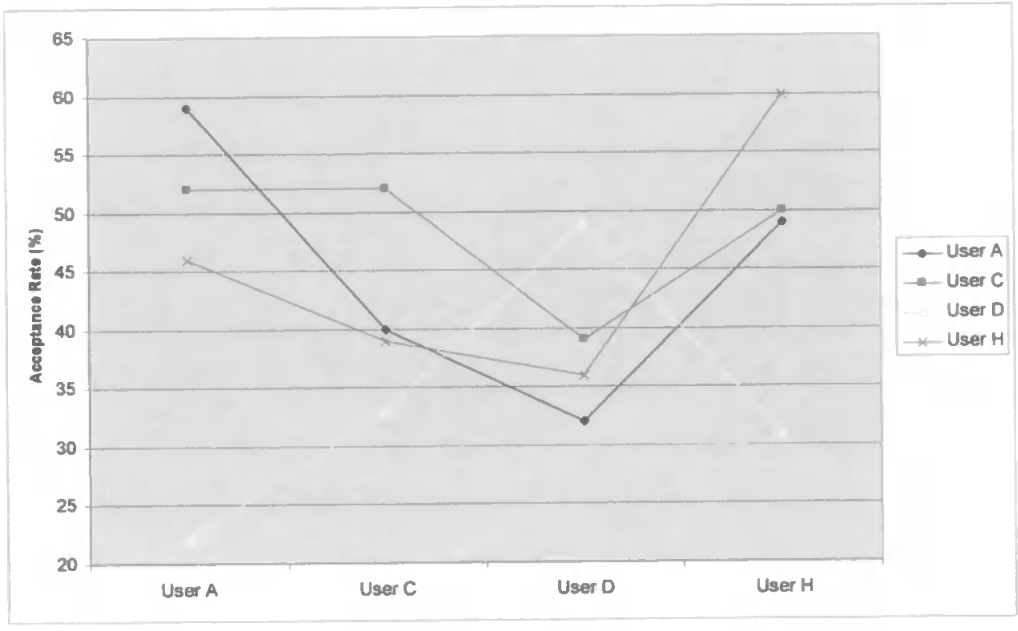
Internet Explorer Results



Outlook Results

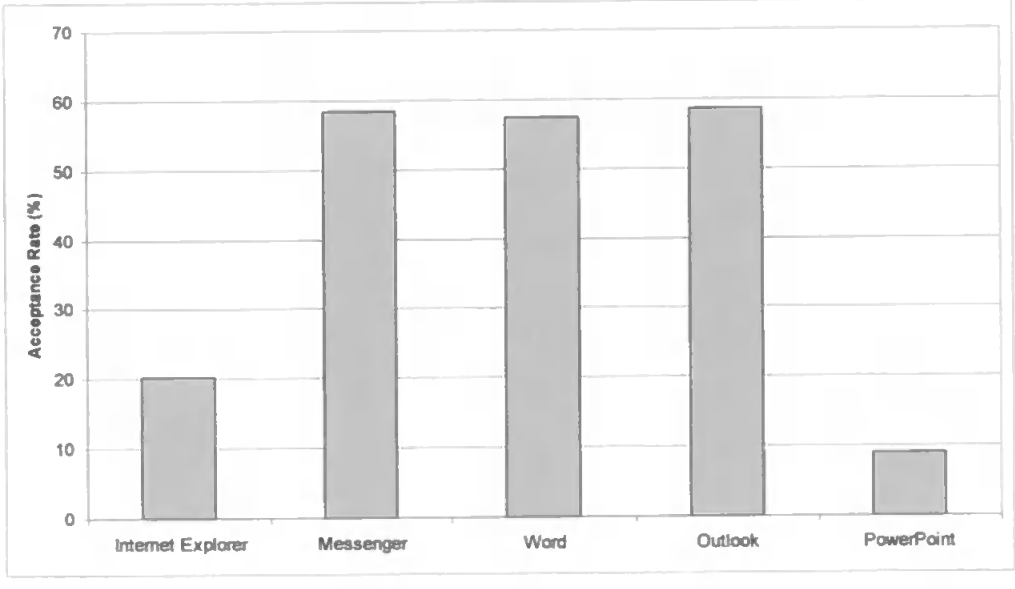


Word Results



Single User (H) Profile Comparisons

(raw keylogger data compared with specific application profiles)



Appendix C

Long-Term Trial Results

Additional Trial Results

This appendix contains additional material to support chapters 7 and 8. This appendix contains the list of keywords used for keyword latency monitoring. This appendix also includes the acceptance rate tables (in summarised form) for each of the main settings for digraph, trigraph and keyword. Finally, tables are presented summarising the effectiveness of the techniques by determining the number of keypresses required before an impostor (or the genuine user) is challenged. In all of these summarised tables, the reference profile is shown running across the table rows with the comparisons running vertically down the columns. Where the genuine user is compared against their own reference profile, the table cell is highlighted. For the challenge point tables (the last three) the optimum setting is achieved when there is a clear distinction between genuine users (with a high number of keypresses) and an impostor (with a low number of keypresses) indicating the number of keystrokes a user can press before the alert level generates a challenge. The other tables present the highest alert level of the system at various settings for the individual metrics and, as such, the aim is to obtain a low alert level for the genuine user and a high alert level for all impostors.

Monitored keywords:

word	little	number	keep	paper
network	very	great	children	hard
research	after	tell	feet	near
group	words	small	land	sentence
that	called	every	side	better
with	just	found	without	best
they	where	still	once	across
this	most	between	animals	during
from	know	name	life	today
have	through	should	enough	others
what	back	home	took	however
were	much	give	sometimes	sure
when	before	line	four	means
there	good	under	head	knew
your	write	read	above	told
which	used	last	kind	young
their	same	never	began	miles
said	right	left	almost	ways
will	look	along	live	thing
each	think	while	page	whole
about	also	might	earth	hear
them	around	next	need	example
then	another	sound	hand	heard
many	came	below	high	several
some	come	something	year	change
these	work	thought	mother	answer
would	three	both	light	room
other	word	those	parts	against
into	must	always	country	turned
more	because	looked	father	three
like	does	show	night	learn
time	part	large	following	point
could	even	often	picture	city
make	place	together	being	play
than	well	asked	study	toward
first	such	house	second	five
been	here	don't	eyes	using
people	take	world	soon	himself
made	things	going	times	usually
over	help	want	story	
down	years	school	boys	
only	different	important	since	
find	away	until	white	
water	again	form	days	
long	went	food	ever	

0.6 SD — Reference Profiles — Digraphs

	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17	User 18	User 19	User 20	User 21	User 22	User 23	User 24	User 25	User 26	User 27	User 28	User 29	User 30	User 31	User 32	User 33	User 34	User 35
User 1	19047	33023	15513	39709	25408	29829	25759	25879	28933	33860	27355	33655	25840	28470	32458	33850	12853	31433	24934	29780	34557	16532	9480	9908	40817	39386	28229	21485	18344	35553	16434	11034	14668	30217	14545
User 2	50452	205	42469	51553	46940	14129	34599	25338	49093	18573	54	42145	2588	32298	42856	14789	27319	53	22012	10998	8809	24643	17019	53226	34590	33365	31910	48239	32548	18801	45872	44883	50	56	6892
User 3	124132	152593	72	190712	116788	137404	112799	104255	114268	184752	119332	171426	120861	141661	154396	148103	70689	129124	69950	128996	183465	65418	66057	58731	197235	196778	133507	73682	61495	164947	73434	52908	75084	123282	70655
User 4	30765	18207	32450	6958	32068	17122	30757	29398	34005	28871	19155	16051	16883	27314	27813	21181	29921	9954	32771	29256	22596	28332	20408	33107	18245	12348	28086	33964	31454	7181	33143	32004	14777	13412	21001
User 5	33831	41905	21713	56434	8769	31228	27535	25987	41149	37765	21181	48228	27771	35484	46834	43072	13272	38891	23758	33292	44818	12813	3500	18498	56858	51462	31388	30882	23015	39657	19127	10536	3929	33919	5884
User 6	44841	25475	36460	49320	36292	5649	28922	24467	51401	34936	11598	38941	11091	39806	46853	34801	27879	14330	17153	26331	30841	20784	21948	38533	44158	39743	30748	44042	32201	22814	40002	30358	7837	15209	13845
User 7	54481	52679	29808	84851	32058	34765	50	27954	51618	49970	10135	65793	34402	45193	74683	61349	14700	43948	4594	15665	50181	50	2977	17677	78490	82498	30578	41731	22529	48187	9320	2377	55	36441	52
User 8	43884	38981	30932	56805	38824	34345	36592	16810	43528	36396	15144	47495	24528	39563	45545	41612	23295	27060	24750	31017	39899	18093	17290	39105	53028	48371	34889	30834	27416	36126	41883	33867	11464	28220	14318
User 9	29937	33078	14980	44797	27745	33721	24414	29185	18396	33931	21275	40715	28580	28802	36283	34771	14412	32158	29810	27822	33054	16866	7781	17399	43934	43932	32306	10064	16325	35828	12952	17538	10488	30121	9964
User 10	64043	38483	49329	73020	53834	50767	39963	38980	55272	212	401	62319	43252	29336	62600	43482	33655	28515	39888	24947	32059	25632	9385	57744	51309	56060	37331	54899	33720	35834	41343	50602	2160	32939	384
User 11	205763	148832	165327	255674	171398	163253	143791	128489	207641	113197	50	226768	129955	144575	209032	174703	93287	122873	128411	115088	153081	88392	59381	186501	201400	223042	156843	198468	125940	154002	173849	165121	21420	106905	22898
User 12	20474	8833	20580	15506	21192	6837	19475	14267	23374	13319	7794	9859	7786	17156	17566	12081	18750	960	14397	14069	12373	15008	12255	20653	12942	9923	18289	21935	18054	3923	20967	19888	5680	6269	10638
User 13	57986	36598	54845	64969	52013	28054	50038	41189	68156	48618	20810	54512	4879	54988	60642	50887	38274	28476	38180	37710	37706	39090	31432	59394	56838	53795	50922	63883	48796	39067	59538	49032	8124	25819	31003
User 14	28593	26133	22528	39578	22915	28258	18289	21862	27149	18909	5360	34739	25711	13048	32694	28902	14426	21352	17218	18107	22394	10805	6568	21273	33725	35675	20456	22810	12849	23073	19889	23199	10088	22885	96
User 15	14135	5728	9774	17017	11328	7506	7860	4679	11863	4393	67	14422	3799	8984	10641	7877	4523	3108	6097	3642	5581	3728	67	12084	15108	13362	7191	11959	5962	7777	9757	10078	67	2209	67
User 16	40078	1902	32348	38387	35849	11542	29454	17104	38537	12139	102	34757	1898	28033	34490	3888	18455	50	21216	12361	4383	18388	9627	43797	27285	21569	25870	38978	28432	16872	37092	38635	71	67	5915
User 17	82411	95771	36818	120968	71811	84906	67072	71889	78957	92680	63452	108877	72200	87824	103178	94855	2752	87081	62231	78371	98090	46891	27799	41143	114897	115436	92076	57155	48390	100073	39853	32770	20695	78145	35903
User 18	87007	47628	65447	91506	81878	59619	67720	46857	83118	67275	25728	77805	48145	71534	76284	61973	56806	367	45449	50560	66513	40656	39363	79077	84724	81292	66497	76866	50588	53776	76728	73095	19754	24232	31592
User 19	89280	82395	42925	121893	79037	73725	54988	44074	95074	70031	23631	100324	60119	68562	98882	88102	47157	53100	82	50345	79172	12480	31520	46853	105847	112081	58825	70120	29585	74273	54456	44046	37187	63647	5924
User 20	65984	31769	41558	86876	47494	37494	22103	32179	59998	23866	67	71401	25802	38705	68074	42620	22234	29953	19537	314	18815	11801	4401	51932	70574	72286	35994	53948	25748	47021	38537	39370	68	19696	167
User 21	77378	18213	61197	77186	64836	37848	48354	44405	70292	17472	84	70599	18618	41387	67280	28888	38058	20822	42103	5952	53	34881	13854	75779	58053	54876	45478	71718	42927	38685	56735	62683	171	7968	1804
User 22	86840	93192	45379	132784	72170	80279	62581	57030	85108	88902	37064	112728	70870	80128	104203	99787	32689	74938	41791	71130	99931	58	7804	27088	133548	124808	70533	59791	30912	89342	28087	38281	20144	77126	1730
User 23	98380	109803	70889	136538	88705	105242	95647	91420	101562	110548	78648	118961	88482	101750	112127	111698	63980	96082	93183	100026	114485	65579	20748	61502	133535	128327	105736	90734	70799	108796	73801	67177	46590	95767	51908
User 24	134325	170331	58372	241778	127352	155824	142162	112493	150457	204813	138008	195813	122100	161603	180115	180207	80594	151105	94130	153742	190847	80771	41669	67	255186	220823	142859	96410	65335	176785	61907	32106	75774	169704	49884
User 25	41323	16763	45697	34969	41502	27818	39024	27329	41912	16523	6165	28631	20584	28618	35952	22899	31928	2642	32518	26020	21763	34078	21375	52723	10199	21765	33523	46823	40435	13313	46787	49295	6952	14801	18188
User 26	49651	8452	54008	31663	51731	18006	52098	39186	53828	19041	11867	24317	10352	37582	43841	15702	42014	53	43740	32028	13897	43616	23765	63607	16149	62	41809	57690	47478	8784	59851	54600	4101	7299	20891
User 27	26331	28102	10809	37255	20093	23636	15094	16277	23820	23581	12203	32482	21294	21174	30825	28095	12733	20692	9903	17105	25882	5417	7438	10321	37530	38252	14280	16187	9843	25580	11394	10320	10012	20881	3396
User 28	54569	66161	10513	81684	52818	59632	48104	51355	46151	64584	43777	73543	54728	57727	68912	66226	24998	58503	30534	51789	65170	24827	23303	22272	85524	84658	53880	14367	18994	65257	22087	23200	35147	58103	19829
User 29	71368	81951	31960	104597	67028	76428	63144	58229	65946	73544	39618	91884	66905	62785	79580	83621	34825	66912	46079	60485	79328	28655	19722	37968	105827	101504	68094	48786	11619	80180	37727	38492	33856	70975	17638
User 30	41526	29218	37715	38878	38323	30180	36939	31469	41908	29019	18082	36120	27144	33798	38439	33712	35176	17198	31001	31321	33075	28042	23943	38415	35578	35862	30793	40458	31968	15406	40131	37909	20698	24487	19333
User 31	184308	253547	8488	367643	149186	235563	111387	171822	153428	280580	122038	316033	207803	176395	288198	253092	32706	241077	118073	153834	247559	72	69	84	391719	376938	150627	32349	14285	277459	50	9181	52915	212707	72
User 32	257485	355261	150798	429581	181358	297123	220529	230882	297884	339170	177697	374702	259712	273590	355291	372897	140871	296865	118938	274106	383888	85458	72934	73002	441819	435760	224218	171654	105060	309829	114760	62	180401	277019	84287
User 33	234802	233238	197230	282181	190143	208304	188655	200682	255299	225072	135058	284816	187648	220351	248418	259494	115232	211796	214053	199623	234468	161772	105020	199364	297802	283374	217959	231914	175755	214800	188924	169144	250	168512	122621
User 34	117981	83213	94940	122210	91885	56619	83139	86224	118926	82716	19709	103541	42941	107900	109328	80219	63683	39938	75444	61834	89900	73752	52285	106276	97987	98375	94195	112897	83657	67184	87674	62725	9752	280	42768
User 35	113231	112733	61550	159163	103016	111414	93790	93892	107550	105287	52163	139568	95986	99857	129032	118107	55515	92507	75718	90829	115385	44148	26513	67014	149138	148558	105530	84820	55901	111529	61088	73951	34293	91770	2890

total	1	0	0	0	0	0	0	4	1	0	0	0	3	1	0	0	0	3	0	0	0	1	24	1	0	0	1	1	2	5	0	0	6	2	8
FAR	2.86	0.00	0.00	0.00	0.00	0.00	0.00	11.43	2.86	0.00	0.00	0.00	8.57	2.86	0.00	0.00	0.00	8.57	0.00	0.00	0.00	2.86	68.57	2.86	0.00	0.00	2.86	2.86	5.71	14.29	0.00	0.00	17.14	5.71	22.86

35 Users
1190 Impostors

Accept Count 84
FRR 0% Fixed
FAR 5.4%

0.6 SD Reference Profiles — Digraph

Unmatche	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17	User 18	User 19	User 20	User 21	User 22	User 23	User 24	User 25	User 26	User 27	User 28	User 29	User 30	User 31	User 32	User 33	User 34	User 35
User 1	170	16230	9324	15224	7832	12552	12440	9150	7976	19294	21608	6578	10892	6522	3186	15052	4150	17564	14814	15938	19840	9956	2432	5868	20280	21274	6338	7844	6342	15740	11018	7650	10920	18400	7812
User 2	15466	50	35060	19982	24470	54	22956	4508	21124	5098	54	5316	60	3280	1170	68	15422	50	12786	416	76	15716	7606	48888	9984	13440	3960	33572	20544	66	43310	42712	50	54	2748
User 3	15770	65108	58	83982	36970	68344	64414	31820	25746	109514	101254	51294	59952	47228	23098	70292	38102	78800	33236	75966	107064	32360	36212	42038	108172	120882	36944	21204	17112	83266	57010	42250	62346	77794	46724
User 4	13042	9240	29360	50	20930	7820	25090	18996	19782	20130	16596	52	9038	12434	6710	10710	23810	3218	27718	22584	15036	23770	15548	30874	5920	2424	13970	26460	25676	62	31998	30868	13466	7252	17940
User 5	724	19176	13002	22650	60	6820	11128	2932	10562	19328	15410	8798	8918	2700	5388	18086	1880	22544	11770	15408	25978	624	50	12420	28338	27158	718	12806	7674	12912	13550	6546	150	19500	70
User 6	11186	7290	29692	18330	15108	54	16520	4016	24502	20752	6700	3068	52	10848	6784	13742	16210	1058	7928	13190	16140	11572	12018	31838	19816	19884	3190	29632	19750	74	36740	27890	4768	3456	6992
User 7	3028	27214	20720	38814	484	7568	50	98	10748	31020	4360	12346	10838	2682	14042	30946	58	25552	50	82	28820	50	72	11828	40920	53672	76	20606	5826	14068	6286	586	54	18618	52
User 8	12422	19144	25186	24900	16470	14208	22898	620	17172	21106	10840	10692	7818	10942	5260	19210	12404	12182	15242	16832	23908	9608	9502	35222	27774	28856	5888	24538	15676	11618	38078	30816	8122	14508	6038
User 9	3740	16684	9182	18356	6590	17068	12054	11824	142	19892	16968	11632	14526	5946	5340	15578	6130	19230	20778	14722	19756	8280	656	13380	22902	25570	6482	7150	5312	15842	6844	14684	6494	17808	4140
User 10	17590	15802	41106	32264	24918	26410	25588	12354	18412	54	170	14090	22388	66	7704	16344	18398	11944	28382	8734	13032	15028	164	52598	20046	30380	1072	36140	18854	5312	38842	48244	248	17374	186
User 11	57330	49950	132354	104718	65956	61132	73382	26268	82642	33932	50	53382	46904	7074	22806	58314	44218	47384	74764	41070	69786	43430	19832	101762	76238	113082	16764	123036	63714	32806	153746	170078	3368	40906	86
User 12	7142	1928	17938	3566	12782	1876	15288	6596	12478	8210	5854	52	1728	5794	1990	4070	12282	54	10912	9192	6528	11534	6372	18984	3924	2428	5778	16536	13714	60	19790	18730	4408	1524	8282
User 13	12314	9126	42926	21262	21206	848	29934	11004	32042	26530	12118	5880	58	14494	6136	19460	21574	5548	22202	16258	15678	24066	16040	51806	21442	24310	11150	41840	29548	5074	53818	44970	1978	6160	20708
User 14	6914	14204	18774	18490	7930	15496	9866	8440	10238	9212	2972	10056	14916	58	5844	14218	7784	12252	11082	8822	12222	5090	1328	18422	17068	21848	1838	13552	5074	7572	17934	21554	7836	14130	52
User 15	3934	528	7860	7550	4938	2012	4320	70	3970	572	72	3442	94	512	68	1464	1128	64	3574	150	1484	1160	64	10988	7984	7412	72	7730	2228	618	8978	9588	64	108	60
User 16	11992	52	26400	13114	18450	66	19942	926	16420	1840	92	5080	52	4406	934	60	8624	50	13976	2296	80	10960	1930	39944	7624	5920	3646	27314	16586	54	34706	34554	60	52	702
User 17	9774	39806	18376	41970	16318	28828	23512	15448	13704	43858	48506	20930	24452	16092	9924	32158	52	42910	28282	31774	49292	16126	3032	25194	46092	54528	16964	12672	8940	34662	22470	20436	7752	38678	13562
User 18	27560	17352	54068	37872	44110	27528	40122	10904	36738	42920	17338	15912	17454	22022	5166	25722	37918	56	29344	26554	41406	30168	23424	71780	42172	46272	19718	50830	29246	14562	71178	69004	14032	3558	18604
User 19	20186	40632	28374	51068	29076	30368	24834	1418	38146	36384	11908	21810	22414	6910	12118	37362	22478	21812	52	19018	44480	54	14002	35732	51338	65778	78	37558	3530	20770	43834	37058	28856	33064	62
User 20	14678	5168	31578	39932	14678	9710	5070	2318	18902	4144	56	17772	1548	50	5826	11430	4182	10256	6924	54	52	1584	214	45562	35188	43254	80	32646	7648	11680	34520	36022	62	678	98
User 21	24442	50	50716	28838	32390	9706	28458	12786	29774	366	54	14436	52	162	4218	52	20292	2902	28456	70	50	20918	560	69280	20566	24354	2846	49686	24540	4976	51876	58636	100	50	76
User 22	9890	46108	28222	57690	15818	32268	29532	7272	20348	51412	23024	27788	28784	11042	10598	45030	4756	38808	16380	35736	59060	58	52	14810	73654	73206	2878	21812	1266	30802	17956	27052	10170	44284	54
User 23	18494	49950	49730	51300	27110	45136	49518	33984	31556	60232	62664	22862	37216	24474	10498	44856	33664	49336	59682	51114	64398	38006	7930	47060	60974	64028	26720	43438	31402	39802	58114	54536	33858	53674	29366
User 24	5276	79828	24562	107990	33178	63834	75092	22568	35044	131352	111776	45984	42262	34400	19412	78584	32684	61334	46062	63910	115208	18064	5960	50	143530	133104	20602	23918	7850	72710	28960	14920	56960	105678	16098
User 25	15602	3518	40598	11250	25154	13718	31234	11400	20740	6242	3358	1424	8374	6604	5588	7130	22774	70	25888	16698	10752	27370	14498	49492	62	7268	12522	35942	31218	50	44968	47776	4982	5884	13384
User 26	17906	54	47820	2764	32078	1650	41916	20584	28722	7148	7846	50	52	11190	6508	52	31544	52	35548	20740	984	38020	15522	59734	62	50	16868	44648	38980	50	57810	52880	1830	58	15278
User 27	4152	15762	6038	16296	5650	10756	6640	2140	5438	13864	8528	8506	10076	1906	4156	13382	4672	11534	3030	7776	15274	280	1118	7028	21002	24518	62	5842	1492	10108	9212	8308	7560	12858	196
User 28	4938	36042	370	32962	18718	28748	28202	19138	6610	40134	35574	19622	27144	14958	11012	31456	7792	34942	14702	28810	39652	9728	6852	14230	46424	50792	9442	100	74	28456	15430	17614	29210	38568	8486
User 29	11566	45232	19596	46114	23824	40560	37298	19840	16024	43984	29488	26854	34828	10632	7568	41472	13672	40196	26934	32360	47088	11824	3530	28712	58908	61604	14048	20482	160	34460	30276	31846	26504	44402	5264
User 30	13332	13424	31350	12072	17238	12992	25722	13114	18706	18098	13200	5202	12684	9260	4982	15024	24892	5108	21934	18954	20110	19772	15546	33806	13986	17590	6894	27408	20784	58	36374	34630	17146	12606	13138
User 31	88	140638	62	170838	14478	117458	37140	44728	72	175608	93138	102630	105360	84	44410	120724	68	157452	58250	73400	150938	54	64	54	245770	255280	56	56	50	130610	50	184	34962	139642	60
User 32	15828	216358	101760	197714	22940	153548	122110	77022	108706	228396	143134	118104	140016	62114	78366	212290	57846	192588	46200	168128	244854	12800	6464	39352	260736	282992	17382	65978	13910	133102	66190	62	138068	181978	33932
User 33	33976	69336	137136	69450	54072	66432	88220	52980	75876	103818	93614	43664	44312	29710	11462	101988	34704	102508	124674	81058	111648	77974	34444	150440	124176	128476	28028	114908	71572	50872	149148	132254	86	79204	54328
User 34	20000	6840	71368	38190	38038	1534	43556	21688	38590	38502	3810	10080	244	26134	7376	14318	31200	304	40348	16316	24578	37280	21804	66074	28518	37140	16820	64774	39292	1156	75356	69930	1742	50	16954
User 35	17938	48984	38726	62600	30766	46318	46890	27912	26016	53180	33540	30938	40468	12528	11714	44828	20188	42102	39406	41282	60312	14926	850	49674	70438	79242	17658	33378	15374	34842	45514	60782	20030	46030	60

total	1	1	0	0	0	1	0	2	1	0	0	2	4	1	0	2	0	4	1	0	0	3	19	0	1	0	1	1	2	3	0	0	5	1	5
FAR	2.86	2.86	0.00	0.00	0.00	2.86	0.00	5.71	2.86	0.00	0.00	5.71	11.43	2.86	0.00	5.71	0.00	11.43	2.86	0.00	0.00	8.57	54.29	0.00	2.86	0.00	2.86	2.86	5.71	8.57	0.00	0.00	14.29	2.86	14.29

35 Users
1190 Impostors

Accept Count 61
FRR 0% Fixed
FAR 5.1%

0.7 SD	Reference Profiles			Digraph																													
Unmatched	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17	User 18	User 19	User 20	User 21	User 22	User 23	User 24	User 25	User 26	User 27	User 28	User 29	User 30	User 31	User 32	User 33	User 34	User 35
User 1	52	12834	2178	14432	2472	8808	6706	4714	3980	15278	13628	5142	6512	2800	1930	12260	54	13236	9904	11170	15800	2804	74	62	18412	16898	3082	3026	1094	12056	3842	752	828	12480	342
User 2	12866	50	28444	15618	19194	52	14338	252	17356	50	52	1114	50	98	88	62	8038	50	1564	50	56	5956	380	40528	2324	4846	342	28184	12516	52	35654	34880	50	50	54
User 3	3054	63868	50	81002	15978	46284	35038	7760	5288	66466	58278	46862	33090	25248	15114	49798	204	49218	4182	47540	84000	160	84	8898	99064	110894	19074	84	62	64844	17800	6060	9026	49294	84
User 4	12190	6380	27656	50	19408	5344	23078	17182	18974	15930	12004	52	8426	10594	5990	7878	20446	54	25278	19380	12032	21108	12960	28722	1304	58	12746	25216	23576	54	29530	29132	9810	3776	14242
User 5	72	12444	3826	21498	58	2166	3068	58	5824	10852	2894	6786	1272	70	3162	12358	50	13342	3728	7282	18606	54	50	2022	24350	22630	70	5634	586	7040	2276	66	54	9018	60
User 6	8006	84	22582	14518	10260	52	8512	58	20422	11556	54	88	52	5708	3694	5446	7002	52	228	3454	7432	2048	3420	24318	13734	11136	98	23958	12786	52	29200	19860	58	58	166
User 7	50	11452	3000	35828	52	62	50	56	954	12000	50	8818	50	60	6654	17754	50	5488	50	66	11680	50	50	50	32396	44112	70	6358	52	3982	50	50	52	50	50
User 8	8414	11160	16374	23100	10410	6452	14834	58	11652	12388	54	6004	130	4970	2328	11848	2072	1326	5018	7132	15960	62	60	25144	22754	21212	1000	17538	6730	4850	28492	21028	60	3934	58
User 9	1266	12618	1122	17660	4284	13138	6282	6052	58	13812	6140	10724	9378	1050	3816	11934	56	13428	13214	8942	14692	330	58	4512	20610	23196	4894	856	58	12002	474	6186	70	12394	72
User 10	13906	4702	29166	29420	16600	17704	14344	976	12034	50	70	9066	10572	74	3556	6900	5992	232	13942	212	338	880	94	39530	9502	21612	50	26380	7038	112	25738	37230	68	4070	62
User 11	44142	18896	96272	97366	41348	35148	41482	170	63230	50	50	42788	12236	70	13496	34562	5594	2304	31126	3380	35986	228	56	122842	54210	92692	96	97104	29288	6118	117050	135098	56	3432	60
User 12	6316	56	15858	1546	11150	102	12908	4516	11494	4722	1776	52	56	4132	1106	1474	9310	52	8244	6508	3694	9090	5336	17184	670	76	4148	14808	11662	60	17492	16574	68	54	4916
User 13	8338	70	34016	16782	13688	68	19546	1842	26506	17134	50	2032	50	8582	4284	9724	11148	54	10382	4078	5868	11330	5216	41382	14918	15394	5030	34896	19830	50	43190	34058	50	62	6060
User 14	4892	0896	12878	17606	4012	11052	4674	2556	6862	2130	56	6824	6396	54	4420	10330	1622	6232	4190	2910	7042	134	54	12446	13980	18792	54	8632	404	2872	11650	15834	84	8826	52
User 15	2866	66	5600	6586	2712	114	1708	66	2542	66	66	1906	68	60	50	92	66	58	452	84	66	68	64	8188	5548	4808	62	5762	68	52	8178	6976	64	60	60
User 16	9508	50	19774	10350	13670	62	13602	64	13216	52	60	1140	50	56	50	50	2556	50	4136	116	60	2480	52	33870	588	52	102	22590	10482	50	28110	27078	52	50	68
User 17	1874	27980	98	40182	2368	17220	6644	462	1460	30060	20556	17934	7840	4022	6478	22362	52	27822	11262	16530	38920	84	52	2510	39128	48880	7908	130	56	27188	270	60	50	22038	52
User 18	23184	3048	42168	32988	36050	14120	34570	122	29916	28476	70	9040	2512	9476	1090	12898	24478	50	13326	9662	24052	14170	6762	59316	30396	35660	7724	41228	15038	738	57466	55480	60	50	742
User 19	13178	22484	8616	48976	17500	13292	6128	74	27622	16840	76	16566	2224	68	7626	23166	260	946	52	132	26770	54	52	18278	41444	54994	52	21754	52	7706	22590	17224	62	10092	54
User 20	8434	54	16486	36752	5064	58	108	78	11218	60	50	11726	50	50	368	66	194	58	106	52	52	94	62	29488	25996	32770	52	21552	154	138	19048	21106	54	54	50
User 21	19194	50	36504	23856	22488	52	14170	850	23052	50	50	6826	52	52	258	50	8048	50	6758	50	50	2926	68	55900	8578	11988	66	39724	9996	50	38516	45120	50	50	50
User 22	64	28068	5328	54398	56	16532	8502	58	5688	28736	50	22820	7564	50	3530	29174	52	12992	50	12168	40840	52	50	56	62588	61694	50	3914	50	13746	52	466	52	19074	54
User 23	9792	38626	27894	48712	12982	33512	32638	18228	19578	44072	35422	19276	21264	12020	6166	35096	10480	32436	41514	34688	50390	13644	90	23348	53706	58114	16590	28152	13504	27968	31870	31686	5038	35742	4540
User 24	52	56692	50	98028	1230	32114	33472	112	9816	99574	56662	35636	8588	10626	9352	59504	50	47092	10222	50408	91172	50	60	50	129538	114388	82	50	72	46070	64	52	62	61592	52
User 25	14562	72	37316	9010	22286	9584	27354	6772	18744	368	64	78	3048	3108	3948	2342	18310	64	20078	11158	4828	21870	9248	45388	58	2000	9014	33454	27302	50	41038	44324	184	510	7512
User 26	16238	54	43304	52	28442	52	36784	13486	26554	254	58	50	52	5682	4118	52	25676	52	27212	11928	58	28328	9248	53842	54	50	12064	41796	31592	50	51826	48344	58	58	5704
User 27	1278	9928	74	14936	1136	5064	306	54	1790	5744	188	6838	3160	58	2012	6566	60	3528	92	500	9266	64	64	634	17972	20094	58	568	54	4470	2002	952	62	5004	64
User 28	344	27650	56	31318	7274	20164	14670	8814	74	28868	18214	17162	17196	5290	7624	24176	84	24534	2822	17104	29896	72	60	176	41582	45424	1974	58	72	19040	106	2622	6906	24552	80
User 29	4770	33244	1432	44534	13510	29868	22074	6340	5984	29296	7010	23010	20374	250	2584	31692	122	24742	0146	16448	33914	96	58	9304	52776	55072	6110	6398	60	23572	10896	14022	678	28120	54
User 30	11448	7828	26138	9720	13806	6128	20742	6578	16228	9624	4522	2936	7064	5468	3580	10572	19172	92	15846	12982	14464	13308	8626	28854	9530	13162	3654	23368	15904	50	30834	20038	8032	5690	5614
User 31	50	100082	50	169638	58	78710	76	124	64	95842	9942	86894	55934	58	29128	83920	62	95350	124	5892	94022	50	60	50	213300	216696	52	50	50	76026	50	54	92	76984	58
User 32	72	169322	25084	187378	72	99420	62310	14924	57258	154876	48818	98600	77164	12096	60562	173388	54	130526	80	103532	198714	66	58	82	241068	251600	78	4408	56	81210	124	56	8180	112424	54
User 33	17712	57464	95044	80486	20668	37912	52460	20946	55928	70588	34368	31520	6536	6746	2820	76152	244	57948	84970	43884	77956	31124	668	102844	105064	105800	6582	66120	33166	25636	102762	83630	50	31938	2486
User 34	22442	54	49692	31154	22118	54	26212	6276	31230	18810	52	2222	52	14514	1468	714	11418	54	18332	304	4898	15208	1960	65790	16074	23244	4930	51214	20348	62	55250	48374	50	50	50
User 35	8772	31424	13824	59104	15698	31368	25056	6270	11982	29360	60	25358	10934	58	5770	30332	568	17422	14222	17576	40068	344	56	23232	57698	67208	3706	14834	824	17978	16324	34738	70	22072	54
total	3	2	2	0	3	3	0	2	0	3	5	2	4	3	1	1	4	3	2	2	0	3	22	2	1	0	6	2	8	5	1	3	5	5	12
FAR	8.57	5.71	5.71	0.00	8.57	8.57	0.00	5.71	0.00	8.57	14.29	5.71	11.43	8.57	2.86	2.86	11.43	8.57	5.71	5.71	0.00	8.57	62.86	5.71	2.86	0.00	17.14	5.71	22.86	14.29	2.86	8.57	14.29	14.29	34.29

35 Users
1190 Impostors

Accept Count 120
FRR 0% Fixed
FAR 10.1%

0.5 SD	Reference Profiles			Trigraph																															
	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17	User 18	User 19	User 20	User 21	User 22	User 23	User 24	User 25	User 26	User 27	User 28	User 29	User 30	User 31	User 32	User 33	User 34	User 35
User 1	20508	24715	19824	24488	22382	24829	23887	22683	22965	24944	23855	23817	24390	23921	23814	24729	20278	24485	22759	25034	24664	23090	19286	16827	24577	24906	24072	21098	23064	24327	20387	15857	23990	25228	23011
User 2	37720	52816	37267	38696	37557	32591	36983	36517	36564	33306	29673	36307	32634	35368	37224	34941	40318	31478	35011	34493	32939	34789	33835	41182	37105	37393	36118	37815	35033	34977	43277	43844	23544	32240	34515
User 3	105173	111058	65769	109985	106434	100566	108468	107946	107789	112884	110248	108195	109154	108926	108110	110691	91931	112571	106949	112487	112151	100891	96030	85722	110603	112025	108752	102082	104542	109929	88141	79412	99244	112804	106298
User 4	19351	19637	21960	16808	19793	19795	21443	20090	19600	20805	22541	19153	20303	19708	19188	19623	21147	19394	21816	21438	21108	22184	21141	24612	18472	19205	19531	20706	20732	19302	28532	27882	19178	20229	20465
User 5	35112	38156	29882	36096	33254	38125	36311	34947	36014	39086	35546	37351	37209	36902	36987	38594	33259	39156	35339	36239	38737	33692	31562	28895	38370	38905	36748	32418	35589	37526	29406	21385	32999	38685	34146
User 6	34823	32038	32887	35608	34596	30381	33855	34460	35510	33228	30418	34172	31299	33074	34650	33781	36427	30616	32548	32984	31579	31887	31840	35907	35272	35361	33055	34484	32802	33955	37706	37113	23576	31150	32430
User 7	53940	57650	45359	58163	52592	57485	35760	51691	53985	53319	45357	55877	55571	49650	52524	58680	55508	53253	46805	53666	57517	42240	41376	45047	58408	60559	49801	47201	46082	55966	33872	27313	42794	58553	44237
User 8	34641	35946	31329	36830	34510	35332	33876	35029	35533	36029	30361	35404	34441	34340	35162	35786	33454	35800	31864	34359	34563	28920	28598	34197	30175	36977	34786	33575	32679	35578	33432	31686	28925	35908	30749
User 9	23520	25824	19248	25803	23952	25485	24947	23472	23204	26590	23533	25107	25174	24982	24865	25733	23135	25938	23418	26289	25761	22895	19982	19817	25752	26144	24663	21898	23171	25508	18546	15865	23805	26546	22019
User 10	48920	49344	49335	51277	48153	48179	44530	45787	50466	35992	34417	49339	46984	43681	47054	49928	54286	42708	43500	42709	44739	41952	40725	57088	50663	51848	45100	48313	43845	48336	60580	56010	33051	47735	39707
User 11	145255	148055	136129	151334	144825	146528	146355	137100	147233	147970	81153	147290	147836	143802	146287	147345	148552	144855	138753	144860	147758	137141	131047	151618	147564	150002	142809	142606	142828	146274	151828	137114	112994	147979	133514
User 12	14538	13809	14918	14780	14898	13698	14919	14498	14725	14141	13836	13123	13654	13861	14148	13868	15562	12519	14910	14572	14359	14749	13986	16290	13775	14137	13978	14815	14243	13554	17590	17745	11764	13251	13597
User 13	43000	42379	41915	44564	42083	42263	41800	43608	44201	42748	37423	42984	30297	42430	43074	43815	44197	42871	42165	42821	41589	42558	39707	43638	44228	44797	42030	42070	42635	42503	48536	40244	33857	42185	41995
User 14	22084	24083	20009	24401	22529	23810	22016	21197	22819	23367	17296	23267	23030	21214	22927	24084	24081	23080	21411	22127	22665	20833	19064	23717	24284	24660	21599	20844	21185	23351	23459	22427	17226	24204	18372
User 15	10453	11278	10359	11791	10841	11150	10573	9879	10915	10246	8688	11429	10705	10254	10744	11189	12353	10570	9517	9815	10004	8869	7904	11639	11851	11856	10717	10185	10218	11073	12101	11323	7508	11010	8609
User 16	30710	26851	29806	31460	30842	27117	30631	30164	31517	27862	25574	29195	26994	28906	30851	26473	32971	25728	29210	29703	26804	28035	28590	34668	29437	29940	30108	31530	30226	28433	35941	37194	16512	25160	28307
User 17	68270	70906	62397	70386	68771	70313	72011	69972	70077	73617	78025	68961	70997	70376	68738	70890	60722	72228	71233	72870	72080	68019	65643	62277	70819	71386	69780	67969	68373	70305	60363	64060	71230	73081	70888
User 18	58988	57676	61787	61963	60483	55629	59222	56889	60276	57970	49924	50931	57878	55465	56753	59078	64417	45193	56062	57648	57410	54032	52154	65409	60275	61758	55573	58020	54142	59063	70537	69954	45242	55060	53298
User 19	70688	73845	63142	73719	71315	73407	67694	70110	71391	75210	62705	71805	72498	71333	71713	73481	69088	73635	57753	71979	73589	62969	59964	63000	73441	75526	70424	68299	66348	73367	64005	61005	61329	74308	65151
User 20	53289	53822	47328	56346	52821	52301	46373	48970	54511	49455	39455	53014	49562	47512	52567	55241	57282	50683	44742	40008	47415	42163	41706	55315	56877	57873	49065	47863	45258	52663	47702	41841	33475	52229	42864
User 21	58999	52908	51737	59001	56882	50901	51724	54412	58712	50774	47244	53052	49479	48978	56752	54258	60282	52169	51198	45004	35269	48754	48202	60365	57183	58079	52749	54888	51604	51488	57617	56947	32778	51633	49364
User 22	74761	83017	65452	83123	76400	81997	75784	72810	78450	84558	70253	80855	80397	77594	76906	82994	77413	80695	68440	82504	82585	54434	56600	62383	83339	84788	77802	71828	69745	82440	55647	51580	67053	84391	64810
User 23	75889	79034	69545	78518	76194	78557	78243	76817	77115	81026	76706	77575	78232	77915	77156	78820	72314	79533	75924	79827	80400	73269	63019	65062	76867	76545	77405	74352	75776	78345	69943	62540	73063	80858	75258
User 24	123061	138848	110888	138102	126864	137284	135154	127860	128576	144158	129933	134317	136909	135653	132490	138469	117688	137944	124298	139638	142375	129055	113036	61371	138360	141352	134345	119580	127256	136829	103016	84162	134519	141975	128338
User 25	27173	25980	30477	27186	27499	26127	29963	27803	27660	26212	27756	26078	26207	26719	27111	25532	29786	24413	29103	28891	27608	30434	28364	33368	24297	26386	26996	28810	28842	25727	38849	39502	22377	25373	28483
User 26	34829	31130	38138	34333	35270	31990	37025	34458	35394	31994	33781	33399	31546	33627	34541	31602	37550	29521	36681	34515	32579	36242	34118	42094	33296	29641	34505	36627	35810	32580	46828	47344	25200	30875	32662
User 27	22311	23947	20394	23988	22822	23964	21012	21888	22556	22985	19898	23402	22840	21589	22268	23932	23624	22889	20596	22164	23090	18708	18670	21555	23923	24579	21844	20872	20254	23355	19739	17151	18909	23540	19732
User 28	46423	48149	38675	47935	46810	47697	47140	46171	46309	49740	49068	47305	47972	47485	47181	48416	42963	49882	46244	49049	49008	45137	43083	38948	48021	48710	47069	41210	45791	47702	37150	33822	46557	50032	46097
User 29	53406	57685	47212	57691	53944	57212	54169	52657	54369	58945	53435	56394	55999	53705	54384	57074	52201	56793	50748	57187	57283	47271	44781	45734	57979	59491	54557	50165	44147	57001	43857	39728	50982	58748	49397
User 30	29076	28515	30079	29590	29308	28290	29484	29012	29446	28171	27621	28659	28110	28363	28859	28778	30655	26884	29792	28913	27969	29588	28419	32893	28790	29353	28535	29439	29129	27201	33952	33978	24978	27945	28509
User 31	189331	227288	140118	222925	193063	226766	205856	188831	193891	231264	208463	216084	223952	213944	210402	226390	189543	221558	185747	222157	221094	186375	153048	111688	226753	233893	212113	179702	181288	219385	10292	112140	207080	234550	169291
User 32	232940	249409	195418	248307	229930	248211	232987	232390	234211	247148	232834	239644	243154	235333	238833	251559	213187	257210	228160	248651	251200	233054	210650	188274	247553	253008	238607	215989	226284	238173	181526	40583	244013	258473	232712
User 33	182815	183358	176350	185855	182146	184066	182592	182370	182959	188197	173004	183832	183264	183258	183242	184858	163969	185354	180753	183381	187231	178168	170704	181433	186237	187407	180714	180399	182717	185723	180155	162741	114175	186551	177029
User 34	87607	85037	86694	88098	86998	84407	86843	87340	88276	85725	80776	86823	83492	86000	87385	85947	90448	81733	86586	84321	82458	82878	81873	91393	86801	88930	86398	88261	85494	85657	94514	92567	84292	57805	83799
User 35	81870	87552	72548	87982	82985	86986	85700	81731	84437	88550	76937	86143	86395	85002	84941	87442	82630	85694	80684	86434	86126	72493	66190	74232	87620	88772	85338	81391	78852	86955	71530	72915	75458	87455	80748
total	3	12	26	2	10	10	12	10	6	13	29	1	14	3	0	8	17	17	23	15	14	23	25	22	5	8	4	15	17	8	0	19	31	21	25
FAR	8.57	34.29	74.29	5.71	28.57	28.57	34.29	28.57	17.14	37.14	82.86	2.86	40.00	8.57	0.00	22.86	48.57	48.57	65.71	42.86	40.00	65.71	71.43	62.86	14.29	22.86	11.43	42.86	48.57	22.86	0.00	54.29	88.57	60.00	71.43

35 Users
1190 Impostors

Accept Count	468
FRR	0% Fixed
FAR	39.3%

0.6 SD	Reference Profiles			Trigraph																													
	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17	User 18	User 19	User 20	User 21	User 22	User 23	User 24	User 25	User 26	User 27	User 28	User 29	User 30	User 31	User 32	User 33	User 34	User 35
User 1	20104	24663	18520	24482	22050	24485	23367	22253	22537	24756	22983	23793	23930	23613	23498	24669	19294	24193	21499	24734	24420	21630	17302	14867	24545	24848	23716	20438	22500	24279	18387	13181	22218	24956	21319
User 2	37548	31875	35643	38824	37205	31699	35967	36001	38520	31866	26091	36011	31598	35096	36944	33725	40068	29640	33923	33145	31155	33021	32015	39970	36281	36545	35702	30859	33981	34237	41149	41192	20468	30310	32099
User 3	104093	110888	52369	109949	105282	109002	106364	106202	107089	111908	103944	108123	108194	108114	108022	110243	69103	110403	104001	111275	110023	95775	88894	79128	110483	111881	108392	98038	102430	109841	74997	67144	62176	111220	101160
User 4	19351	19501	21772	18242	19785	18575	21315	20878	19600	20225	22013	19063	20139	19688	19188	19459	21095	18814	21796	21342	20932	21968	20917	24364	18128	18783	19523	20694	20868	19166	26196	27630	17954	19993	22261
User 5	34828	37038	28334	38096	32242	37769	35331	34315	35554	38530	33203	37343	36881	36226	36811	36394	31979	38584	33839	37515	38013	31588	29462	26355	38302	38845	36428	31414	34749	37350	28102	17037	29083	37841	32122
User 6	34531	31298	31591	35560	34304	26997	32623	34016	35472	31974	27206	33796	30259	32722	34462	32849	36079	28824	31556	31732	30047	29031	30276	34371	34824	34789	32623	33782	32122	33315	35606	34817	20588	29530	30986
User 7	53018	56870	41863	58143	51720	58141	31914	49431	53149	50531	40385	55353	53395	48542	52320	58256	54188	51481	44173	50390	55505	37320	38984	40983	58178	60359	49457	45813	44206	55158	26492	19657	34520	56877	40825
User 8	34485	35662	29621	36826	34150	34780	32496	30673	35425	35017	27273	35358	33749	33968	35122	35446	32638	34752	30348	33043	33475	26848	27176	32093	35987	36725	34600	32623	31727	35354	30488	28618	23673	34944	28849
User 9	23356	25744	17750	25599	23736	25313	24287	23184	22680	26354	22225	25087	24822	24454	24481	25625	22271	25534	22574	25905	25389	21251	18690	18137	25728	26104	24407	21320	22403	25472	16094	12793	22045	26214	20379
User 10	48085	48208	47419	51133	47881	48831	42210	44159	49830	31464	29285	48947	45784	43241	46834	49132	53838	40326	41708	39909	42599	39378	39049	55112	50099	51116	44720	45133	42445	47448	56832	51362	27803	45811	37327
User 11	144523	146715	129897	151142	143453	145020	143475	134248	146221	144724	57049	147166	146036	141822	145915	146221	145592	141383	131165	141040	145004	130505	124987	145428	146338	146822	141857	139674	140004	144808	141408	124388	100626	144787	125382
User 12	14458	13497	14494	14660	14854	13298	14591	14250	14689	13669	12584	12739	13214	13685	14088	13472	15498	11883	14674	14184	13827	14321	13528	15786	13563	13825	13858	14687	13983	13302	18926	17105	10620	12675	13137
User 13	42868	41591	40303	44508	41779	41383	40628	43000	43993	41408	34331	42884	39913	41786	42990	43275	43589	41743	40985	41317	40585	40942	38311	41772	43810	44425	41558	42059	41987	41707	43852	36536	29029	40781	40347
User 14	21952	23883	19309	24385	22209	23366	21196	20509	22455	22755	14912	23203	22514	20506	22799	23788	23633	22332	20271	21239	21953	18377	18252	22469	24112	24508	21307	20384	20421	23175	21847	20219	14618	23668	16740
User 15	10377	11122	9535	11767	10673	10930	10301	9539	10863	9590	5436	11413	10473	9902	10624	10933	12153	9970	9073	9259	9356	8205	7520	11039	11527	11508	10529	10005	9770	11029	11205	10157	6424	10622	7941
User 16	30658	26007	28286	31340	30630	26217	29619	29832	31493	26664	22874	29027	28238	28866	30843	25281	32675	23952	27994	28519	25292	29659	27146	33590	28633	29100	29676	31070	29358	28213	33901	35068	13668	23568	27007
User 17	67910	70830	57749	70374	68387	70089	71407	69148	69845	73337	75081	68953	70581	70216	68722	70702	45822	71652	70181	72270	71448	65543	61607	57689	70755	71326	69572	68629	66969	70273	65015	58080	67442	72545	68432
User 18	58192	56356	60315	61723	59815	53930	57518	55777	59620	55626	45732	59519	56382	54729	56405	57846	64057	38453	54622	55758	55254	51808	50274	63291	59399	60858	54853	56940	53056	57905	67661	66262	39242	52684	51398
User 19	70452	73385	59318	73715	70739	72583	65590	68138	71143	74110	57603	71713	71214	70589	71677	73113	67324	72319	62865	70171	71913	57817	58524	58472	73241	75318	70112	67075	64670	73235	57629	53201	55213	72948	61071
User 20	52733	52854	44382	56306	52177	51017	43657	47554	53891	46503	33159	52350	48078	46716	52267	54197	56534	48427	42110	35584	43943	38595	39370	51755	56333	57353	48489	46015	43306	51759	41986	34929	27827	49925	38980
User 21	56899	51384	48697	58925	56342	49149	50100	53088	58508	48258	42448	52436	47751	47958	56448	52644	59606	49425	48560	41874	26216	45874	45506	58153	55775	57100	51755	53418	49932	50284	53301	53943	27884	49063	46724
User 22	74365	82445	59642	83083	75180	81017	73348	70938	78022	83158	64829	80699	78913	75246	76246	82438	74901	78327	65400	79880	80830	43770	52480	55375	83151	84436	78366	70392	65883	82100	46215	42780	69121	82859	58886
User 23	75477	78906	65637	78470	75670	78333	76895	75777	76815	80454	73898	77523	77496	77451	77016	78678	70222	78529	73960	78943	79428	69389	66527	60534	78723	79441	77087	73224	74500	78233	63859	55740	67559	70066	71652
User 24	121677	138672	102846	138046	125580	136540	131766	126460	126580	143266	124209	134213	135213	133689	131582	137037	111956	135868	118882	137962	140579	121271	104128	44170	136236	141208	133509	115944	123244	136689	66324	67326	125823	140307	120730
User 25	27157	25484	30045	27038	27335	25595	29603	27463	27652	25328	26420	25990	25743	26355	27103	25080	29670	23321	28603	28335	26936	29978	27804	33048	23267	25738	26840	28626	28550	25511	38321	39052	20581	24349	27823
User 26	34789	30468	37280	33937	35118	31090	36381	34180	35390	30838	31389	33207	30918	33399	34501	30754	37366	28285	35873	33695	31571	35226	32930	41442	32356	27613	34289	38247	35082	32268	45512	46312	22940	29447	31570
User 27	22127	23747	18994	23952	22508	23600	20100	21432	22376	22337	18034	23338	22404	21380	22240	23732	23252	21977	19744	21144	22330	17372	17958	19943	23823	24463	21560	20504	19734	23203	17467	14709	10697	22776	18518
User 28	48199	48021	35607	47911	46218	47349	46180	45587	46061	49384	48464	47371	47580	47181	47097	48324	41971	49102	44940	48521	49244	43045	40612	35696	47049	48630	46841	39271	44763	47646	32022	27602	43633	49452	44141
User 29	53050	57297	43628	57071	53308	58478	52549	51441	53833	57733	50111	56230	54951	52629	54048	57614	50819	55205	48852	55863	56207	44303	42333	42030	57855	59303	53953	48941	40903	56885	37867	33598	46506	57876	46025
User 30	29028	28219	29303	29498	29220	27878	28984	28668	29374	27535	25957	28531	27674	28155	28799	28406	30515	26084	29228	28353	27049	28672	27695	32189	28510	29077	28303	29075	28777	28485	32720	32498	23148	27153	27601
User 31	168511	226986	125226	222881	189319	225728	194956	184475	187610	227828	195104	215784	219308	205624	206354	225782	179603	214478	174727	216841	217402	145523	138718	91764	226617	233693	206629	173458	170732	219169	66	84468	188600	231698	149329
User 32	231378	248557	181482	246321	228686	246807	225011	229218	232639	243808	215974	239432	239448	233217	238093	251068	206559	253934	221124	242935	247388	220686	194182	169610	248963	252512	237487	208949	220472	235457	151648	52	224073	253045	222492
User 33	182215	182610	170298	185599	181118	182854	180356	181406	182375	188185	165136	183724	181888	182306	183150	183692	181873	182418	177561	181089	185523	172954	166348	175721	185781	186811	180186	178819	181125	185367	171383	149385	26539	183087	173453
User 34	87523	84269	84246	87986	86790	83427	85467	86960	88248	84253	76608	86283	82564	85724	87233	85043	90008	79513	85478	82617	80598	80526	79925	69337	86345	86288	86202	87477	84566	85213	91774	88575	58316	48886	82155
User 35	81382	87172	68272	87866	82253	86454	83820	86547	83777	87198	70837	85999	85319	83782	84549	87094	80210	83966	77858	84566	84622	87041	81858	69360	87368	86472	84454	70947	76000	86727	62499	65555	69234	86007	58600
total	3	12	22	2	10	10	10	6	11	23	1	14	3	0	7	17	15	22	14	10	20	25	19	3	8	4	15	15	8	0	0	29	17	23
FAR	8.57	34.29	62.86	5.71	28.57	28.57	28.57	28.57	17.14	31.43	65.71	2.86	40.00	8.57	0.00	20.00	48.57	42.86	62.86	40.00	28.57	57.14	71.43	54.29	8.57	22.86	11.43	42.86	42.86	22.86	0.00	0.00	82.86	48.57	65.71

35 Users
1190 Impostors

Accept Count 406
FRR 0% Fixed
FAR 34.3%

0.6 SD	Reference Profiles				Trigraph																											
Unmatche	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17	User 18	User 19	User 20	User 21	User 22	User 23	User 24	User 25	User 26	User 27	User 28	User 29	User 30	User 31	User 32	User 33	User 34	User 35
User 1	50	2840	210	2250	134	2792	2792	630	210	4026	6076	964	2478	1530	626	2776	54	3456	1286	4082	3756	2962	68	60	2524	3142	1556	54	1118	2038	3132	306	5382	4470	2200
User 2	1750	50	7108	3770	2308	54	5358	2142	3330	1190	1274	52	50	394	626	116	7594	58	2802	2136	58	4216	800	13994	2060	3372	196	3900	1222	54	21984	22816	58	66	2448
User 3	60	6524	56	5914	1884	7820	9440	4576	3162	15636	26924	2448	6886	4712	2104	8114	50	15946	10570	14246	15810	7624	52	292	7370	10250	3842	106	3296	5782	8482	7220	13020	16892	10354
User 4	1276	2508	7350	54	2228	2930	6336	4034	1774	5534	9668	832	3524	2202	950	2636	4688	3432	6290	5774	5282	7614	5212	12406	942	2316	1632	4022	4358	1512	17014	10014	4882	4468	5106
User 5	50	3352	52	2702	56	3508	2716	66	54	6084	9510	1262	2626	1060	686	4164	142	6272	1506	4846	5334	1286	50	58	3476	4870	744	54	700	1920	2974	60	3274	5090	544
User 6	748	50	4534	2418	1374	52	3200	1862	2174	2260	3202	50	60	56	406	1136	5296	64	2150	2258	1180	2602	478	9812	2480	3436	52	2446	868	284	16258	17176	52	380	1760
User 7	94	6924	222	6442	176	6776	50	116	518	6022	4446	1614	4254	50	50	8412	5900	5730	50	5216	9186	52	52	3606	7296	11890	50	50	50	2904	488	50	50	10926	50
User 8	158	2718	1666	3522	510	2174	2538	58	1460	5128	3490	710	1528	544	402	2498	1776	4108	814	3068	3130	50	50	8032	2740	4388	468	932	90	1598	10668	10648	60	4978	218
User 9	50	2276	50	1678	54	2060	2354	56	52	4436	4838	794	2048	736	136	2214	524	3640	1296	4144	3668	834	56	52	2048	2900	414	52	64	1568	550	80	3420	4020	62
User 10	744	3634	10048	5218	1822	2780	2970	232	3258	60	76	1334	2172	54	50	4826	11690	132	1324	246	1666	2806	68	23270	4848	7878	50	1578	160	1062	34378	28724	56	4874	68
User 11	1768	9418	13420	11778	2896	8226	15338	118	6592	18050	50	3288	11026	2936	1700	8744	20806	11292	6872	13478	14410	13748	2408	37352	6530	12790	490	6450	6518	5628	61852	46594	54	17512	7246
User 12	862	626	3482	1462	1842	574	3002	1454	1228	2184	3318	60	506	370	348	658	3030	62	2702	2458	2018	3384	1568	8406	520	1250	336	2028	1504	54	9646	10100	426	780	1612
User 13	380	1450	5182	3124	544	1848	3284	2844	2878	4064	4298	160	52	676	182	3230	5238	3954	3908	3658	2338	6196	1080	10330	2996	4734	56	2074	2688	234	16284	14716	60	3448	4150
User 14	64	2744	2238	2696	304	2438	1868	66	708	3260	612	584	1910	52	152	2638	3940	3034	1292	1620	2168	1678	110	6228	2742	3702	54	160	196	1516	9816	9468	52	4274	60
User 15	50	1008	1244	1242	222	892	1156	64	274	526	58	530	514	50	50	818	2594	848	100	300	284	184	52	3574	1302	1556	94	158	92	514	5494	5214	52	1338	58
User 16	1196	50	4948	2578	2160	56	4846	2284	2446	1798	2472	50	72	290	1058	50	6086	70	2976	3678	472	3752	1814	13294	548	2168	648	4214	2644	52	18650	21048	56	66	2030
User 17	1086	6042	2428	4854	2228	5732	9692	4914	4364	12232	27424	1808	7084	5098	1442	5982	54	9728	9738	10552	9508	8120	3140	5784	5728	6834	3874	3408	4144	4568	19324	18166	15464	11984	11314
User 18	1166	2322	14428	9092	4204	1222	8148	1784	3492	5966	4320	1918	3140	176	50	4302	11036	50	4840	5120	6058	5310	424	21112	4484	8118	50	3972	776	2324	34852	38078	134	3658	1764
User 19	50	5362	2454	4158	1468	5278	2550	808	992	10716	6590	610	4316	1704	434	5048	2680	9112	52	7388	6964	84	52	5532	4482	8516	54	1208	52	3806	14810	13850	2498	10612	700
User 20	1082	4058	3386	5752	1946	2788	512	62	3022	2744	64	292	98	50	232	5858	10040	3068	66	54	704	92	50	15506	7058	9804	50	122	52	1126	14884	9414	50	4582	60
User 21	2942	664	7158	6422	4098	98	3038	2480	5732	672	3722	50	58	50	2284	2166	10864	3498	2254	70	50	2504	104	19058	4414	6922	60	4374	354	50	22820	26462	54	2586	1050
User 22	52	8092	64	7332	60	7284	5684	52	50	15282	11718	2924	5530	58	50	8080	5624	9286	56	11990	11410	50	62	680	8244	11814	52	58	50	6298	5164	1490	2124	14942	62
User 23	58	4408	3546	3076	472	4372	6726	2272	1370	9876	18894	1320	4336	2480	622	4128	1650	8222	5688	6890	9478	4928	50	2708	3994	5498	1690	1192	2350	2914	13718	10020	9206	11248	6744
User 24	50	14354	82	12390	52	13490	16856	3716	52	28764	28430	5060	12916	7928	2134	14228	74	19694	4870	22296	24760	15344	50	54	13210	19250	5680	54	2540	9944	7200	72	27840	25452	10114
User 25	1142	1234	9198	1560	2150	1602	7878	3148	2044	3508	6888	50	1878	1176	690	866	6376	776	6426	6562	4740	9502	5384	14956	64	2424	1208	4780	5058	64	25714	27252	1080	1800	6186
User 26	1794	60	10860	1744	2944	510	8530	3108	2920	2878	8414	110	644	1378	1258	118	7468	284	7854	5478	2722	6392	4080	18088	1244	50	1638	5766	6444	98	28708	30572	126	1092	3704
User 27	54	2190	802	1744	462	2348	336	52	64	2490	1732	712	1174	56	50	2276	2564	2038	60	1008	2056	52	56	3098	1988	3324	52	54	52	084	4746	2290	62	2868	58
User 28	50	2670	64	1854	308	2948	3824	632	90	7324	12692	722	2742	1246	398	3032	64	7176	3872	6274	7348	3760	52	270	2178	3668	526	52	856	1458	3394	2588	6782	7768	3518
User 29	60	4938	182	4682	84	4516	4278	62	134	9558	10146	1646	3254	50	50	5328	1380	6546	770	7170	6822	148	50	710	4910	7886	54	54	52	3018	5854	2784	4456	9228	352
User 30	652	958	6440	1820	1306	1120	3476	1472	1440	2062	4836	68	632	496	250	1284	4314	696	4132	2694	2142	5348	2682	10994	1030	2088	270	2698	2114	54	16116	16228	1372	1998	2960
User 31	50	32740	50	23130	50	33076	21052	60	52	54820	59512	9782	28068	6732	74	32038	68	35008	102	41230	35016	52	50	50	31070	45590	5518	50	50	19118	52	50	33706	53604	58
User 32	58	26088	50	18408	54	26504	21200	7234	1190	40450	53198	5514	22202	5216	3666	29774	50	47788	16318	38060	38262	29354	54	106	21848	32774	4972	52	4540	8418	15380	50	56436	48152	22162
User 33	76	4682	12406	5342	1748	6552	8420	3586	1438	15078	19558	1020	6448	3114	166	6884	11248	14018	10022	8650	13826	10490	260	24990	6466	9926	50	3054	4966	5210	38202	33994	68	19642	5906
User 34	1182	54	7976	2280	1324	54	4806	2584	2348	3170	4358	50	50	262	600	538	8580	186	5030	1796	502	4394	618	16892	1004	2422	196	4424	1414	50	27272	28014	52	50	2888
User 35	62	5576	1284	5340	108	6180	6228	290	1126	11548	12660	1838	5258	1500	224	5354	5058	6782	3230	8890	7458	296	58	6352	5442	7976	1588	1110	54	4076	15702	16916	4042	10694	54
total	7	2	4	0	4	0	0	3	3	0	0	6	2	5	6	0	3	0	1	0	0	1	6	2	0	0	7	4	6	5	0	2	14	0	1
FAR	20.00	5.71	11.43	0.00	11.43	0.00	0.00	8.57	8.57	0.00	0.00	17.14	5.71	14.29	17.14	0.00	8.57	0.00	2.86	0.00	0.00	2.86	17.14	5.71	0.00	0.00	20.00	11.43	17.14	14.29	0.00	5.71	40.00	0.00	2.86

35 Users
1190 Impostors

Accept Count 94
FRR 0% Fixed
FAR 7.0%

0.7 SD	Reference Profiles			Trigraph																													
	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17	User 18	User 19	User 20	User 21	User 22	User 23	User 24	User 25	User 26	User 27	User 28	User 29	User 30	User 31	User 32	User 33	User 34	User 35
User 1	10048	24843	17204	24482	21742	24217	22783	21795	22149	24536	22015	23741	23294	22945	23334	24609	18398	23921	20675	24342	24148	20034	15854	13095	24517	24826	23066	19982	21708	24251	16511	10997	19818	24852	19259
User 2	37372	30807	34159	38580	36945	36943	34863	35425	38472	30666	22127	35615	30698	34764	36816	32581	39806	27744	32971	31901	29703	31369	30799	38718	35293	35869	35170	35999	33037	33785	38973	38556	17884	28384	30959
User 3	103121	110836	41197	109893	103568	108242	103928	103834	106373	111016	97308	108043	106766	106834	107854	109799	86931	108343	100285	110147	108895	90531	82590	73242	110383	111689	107040	94146	100134	109769	64417	56168	84796	109512	95272
User 4	19351	19341	21564	15700	19749	19331	21227	20666	19800	19785	21453	18921	19883	19680	19168	19299	21063	18314	21780	21214	20748	21748	20789	24052	17872	18433	18519	20674	20624	18958	25820	27386	17030	19753	20045
User 5	34584	37728	26922	38068	31200	37413	34527	33823	35214	37902	30735	37327	38141	35362	38591	38010	31012	37628	32491	36571	37269	29672	27694	23783	38222	38737	38044	30770	33853	37182	22978	13493	24759	36857	29999
User 6	34427	30604	30187	35504	34116	27741	31571	33272	35428	30646	24026	33448	29551	32370	34278	31865	35715	27068	30680	30456	28515	28187	29080	32911	34336	34165	32303	32976	31534	32695	33482	32914	17981	28014	29702
User 7	52176	55458	38703	58131	51188	54441	28914	47547	52393	48055	35585	54901	50639	47702	52092	57660	52998	49821	41543	47148	53245	33380	37517	37019	57872	60071	40237	44737	42482	54046	19960	12428	27940	65101	37589
User 8	34225	35346	28069	36798	33794	34284	31632	29206	35249	34045	24037	35288	33281	33580	35066	35054	31758	33720	28728	31819	32375	25240	25828	30088	35791	36477	34370	31647	30959	35162	27976	25648	21013	33912	27065
User 9	23276	25648	18202	25595	23500	25137	23731	22832	22272	26062	20581	25051	24450	23878	24341	25501	21487	25038	21746	25485	24905	19835	17694	16737	25720	26076	24083	20800	21479	25448	13538	10101	20001	25830	18855
User 10	47305	46908	45335	50961	47461	45547	40278	42907	49314	27704	24849	48423	44416	42857	46578	48224	53290	38118	40196	37325	40380	37292	37829	52996	49403	50344	44404	44225	41425	46324	53120	46586	23803	43787	35359
User 11	143739	145563	123829	150878	142521	143764	140819	131588	145207	141172	35259	148994	144378	139662	145523	145245	142804	137599	125857	137456	142250	124445	119175	139524	145032	147626	140801	136970	137408	143446	131128	112068	89318	141443	117030
User 12	14358	13157	14022	14588	14782	12930	14291	14062	14625	13153	11641	12307	12850	13585	14052	13120	15428	11351	14442	13824	13515	13925	13078	15346	13423	13509	13786	14471	13731	13054	16458	16489	8824	12119	12697
User 13	42756	40571	38747	44488	41523	40471	39524	42152	43781	40080	31147	42788	31625	41248	42926	42667	42963	40411	39809	40017	39513	39390	37143	40000	43288	43957	41250	41483	41263	40923	41272	32864	24517	39417	38907
User 14	21876	23663	18093	24369	21945	22906	20428	19953	22139	22095	12596	23147	22034	19782	22675	23540	23293	21432	19291	20407	21217	18117	17508	21429	23924	24288	21015	19984	19649	22955	19747	18095	12582	23128	15296
User 15	10317	10970	8655	11743	10557	10694	9925	9275	10831	9078	4498	11401	10277	9738	10572	10721	11925	9398	8609	8875	8692	7925	7156	10539	11471	11318	10313	9821	9554	10949	10389	8871	5096	10134	7421
User 16	30598	25307	26930	31224	30414	25409	28783	29604	31457	25440	20120	29007	25842	28398	30778	24265	32355	22384	26998	27399	23720	25339	25754	32288	27829	28258	29272	30442	28486	28081	32091	32952	11588	22160	25903
User 17	67626	70746	53285	70354	67971	69889	70799	67092	69821	73081	71289	68921	70073	70032	69874	70514	41670	70992	66841	71834	70872	63031	56723	53597	70679	71234	69172	64817	65597	70229	60675	51908	63114	71905	65492
User 18	57428	55052	50653	61455	59007	52438	55878	54709	59188	53338	42124	58983	54658	54233	56125	56830	63681	34209	53282	53810	53494	49572	48570	61271	58367	59910	54349	55992	52090	56857	64833	62214	34226	50388	49522
User 19	70188	72833	55530	73691	70359	71615	63742	66122	70847	72802	52515	71609	69998	70025	71641	72605	65944	70815	48665	68171	70101	53145	53752	54024	73037	75090	69776	65093	63178	73067	51617	46017	49877	71560	57639
User 20	52141	51690	41460	56286	51737	50025	41881	46170	53367	43635	27399	51770	45766	48060	51839	53149	55550	46051	39650	31604	40471	35379	37522	48259	55065	56549	47681	44455	41714	50835	38776	28885	23131	47697	35478
User 21	56459	49964	45741	58817	56062	47805	48880	51772	58320	46142	37680	51800	46511	47130	56128	51028	58858	46483	48006	38954	24036	43306	43280	55861	54357	55951	50767	51718	48418	49524	48957	48575	24092	46873	44088
User 22	73977	81873	64050	83031	74376	79949	71079	68934	75590	61038	59169	80503	77725	73878	75702	81742	72285	75499	62272	77108	78882	34542	49405	49091	82899	84048	74874	69280	62759	81558	37967	35158	50273	80495	53274
User 23	75197	78786	62081	78422	75126	78089	75555	74829	76603	79822	69246	77439	76840	76807	76888	78512	68142	77441	71916	77959	78480	65545	51443	56130	78663	78281	76599	72280	72956	78141	58219	49544	61807	79026	67620
User 24	121045	138472	94790	137078	124080	135636	127570	124860	125008	142062	118473	134073	133253	129845	130382	137421	107328	133052	114688	136154	138863	112803	94832	30583	138120	141004	131677	112584	118580	136533	71092	52330	114399	138683	112490
User 25	27129	24916	29557	26826	27187	25091	29223	27279	27644	24408	25020	25922	25275	26059	27087	24598	29568	22373	27999	27607	28188	29374	27116	32692	22421	25188	26568	28386	28262	25295	37657	38520	18965	23389	27063
User 26	34577	29968	38416	33625	34990	30350	35845	34004	35386	29854	28901	32955	30442	33275	34425	30114	37206	27177	35061	32747	30563	34206	31870	40706	31484	25589	34025	35731	34278	31916	44224	45212	20824	28175	30562
User 27	21995	23487	17678	23940	22134	23160	19300	21024	22212	21381	18054	23234	22120	21233	22212	23556	22840	21053	19116	20232	21506	18192	17390	18879	23887	24387	21284	20144	19268	23019	15299	12749	14205	22124	17352
User 28	48031	47901	32735	47891	45818	47005	45129	45011	45785	48920	43768	47343	46938	46861	47025	48224	41159	48318	43620	47905	48392	41077	38268	32712	47877	48506	46565	37923	43787	47590	27347	21850	40293	48780	41885
User 29	52722	56785	40384	57051	52648	55828	51089	50125	53409	58597	46811	56002	54047	51837	53812	57202	49671	53709	46616	54427	55047	41391	40243	38978	57691	59143	53269	48031	38471	56313	32407	28102	41814	56892	42853
User 30	28960	27943	28483	29434	29160	27478	28580	28340	29338	26815	24253	28431	27282	27931	28747	28030	30363	25220	28636	27697	26193	27812	27019	31565	28188	28753	28075	28723	28461	25813	31392	31162	21328	28321	26781
User 31	188003	226422	112682	222861	184443	223628	185280	179999	185027	224148	180624	215552	213628	198268	203034	225246	171047	207092	164523	210141	213270	123971	128924	78188	226397	233493	198777	169154	156740	216957	55	59304	168848	228702	129777
User 32	229756	247333	167522	246277	223202	245235	216295	225794	231131	240756	197274	239212	235706	230537	237425	250412	201219	249006	214872	236375	243184	207814	179662	153462	246185	252020	236323	203269	213800	234861	123622	52	199278	249505	210900
User 33	182011	182006	164814	185399	180294	181882	178124	180618	181851	183893	156832	183644	180428	181322	183054	182738	176793	179258	174409	178853	183639	168178	162844	159937	185293	186151	179890	177391	179521	184811	162843	136377	48699	179483	166893
User 34	87427	83725	81918	87870	86614	82531	84119	86472	88228	82941	72650	86079	81744	85540	87029	84227	89482	77177	84478	81425	76786	78568	78561	87497	85893	85634	85988	86857	83790	84833	88950	84967	53544	38100	80611
User 35	81146	86732	84236	87770	81561	85742	81948	79211	83185	85878	64829	85863	84223	82458	84217	88722	78138	82142	75188	82766	83070	82013	58582	85108	87060	88132	83350	78787	73480	88507	54170	58507	65142	84591	50880
total	3	12	19	2	10	10	9	9	6	10	18	1	13	3	0	5	16	14	22	11	6	16	24	9	3	7	4	15	15	8	0	0	23	15	23
FAR	8.57	34.29	54.29	5.71	28.57	28.57	25.71	25.71	17.14	28.57	51.43	2.86	37.14	8.57	0.00	14.29	45.71	40.00	62.86	31.43	17.14	45.71	68.57	25.71	8.57	20.00	11.43	42.86	42.86	22.86	0.00	0.00	65.71	42.86	65.71
35 Users																																			
1190 Impostors																																			
Accept Count	361																																		
FRR	0% Fixed																																		
FAR	30.3%																																		

0.7 SD Reference Profiles Trigraph

Unmatche	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17	User 18	User 19	User 20	User 21	User 22	User 23	User 24	User 25	User 26	User 27	User 28	User 29	User 30	User 31	User 32	User 33	User 34	User 35
User 1	50	2620	64	2250	64	2524	2212	232	60	3606	5108	912	1844	874	466	2716	54	3184	522	3890	3484	1388	52	56	2496	3122	952	54	348	2008	1256	54	2996	4166	398
User 2	1574	50	5624	3706	2048	54	4258	1568	3282	78	84	50	50	126	498	52	7334	54	1658	892	54	2570	94	12742	1072	2496	60	2944	380	50	19788	20180	50	56	726
User 3	54	8272	52	5858	64	7060	7004	2208	2446	14744	20288	2368	5458	3432	2036	7670	50	13886	6870	13118	14084	2380	50	50	7270	10058	3390	52	1012	5710	68	60	6640	14984	4476
User 4	1276	2350	7142	54	2192	2686	6248	4022	1774	5094	9108	700	3272	2194	950	2476	4856	2932	6254	5646	5078	7394	5064	12094	686	1956	1628	4002	4294	1304	16638	18770	3958	4228	4890
User 5	50	3144	60	2764	52	3150	1912	58	54	5456	4044	1246	2086	198	466	3782	94	5336	318	3902	4590	82	50	58	3396	4562	360	50	54	1732	182	52	136	5006	54
User 6	644	60	3170	2362	1188	52	2148	1120	2130	934	224	50	56	56	224	238	4932	54	1274	982	162	858	78	6352	1872	2814	52	1630	278	50	14134	14976	52	54	476
User 7	50	5516	60	6430	114	5078	50	52	82	3554	64	1164	1506	50	50	7816	4708	4078	50	1984	6926	52	52	158	6992	11402	50	50	50	1796	52	50	50	9150	50
User 8	74	2402	308	3494	190	1678	1676	52	1286	4156	380	642	1060	188	346	2106	896	3076	54	1864	2030	50	50	6026	2544	4140	248	120	52	1406	8156	7674	58	3646	50
User 9	50	2180	50	1674	50	1904	1788	52	50	4144	3194	758	1676	170	72	2090	70	3144	504	3724	3182	80	56	50	2040	2872	144	62	50	1544	76	54	1384	4236	52
User 10	224	2334	7964	5044	1402	1498	1074	52	2742	58	70	810	814	54	50	3918	11142	62	146	60	58	742	56	21154	4152	7106	50	770	68	56	30684	23948	52	2650	64
User 11	1010	8264	7354	11514	1964	6900	12482	58	5668	14498	50	3114	9366	650	1308	7768	18018	7508	1662	9898	11570	7690	50	31448	7226	11594	68	3748	3920	4270	51572	34274	54	14170	108
User 12	762	290	3010	1390	1570	220	2714	1266	1164	1876	2376	52	164	282	312	312	2958	56	2474	2126	1606	2992	1116	5966	380	934	272	1816	1256	54	9178	9484	50	230	1176
User 13	272	430	3626	3104	338	754	2180	1996	2470	2736	1118	90	52	182	122	2622	4640	2822	2744	2358	1286	4846	256	6562	2472	4266	54	1506	1962	54	15684	10856	52	2082	2710
User 14	64	2524	1068	2680	150	1976	1100	62	392	2600	52	528	1430	52	60	2390	3600	2138	312	768	1434	476	64	5188	2554	3462	54	116	72	1296	7916	7364	50	3734	50
User 15	60	856	386	1216	106	658	780	50	242	82	56	516	318	50	50	610	2366	166	60	102	60	112	52	3074	1246	1364	70	98	50	434	4658	3918	62	650	52
User 16	1136	60	3592	2462	1944	52	4010	2036	2410	574	124	50	66	100	894	50	5768	70	1982	2558	78	2432	286	11990	56	1324	258	3586	1772	50	16844	18914	56	62	928
User 17	606	5956	56	4634	1812	6532	9084	3758	4140	11956	23812	1776	6576	4914	1394	5794	52	9068	8398	10116	8934	5606	50	1672	5652	6742	3474	1818	2774	4524	14984	9994	11136	11344	8374
User 18	484	1030	11768	5824	3396	68	6810	620	3060	3692	790	1380	1420	50	50	3086	11260	50	3480	3178	4298	3288	108	19092	3452	7170	50	3024	406	1076	32024	32030	50	1382	272
User 19	60	4810	56	4134	1090	4310	758	50	696	9408	1512	506	3100	1154	398	4538	1304	7608	52	5388	7152	52	60	1278	4258	8288	52	168	52	3038	6798	6684	50	9224	56
User 20	546	2892	658	5732	1522	1796	124	52	2500	72	54	50	60	50	80	4610	9058	724	52	54	50	58	50	12010	6390	9000	50	98	50	202	9674	4348	60	2354	50
User 21	2710	52	4216	6314	3822	52	1656	1212	5544	50	96	50	52	50	1964	554	10116	536	158	50	50	606	88	16786	2996	5770	52	2678	62	50	18476	21094	64	222	78
User 22	52	7520	52	7280	58	6216	3404	52	50	13102	6056	2728	4342	50	50	7384	3016	6458	58	9222	9462	50	62	68	7992	11426	50	54	50	5754	54	64	64	12780	62
User 23	62	4288	874	3028	162	4108	5386	1332	1158	9244	14242	1238	3480	1838	474	3964	398	7138	3654	7708	8528	1150	50	768	3934	6318	1202	424	902	2822	8086	4240	3534	10310	2738
User 24	50	14154	52	12322	52	12588	12460	2146	50	27560	22694	4920	10956	4104	954	13712	56	16878	482	20488	22864	6902	50	54	13094	19046	3846	54	64	9788	50	56	16216	23828	1928
User 25	1114	666	6710	1348	2002	1098	7498	2962	2036	2588	7468	50	1410	880	674	402	6272	134	5822	6034	3992	6988	4696	14600	58	1872	936	4540	4768	60	25250	26720	338	842	5426
User 26	1682	66	9998	1432	2816	80	7094	2932	2916	1892	5952	64	252	1254	1182	50	7308	84	6842	4530	1714	6372	3022	17352	360	50	1376	5250	4640	50	27480	29472	68	116	2764
User 27	54	1932	84	1732	168	1906	68	52	52	1544	68	606	904	52	50	2100	2154	1114	54	174	1276	52	56	1838	1850	3248	52	54	50	800	2578	334	54	2236	54
User 28	50	2550	52	1634	50	2302	2774	126	50	6660	9996	694	2098	626	326	2932	56	6392	2552	5658	6496	1792	50	50	2106	3544	250	52	88	1400	76	50	5442	7098	1262
User 29	64	4426	72	4642	66	3668	2818	50	70	8422	6846	1418	2350	50	50	4916	280	5050	60	5734	5462	54	50	72	4746	7726	50	54	50	2646	702	78	396	6244	56
User 30	586	682	5620	1758	1246	722	3052	1144	1404	1344	3132	52	242	272	196	912	4162	58	3540	2038	1286	4488	2006	10370	708	1764	96	2346	1798	54	14788	14892	98	1166	2140
User 31	50	32196	50	23110	50	30976	11380	50	50	51140	44932	8580	22390	54	52	31500	50	27624	52	34530	30884	50	50	50	30850	45390	54	50	50	18906	50	50	13958	50608	50
User 32	54	24664	50	18384	50	24932	12484	3814	56	37398	34498	5294	18462	2564	3006	29118	50	42858	10074	29500	34058	16498	50	50	20868	32282	3832	50	72	4822	64	50	31642	44612	10588
User 33	50	4080	6732	5142	924	5580	6194	2778	924	12806	11254	942	5008	2130	100	5708	9166	10858	6872	6434	11942	5716	52	19206	5978	9268	50	1636	3362	4654	29684	21046	64	16038	1352
User 34	1086	54	5648	2168	1148	54	3458	2096	2328	1858	954	50	50	116	400	52	8056	54	4030	436	54	2434	82	14852	564	1772	54	3604	696	50	24448	24406	52	50	1322
User 35	58	5136	176	5244	60	4468	6356	70	570	10228	6672	1702	4164	176	56	4986	2986	4970	690	7092	5906	68	58	2868	5134	7636	488	458	52	3856	7606	9872	54	9280	54
total	9	2	7	0	5	2	0	10	4	1	0	8	4	8	6	1	3	0	3	2	1	2	12	5	1	0	10	6	7	8	1	3	19	0	9
FAR	25.71	5.71	20.00	0.00	14.29	5.71	0.00	28.57	11.43	2.86	0.00	22.86	11.43	22.86	17.14	2.86	8.57	0.00	8.57	5.71	2.86	5.71	34.29	14.29	2.86	0.00	28.57	17.14	20.00	22.86	2.86	8.57	54.29	0.00	25.71

35 Users
1190 Impostors

Accept Count	159
FRR	0% Fixed
FAR	13.4%

0.6 SD	Reference Profiles				Trigraph																												
	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17	User 18	User 19	User 20	User 21	User 22	User 23	User 24	User 25	User 26	User 27	User 28	User 29	User 30	User 31	User 32	User 33	User 34	User 35
User 1	10864	24583	18109	24470	21382	23925	22219	21351	21737	24328	20823	23721	22850	22053	23094	24505	17858	23549	19931	23902	23828	18290	14670	11683	24457	24794	22352	19568	20860	24211	14719	8978	17146	24340	17367
User 2	37112	30029	32603	38420	36713	30295	33775	34893	38376	29310	18887	35467	30034	34456	38540	31641	39482	28068	32079	30717	28347	29809	29775	37590	34253	34653	34850	35159	32421	33357	36873	36104	15920	20840	29611
User 3	102381	110432	32565	109853	101950	107204	101228	101158	105525	109916	91044	107979	104946	105162	107788	109299	85351	106379	96365	108811	106879	85335	77728	67946	110255	111501	107208	91568	97788	109669	54565	46804	78932	107636	88498
User 4	19351	19217	21316	15256	19733	19135	21091	20662	19600	19441	20765	18741	19515	19648	19168	19075	21015	17834	21692	21122	20552	21578	20577	23712	17668	17985	19515	20650	20508	18754	25512	27158	18251	19449	19829
User 5	34424	37500	25622	38048	30556	37009	33815	33259	34878	37290	28251	37303	35689	34468	36359	37690	30204	36516	31467	35527	36393	27936	26490	21767	38122	38809	35592	30382	32993	36988	20110	10157	20757	35709	28131
User 6	34283	30000	28971	35444	33824	28745	30611	32524	35324	29526	21138	33272	28979	32066	33998	30909	35415	25600	29978	29384	27019	26671	28188	31763	33624	33437	32143	32256	31030	32205	31383	30482	15609	26630	28428
User 7	51588	53818	35455	58107	50720	52283	27068	45807	51457	45895	30677	54097	48167	47106	51976	56732	51920	47941	38899	43818	50877	30275	36517	33091	57412	59783	48961	43901	41314	52104	14201	6213	22820	52473	34913
User 8	33993	34994	26753	36762	33562	33728	30860	28306	35061	32921	21397	35232	32807	33248	35006	34574	31174	32600	27432	30759	31423	23728	24828	27740	35563	36213	34162	30991	30303	34882	25456	22894	18586	32948	25593
User 9	23252	25578	14810	25583	23244	24925	23075	22498	21892	25654	18857	25031	24098	23328	24097	25421	20887	24400	20954	24885	24405	18851	17050	15413	25678	26052	23723	20444	20547	25404	11106	7897	17653	25454	17603
User 10	46877	45380	43387	50829	47025	44291	38310	41047	48690	24840	20945	47651	42856	42581	46442	47320	52658	38130	38936	34953	38884	35232	36481	50864	48823	48516	44168	43545	40533	44812	49028	41838	21111	41815	33887
User 11	142955	144383	118585	150646	141701	142536	137907	129488	144213	137548	17017	146814	142872	137474	145171	144281	139936	133927	121101	134080	139418	118201	114451	133376	143760	146806	139789	134518	135036	141754	120976	100202	79586	138291	108914
User 12	14274	12881	13650	14488	14718	12638	14007	13870	14533	12681	10529	11979	12502	13413	13984	12800	15346	10951	14222	13500	13139	13525	12646	14874	13203	13225	13718	14315	13507	12826	15802	15725	9132	11635	12325
User 13	42612	39759	37339	44440	41227	39527	38348	41388	43545	38960	28059	42876	30241	40582	42870	42043	42409	38995	38661	38657	38469	38082	38071	38340	42634	43477	41030	40931	40675	40027	38636	29124	20353	38105	37619
User 14	21804	23359	17125	24341	21689	22374	19820	19365	21823	21295	10664	23063	21698	19282	22563	23248	22833	20544	18399	19503	20445	16637	17008	20277	23696	24044	20775	19656	19001	22735	18119	16007	10754	22480	14140
User 15	10285	10754	8035	11715	10421	10530	9581	9103	10767	8402	3668	11357	10057	9870	10520	10529	11605	6810	6329	6511	6220	7137	6872	10023	11403	11180	10145	9705	9290	10873	9457	7683	3944	9574	7049
User 16	30542	24827	25686	31026	30122	24709	27875	29264	31421	24336	17520	28975	25082	28190	30675	23328	32031	20800	26082	26323	22412	24075	24550	31118	27049	27356	28796	29566	27618	27857	30223	30752	9924	20688	24803
User 17	67090	70666	49049	70346	67363	69701	70019	66292	69405	72685	67001	68897	69821	69660	68654	70318	39634	70304	67245	71398	70164	60671	52263	49773	70803	71158	68716	63301	64265	70177	56103	46084	57882	71265	62120
User 18	56912	53532	54239	61235	58275	51050	54130	53749	58638	51014	38720	58259	52878	53805	55909	55374	63185	29037	52114	51774	51734	47752	47230	58219	57111	58898	53957	55140	51398	54969	52121	58414	29838	47988	47888
User 19	69980	72265	52234	73691	70131	70555	62286	64622	70591	71642	47447	71517	69042	69505	71601	71977	64820	68967	45473	68131	67913	49145	51620	50192	72741	74854	69548	64999	62128	72571	46297	39089	45097	69984	54563
User 20	51681	50562	38962	56234	51409	49213	40069	44862	52771	41091	22515	51206	45598	45552	51511	51941	54426	43675	37526	28244	37395	32743	38098	44571	54661	55689	47465	43047	40598	49923	31984	22853	19819	45473	32936
User 21	56097	48880	43285	58689	55770	46697	47428	50612	58020	44106	33224	51360	45643	48526	55696	49578	58146	43533	43930	38498	26016	40710	41516	53633	52845	54713	49779	50062	47146	46888	44849	43099	21064	44297	41904
User 22	73545	81149	49006	62083	73804	78629	69323	67170	75214	78958	53377	80155	76905	73234	75418	80898	69945	72131	59584	73984	76918	27480	47049	43631	82627	83640	73870	68204	60527	80788	30683	28300	41981	78279	48458
User 23	74949	78850	58673	76390	74718	77717	74247	73761	76319	79050	64558	77351	75920	76099	76880	78384	68458	76209	70124	76835	77324	62025	47739	52474	78547	79129	75995	71388	71432	77997	53151	44388	56311	77950	63328
User 24	120821	138240	87348	137914	122312	134708	123002	122540	123796	140814	112425	133949	131065	125729	128574	136865	103452	129648	110782	134170	136311	103419	88188	20255	137692	140724	129085	110008	113488	136301	56896	38990	100559	137011	104142
User 25	27077	24488	29017	26634	26999	24579	28755	27055	27640	23560	23472	25854	24883	25859	27067	24192	29458	21489	27427	27135	25564	28758	28416	32384	21601	24538	26268	28134	27808	25075	37317	37936	17381	22417	26319
User 26	34501	29526	35420	33293	34850	29702	35189	33752	35366	28902	28461	32731	29946	33147	34357	29590	37026	26109	34259	31859	29767	33158	30838	39026	30772	24541	33781	35167	33570	31468	42668	44116	19032	28947	29694
User 27	21903	23203	16518	23920	21782	22824	18496	20744	22072	20421	14354	23102	21808	21013	22188	23328	22448	20157	18532	19396	20786	15328	16822	17415	23539	24235	21112	19840	18958	22739	13343	10929	12213	21356	16280
User 28	45883	47769	30155	47883	45450	46657	44205	44399	45469	48520	40860	47307	46260	46441	46965	48024	40371	47294	42228	47117	47524	38921	36240	29872	47837	48418	48229	36903	42903	47500	22739	16310	36685	48012	39328
User 29	52494	56225	37506	57903	52242	54804	49753	48877	52949	55093	43603	55698	53279	51261	53580	56722	48859	52225	44784	52763	53775	38967	38623	36130	57459	58891	52705	47247	30706	55809	27519	22950	37412	55792	40065
User 30	28880	27623	27743	29348	29092	27190	28064	27940	29278	26183	22661	28343	26962	27703	28879	27738	30195	24424	28052	27137	25429	26812	28431	30921	27828	28485	27859	28407	28125	25313	30352	29954	19708	25525	25881
User 31	187787	225598	103398	222825	182331	219958	178656	172559	183599	219924	163578	215300	209480	169732	201254	224614	166111	198720	155551	202365	208694	105397	121932	86732	226189	233097	191181	165466	144884	218533	55	37240	146128	225594	113857
User 32	228320	245409	154822	246225	219518	243527	207119	222090	229339	237492	177882	239040	231254	227193	236589	249812	196787	243250	208700	229179	238752	194022	168410	139590	245485	251412	234851	198925	206428	234377	98314	52	169762	244485	197640
User 33	181783	181390	159354	185111	179376	180878	175828	179910	181303	181197	149364	183500	179192	180350	182994	181660	177769	176346	171501	176733	181495	163814	159918	164325	184841	185343	179282	178171	177757	184119	155011	124937	25525	175303	164845
User 34	87247	83193	79666	87714	88410	81683	82951	85964	86200	81593	68952	85947	81228	85204	86865	83431	88870	75253	83650	80317	77270	76810	77393	85465	85405	85006	85718	85873	83210	84577	86314	81147	49508	30805	79227
User 35	80998	86292	80816	87698	81029	84870	80288	77891	82661	84358	59285	85751	83119	81374	83829	88248	78166	80162	72904	81034	81590	57489	58258	61336	86580	87728	82302	77823	71244	80131	47074	52063	57674	83127	44164
total	3	11	14	2	10	9	7	9	5	8	4	1	13	2	0	5	14	13	22	10	2	11	24	5	3	6	4	15	15	7	0	0	20	13	22
FAR	8.57	31.43	40.00	5.71	28.57	25.71	20.00	25.71	14.29	22.86	11.43	2.86	37.14	5.71	0.00	14.29	40.00	37.14	62.86	28.57	5.71	31.43	68.57	14.29	8.57	17.14	11.43	42.86	42.86	20.00	0.00	0.00	57.14	37.14	62.86

35 Users
1190 Impostors

Accept Count 309
FRR 0% Fixed
FAR 26.0%

0.8 SD
Reference Profiles
Telegraph
User 1
User 2
User 3
User 4
User 5
User 6
User 7
User 8
User 9
User 10
User 11
User 12
User 13
User 14
User 15
User 16
User 17
User 18
User 19
User 20
User 21
User 22
User 23
User 24
User 25
User 26
User 27
User 28
User 29
User 30
User 31
User 32
User 33
User 34
User 35
cum
FAR
35 Users
1100 Impostors
Accept Count
FRR
FAR
232
0% Fixed
19.5%

0.6 SD
Reference Profiles
Word
Unmatche
User 1
User 2
User 3
User 4
User 5
User 6
User 7
User 8
User 9
User 10
User 11
User 12
User 13
User 14
User 15
User 16
User 17
User 18
User 19
User 20
User 21
User 22
User 23
User 24
User 25
User 26
User 27
User 28
User 29
User 30
User 31
User 32
User 33
User 34
User 35
total FAR
35 Users
1180 Impostors
Accept Count
FRR
FAR
188
0% Fixed
15.0%

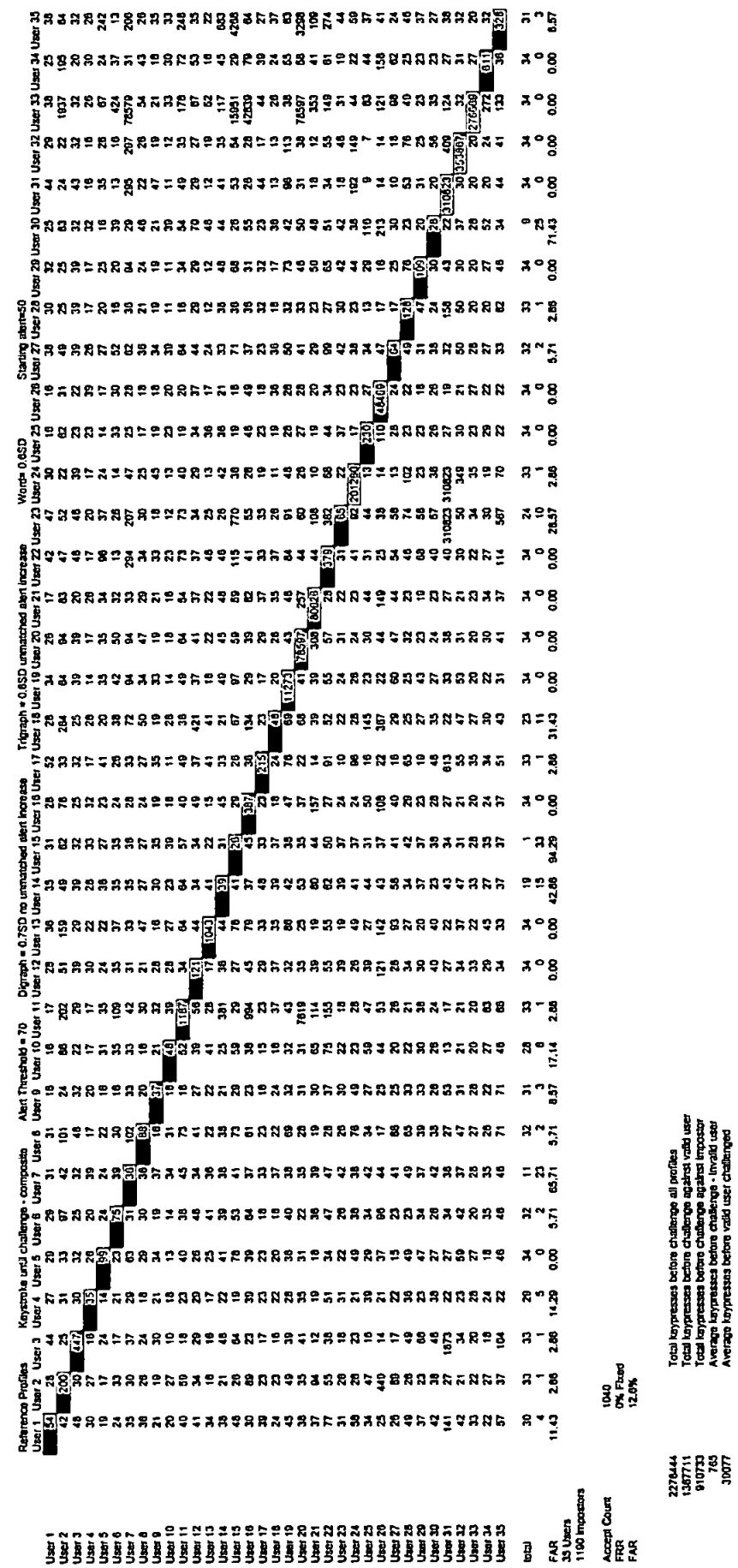
0.7 SD — Reference Profiles

	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17	User 18	User 19	User 20	User 21	User 22	User 23	User 24	User 25	User 26	User 27	User 28	User 29	User 30	User 31	User 32	User 33	User 34	User 35
User 1	56	21	72	22	22	22	50	24	24	38	31	23	23	22	21	21	66	33	43	22	19	94	116	52	18	21	40	41	52	21	16	93	135	18	106
User 2	29	53306	23	18	26	280	65	29	28	320	53306	26	53306	57	54	50	125	53306	79	321	64	27	127	23	49	16	27	24	54	120	27	19	53306	53306	524
User 3	39	15	1587	18	19	26	22	20	30	20	14	28	19	17	21	21	20	127	14	28	15	14	50	98	79	15	19	23	79	29	18	49	325	22	23	25
User 4	22	68	15	300	15	37	16	23	19	55	22	27	53	17	22	32	22	178	18	19	29	18	63	15	80	68	25	16	16	120	15	18	34	53	20
User 5	25	31	34	19	73	31	31	30	35	22	22	26	29	32	27	29	145	17	20	29	21	141	50822	51	20	31	34	24	73	29	74	124	50822	30	34
User 6	18	54	26	22	70	141	43	19	23	60	347	25	156	22	23	32	44	3435	77	104	63	58	76	17	26	26	39	17	40	37	30	43	50167	209	145
User 7	43	101	372	27	145	108	78579	54	50	36	78579	55	247	42	38	39	78579	60	78579	78579	28	78579	78579	2189	16	25	58	205	408	68	78579	78579	78570	280	78570
User 8	21	30	27	26	25	39	29	225	27	30	257	25	96	34	26	24	72	63	128	24	29	30	126	24	34	28	18	77	94	22	27	107	50102	93	489
User 9	46	15	116	19	45	28	40	64	106	19	19	19	27	19	24	20	125	20	27	34	15	296	11107	110	16	16	21	252	250	17	400	316	26	34	14
User 10	23	22	20	18	21	17	25	39	18	70337	83	21	26	32	21	17	41	40	46	41	28	49	29	20	22	22	49	17	21	18	34	26	70337	44	101
User 11	20	84	25	17	19	88	15	63	16	144	227680	43	117	26	21	35	22	113	11	63	65	61	725	26	51	48	19	12	21	60	21	17	227680	145	227680
User 12	26	103	42	33	21	102	40	79	22	38	180	94	120	30	21	46	30	20216	64	59	82	48	83	21	25	18	41	28	32	200	26	49	177	103	48
User 13	24	31	24	33	27	41	23	46	15	30	265	33	107	38	27	28	52	86	28	40	51	97	162	52	20	33	43	23	25	53	15	23	65312	40	172
User 14	38	33	19	16	35	34	64	64	40	102	33639	20	49	79	32	31	217	51	91	150	61	139	1162	59	17	15	63	35	169	71	57	38	541	38	33639
User 15	17	44	12	20	23	29	23	15	17	17	15951	19	15	18	23	17	28	15951	20	48	51	18	15951	20	14	34	32	23	57	62	14	14	15951	62	15951
User 16	19	42839	17	21	18	37	17	46	26	1085	42	43	42839	39	40	60	22	42839	21	23	35	21	1111	15	144	61	42	15	19	642	15	15	42839	42839	31
User 17	33	21	65	24	51	35	89	83	53	17	56	23	49	45	23	21	105543	59	51	108	64	57	209	38	19	20	33	29	55	22	100	214	105543	38	1954
User 18	32	28	51	25	23	30	47	17	15	28	144	32	48	33	23	48	51	166	44	33	48	44	43	21	34	23	22	24	67	33	45	36	191	236	51
User 19	23	38	121	21	31	32	59	50	26	44	372	20	83	51	24	48	104	45	103876	172	33	103876	1606	51	30	22	47	56	1229	30	101	91	157	60	103876
User 20	21	67	54	20	34	64	62	55	48	541	78597	21	102	61	38	28	66	65	35	78597	73391	131	121	47	32	32	63	34	48	26	49	44	78597	132	78597
User 21	19	2555	38	22	39	206	106	129	25	149	80626	48	2850	69	22	648	84	3444	199	211	80620	131	120	51	52	69	77	38	104	370	78	105	80626	80626	80626
User 22	28	32	1160	18	29	90	29	28	24	30	343	23	27	30	19	36	23125	30	315	70	32	117365	117365	51938	19	19	30	37	3808	19	70454	190	117365	91	117365
User 23	20	31	14	20	16	20	14	21	20	29	30	28	18	30	25	16	21	42	30	24	24	19	23	24	29	24	22	13	25	30	32	30	30	31	19
User 24	45	16	602	14	34	48	45	66	33	22	39	19	52	26	28	20	90	18	29	15	18	347	201260	201260	14	16	26	135	29	20	56	201260	153	24	201260
User 25	19	14	13	24	16	16	13	17	15	22	47	22	14	26	19	17	13	15	13	13	13	13	13	13	36	41	19	14	14	49	13	13	44	19	15
User 26	16	16923	15	38	37	162	20	56	22	58	18527	64	902	31	26	489	38	48469	65	158	417	56	148	21	189	48469	30	16	56	2127	17	17	48469	48469	271
User 27	27	22	110	17	38	33	40	41	25	29	55	26	69	22	22	20	100	53	19	41	37	33068	33068	111	13	38	69	20	98	37	2079	65	33068	40	33068
User 28	27	20	70217	16	28	17	51	17	29	19	16	22	16	22	22	21	18	20	18	19	19	15	95	426	18	16	22	46	33	17	102	23	62	15	55
User 29	27	32	58	20	27	24	33	33	17	33	35	20	24	24	20	24	44	32	17	18	17	42	68	27	16	16	18	32	39	27	49	35	58	29	129
User 30	21	42	38	23	27	44	34	53	20	50	43	31	50	37	20	49	34	58	37	35	43	34	36	27	47	31	34	37	37	71	23	41	92	50	45
User 31	43	26	310823	18	61	34	66	25	39	22	14	20	28	17	22	21	310823	37	106	83	20	310823	310823	310823	17	19	38	170	310823	23	310823	310823	96	59	310823
User 32	23	13	52	17	38	35	53	32	25	14	273	21	23	23	25	13	112	39	59	49	13	67	404	32	18	15	32	52	287	31	32	353867	349	14	353867
User 33	39	25	39	23	26	34	32	61	20	25	27	23	26	28	20	33	76	22	22	26	26	67	122	99	17	32	23	25	76	26	74	64	276800	26	26
User 34	26	78	27	19	61	65	52	26	43	64	923	22	96	30	33	32	68	120	54	65	148	68	684	49	38	26	40	39	55	39	90	91	954	900	819
User 35	21	25	20	22	16	19	41	16	24	40	67	27	20	31	21	17	73	57	41	33	33	128	126	52	16	20	24	34	31	54	52	47	59	51	140044
total	34	34	33	34	32	31	34	34	34	34	34	34	26	32	19	32	33	26	34	34	34	33	1	33	27	34	33	25	13	26	34	34	34	30	30
	0	0	1	0	2	3	0	0	0	0	0	0	8	2	15	2	1	8	0	0	0	1	33	1	7	0	1	9	21	8	0	0	0	4	4
FAR	0.00	0.00	2.86	0.00	5.71	8.57	0.00	0.00	0.00	0.00	0.00	0.00	22.86	5.71	42.86	5.71	2.86	22.86	0.00	0.00	0.00	2.86	94.29	2.86	20.00	0.00	2.86	25.71	60.00	17.14	0.00	0.00	0.00	11.43	11.43

35 Users
1190 Impostors

Accept Count 1061
FRR 0% Fixed
FAR 10.8%

10010958 Total keypresses before challenge across all profiles
2406438 Total keypresses before challenge against valid user
7604520 Total keypresses before challenge against impostor
6741 Average keypresses before challenge - invalid user
80215 Average keypresses before valid user challenged


Reference Profiles
Keystroke urd challenge - composite
Alert Threshold = 70
Digraph = 0.7SD no unmatched alert increase
Trigraph = 0.6SD unmatched alert increase
Words 0.6SD
Starting alert=50
Accept Count
FRR
FAR
35 Users
1190 Impostors
1040
0% Fixed
12.6%
Total keypresses before challenge all profiles
Total keypresses before challenge against valid user
Total keypresses before challenge against impostor
Average keypresses before challenge - invalid user
Average keypresses before valid user challenged
2278444
1367711
910733
765
30077

	Reference Profiles			Keystroke until challenge - composite					Alert Threshold = 70			Variable digraph, trigraph and word S.D.'s							Starting alert=50											
	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17	User 19	User 20	User 21	User 22	User 24	User 25	User 26	User 27	User 28	User 29	User 31	User 32	User 34
Digraph SD	0.6	0.6	0.6	0.7	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.6	0.7	0.6	0.7	0.6	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.6	0.7	0.6	0.7	0.7	0.6
Trigraph SD	0.5	0.5	0.5	0.6	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.5	0.6	0.5	0.6	0.5	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.5	0.6	0.5	0.6	0.6	0.5
Word SD	0.7	0.5	0.5	0.6	0.6	0.6	0.6	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
Digraph Unmatched	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Trigraph Unmatched	No	No	No	No	No	No	No	Yes	Yes	No	No	No	Yes	No	Yes	No	No	No	No	Yes	Yes	Yes	No	Yes	No	Yes	No	No	No	No
Threshold	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70
User 1	34352	51	49	53	73	44	50	16	102	38	17	94	36	101	575	34	115	34	26	17	104	54	27	28	74	170	101	44	29	25
User 2	50	53306	27	46	50	161	55	83	34	489	202	75	159	87	53306	390	93	64	94	83	112	40	130	62	361	37	33	24	22	53306
User 3	649	96	156718	90	220	48	52	39	518	53	29	207	29	173	519	78	137	39	39	20	210	70	62	39	98	1706	76	43	32	76
User 4	72	50	16	27324	39	34	37	16	47	28	17	181	22	71	167	47	27	14	17	26	22	25	42	72	55	21	21	16	16	52
User 5	181	39	117	25	50822	54	103	17	71	239	35	238	22	1780	367	39	324	35	35	34	50822	67	30	27	505	110	108	35	26	37
User 6	61	37	20	31	61	320	34	20	21	42	109	175	37	41	67	35	33	42	50	32	17	23	43	42	910	17	23	13	16	48
User 7	744	57	102	52	17450	86	78579	86	78579	102	42	265	33	272	374	43	718	94	94	33	78579	83	43	46	78579	272	223	295	297	39
User 8	45	30	25	24	59	51	58	71	29	47	30	44	47	45	280	40	50	34	47	29	50102	35	24	36	75	35	60	22	26	63
User 9	44	21	31	35	39	35	41	12	37618	34	32	40	16	69	77	29	41	33	19	21	44	102	23	29	58	54	40	47	19	31
User 10	48	69	12	53	32	38	29	28	52	70337	39	79	27	86	194	33	16	14	18	18	35	22	101	34	345	16	18	11	12	53
User 11	79	254	19	36	67	84	51	70	38	522	1187	59	84	447	345	59	72	49	64	54	720	53	38	42	94	34	40	49	35	317
User 12	142	133	61	137	83	72	49	39	71	121	56	20216	44	122	518	133	50	37	41	37	72	39	218	71	119	87	38	29	27	133
User 13	114	87	16	61	82	96	41	18	52	92	28	138	1043	92	216	27	87	18	22	22	215	25	124	46	49	16	24	12	19	18
User 14	137	27	54	48	196	53	87	25	100	46	381	91	44	33639	115	95	39	49	45	48	578	48	77	39	444	50	171	41	35	96
User 15	137	21	97	23	142	93	104	53	65	15951	29	38	78	473	15951	36	36	97	59	59	15951	56	21	27	489	83	142	53	54	50
User 16	56	42839	37	69	100	89	50	41	44	42839	994	100	79	566	42839	42839	53	29	39	62	1059	48	104	64	983	59	58	26	28	42839
User 17	255	44	47	53	51	39	55	17	1599	23	23	83	33	105	175	44	105543	17	29	37	58	68	48	41	58	232	49	44	17	51
User 19	311	94	94	52	87	72	69	49	134	53	43	82	88	153	330	102	220	11273	43	48	103876	97	44	38	533	109	350	96	113	120
User 20	71	89	53	63	57	46	53	25	57	78597	7819	113	25	78597	703	51	28	41	78597	257	78597	38	30	41	78597	52	76	31	38	126
User 21	111	80628	12	66	27	98	44	14	47	148	114	99	19	290	328	330	15	39	308	80628	115	32	47	61	114	30	74	18	12	187
User 22	399	69	180	79	175	89	71	25	256	106	155	119	55	262	735	57	147	55	57	28	117365	152	58	53	528	201	349	34	55	72
User 24	2118	39	25	23	283	50	68	24	281	23	28	59	49	72	111	33	281	26	24	23	135	201260	21	33	93	27	284	192	149	22
User 25	87	129	16	232	41	47	44	21	31	241	47	273	27	85	79	80	31	23	36	44	40	27	38944	50	84	15	32	9	7	53
User 26	73	48469	11	211	44	144	31	14	53	53	53	48469	142	121	173	424	26	22	44	149	37	21	48469	48469	71	21	22	14	14	48469
User 27	65	137	17	44	56	82	78	58	107	32	26	269	93	609	586	266	40	60	47	44	33068	27	98	53	33068	39	60	10	18	269
User 28	505	30	354	79	169	66	49	49	192	34	21	81	27	61	305	54	590	25	32	23	70217	148	37	41	157	70217	179	53	76	29
User 29	474	34	30	52	72	52	52	23	75	52	38	66	20	68	1448	44	309	43	23	19	88059	74	49	35	51	3013	88059	31	25	35
User 31	310823	37	310823	38	38	54	153	26	310823	14	17	59	22	284	213	38	310823	33	38	27	310823	310823	37	37	73	310823	310823	310823	409	57
User 32	1373	69	111	75	131	59	73	37	86	43	21	141	37	114	131	70	61	53	31	21	128	371	64	37	503	3113	95	30	353867	38
User 34	48	73	22	123	24	77	34	18	38	84	83	396	45	73	2490	73	209	22	30	34	82	34	101	42	85	26	48	20	24	124409
total	28	28	28	29	29	29	29	27	27	28	28	28	29	28	27	29	28	29	29	29	28	28	28	29	27	28	28	29	29	29
	1	1	1	0	0	0	0	2	2	1	1	1	0	1	2	0	1	0	0	0	1	1	1	0	2	1	1	0	0	0
FAR	3.33	3.33	3.33	0.00	0.00	0.00	0.00	6.67	6.67	3.33	3.33	3.33	0.00	3.33	6.67	0.00	3.33	0.00	0.00	0.00	3.33	3.33	3.33	0.00	6.67	3.33	3.33	0.00	0.00	0.00

30 Users
870 Impostors

Accept Count	849
FRR	0% Fixed
FAR	2.4%

6426998	Total keypresses before challenge across all profiles
2286842	Total keypresses before challenge against valid user
4140156	Total keypresses before challenge against impostor
4759	Average keypresses before challenge - invalid user
76228	Average keypresses before valid user challenged

Appendix D

Published Papers

Published Papers

The list below presents papers written and published during the PhD work programme. Those highlighted with a '*' are not included in this appendix but are available from the Network Research Group web site (http://www.network-research-group.org).

1. "A Long-Term Trial Of Keystroke Profiling Using Digraph, Trigraph And Keyword Latencies", Dowland P.S. and Furnell S.M., Accepted for publication in the *Proceedings of the IFIP SEC 2004 Conference*, Toulouse, France, August 2004

* 2. "A Correlation Framework for Continuous User Authentication Using Data Mining", Singh H., Furnell S.M., Dowland P.S., Lines B. and Kaur S., Accepted for publication in the *Proceedings of the Fourth International Network Conference (INC 2004)*, Plymouth, UK, July 2004

* 3. "A long-term trial of alternative user authentication technologies", Furnell S.M., Papadopoulos I. and Dowland P.S., Accepted for publication in *Information Management and Computer Security*, 2004

* 4. "Improving Security Awareness And Training Through Computer-Based Training", Furnell S.M., Warren A. and Dowland P.S., *Proceedings of the WISE Conference*, Monterey, USA, July, pp287-301, 2003

* 5. "Assessing IT Security Culture: System Administrator and End-User", Finch J., Furnell S.M. and Dowland P.S., *Proceedings of ISOneWorld Conference 2003*, Las Vegas, USA, April 23-25, 2003

* 6. "A prototype tool for information security awareness and training", Furnell S.M., Gennatou M. and Dowland P.S., *International Journal of Logisitics Information Management*, vol. 15, no. 5, pp352-357, 2002

* 7. "Critical awareness – The problem of monitoring security vulnerabilities", Furnell S.M., Alayed A., Barlow I. and Dowland P.S., *Proceedings of European Conference on Information Warfare and Security*, 8-9 July 2002, Brunel, UK, pp85-92 2002

* 8. "An experimental comparison of secret-based user authentication technologies", Irakleous I., Furnell S.M., Dowland P.S. and Papadaki M., *Information Management & Computer Security*, vol. 10, no. 3, pp100-108, 2002

9. "Keystroke Analysis as a Method of Advanced User Authentication and Response", Dowland P.S., Furnell S.M. and Papadaki M., *Proceedings of IFIP/SEC 2002 - 17th International Conference on Information Security*, Cairo, Egypt, 7-9 May, pp215-226, 2002

10. "A Preliminary Investigation of User Authentication Using Continuous Keystroke Analysis", Dowland P.S., Singh H. and Furnell S.M., *Proceedings of the IFIP 8th Annual Working Conference on Information Security Management & Small Systems Security*, Las Vegas, 27-28 September, 2001

* 11. "Security analysers: Administrator Assistants or Hacker Helpers?", Furnell S.M., Chiliarchaki P. and Dowland P.S., *Information Management and Computer Security*, vol. 9, no.2, pp93-101, 2001

* 12. "Investigating and Evaluating Behavioural Profiling and Intrusion Detection Using Data Mining", Singh H., Furnell S.M., Lines B. and Dowland P.S., *Proceedings of International Workshop on Mathematical Methods, Models and Architectures for Computer Networks Security*, St. Petersburg, Russia, 21-23 May, 2001

* 13. "A Generic Taxonomy for Intrusion Specification and Response", Furnell S.M., Magklaras G.B., Papadaki M. and Dowland P.S., *Proceedings of Euromedia 2001*, Valencia, Spain, 18-20 April, 2001

* 14. "Promoting security awareness and training within small organisations", Furnell S.M., Gennatou M. and Dowland P.S., *Proceedings of the First Australian Information Security Management (AISM) Workshop*, Geelong, Australia, 7 November, 2000

15. "Authentication and Supervision: A survey of user attitudes", Furnell S.M., Dowland P.S., Illingworth H.M. and Reynolds P.L., *Computers & Security*, vol. 19, no. 6, pp529-539, 2000

16. "A conceptual intrusion monitoring architecture and thoughts on practical implementation", Dowland P.S. and Furnell S.M., *Proceedings of the World Computer Congress 2000*, 21-25 August, 2000

17. "Enhancing Operating System Authentication Techniques", Dowland P.S. and Furnell S.M., *Proceedings of the Second International Network Conference (INC 2000)*, Plymouth, UK, pp253-261, 3-6 July, 2000

18. "A conceptual architecture for real-time intrusion monitoring", Furnell S.M. and Dowland P.S., *Information Management & Computer Security*, vol. 8, no. 2, pp65-74, 2000

* 19. "Developing tools to support online distance learning", Furnell S.M., Evans M.P. and Dowland P.S., *Proceedings of EUROMEDIA 2000*, Antwerp, Belgium, 8-10 May, 2000

* 20. "Computer Crime and Abuse: A Survey of Public Attitudes and Awareness", Dowland P.S., Furnell S.M., Illingworth H.M. and Reynolds P.L., *Computers & Security*, vol. 18, no. 8, pp715-726, 1999

* 21. "Dissecting the 'Hacker Manifesto'", Furnell S.M., Dowland P.S. and Sanders P., *Information Management and Computer Security*, vol.7, no.2 pp69-75, 1999

A Long-term Trial of Keystroke Profiling using Digraph, Trigraph and Keyword Latencies

Paul S. Dowland and Steven M. Furnell
Network Research Group, School of Computing, Communications and Electronics, University of Plymouth, Drake Circus, Plymouth, PL4 8AA, United Kingdom, info@network-research-group.org

Abstract: A number of previous studies have investigated the use of keystroke analysis as a means of authenticating users' identities at the point of initial login. By contrast, relatively little research has focused upon the potential of applying the technique for identity verification during the logged-in session. Previous work by the authors has determined that keystroke analysis is a viable metric for continuous monitoring, provided that sufficient data is captured to create a reliable profile. This paper presents a series of results from a three-month trial in which profiles were created using digraph, trigraph and keyword-based keystroke latencies. The profiles were based upon a total of over 5 million keystroke samples, collected from 35 participants. The results demonstrate that the techniques offer significant promise as a means of non-intrusive identity verification during keyboard-related activities, with an optimum false acceptance rate of 4.9% being observed at a rate of 0% false rejection.

Key words: Authentication, Misuse Detection

1. INTRODUCTION

Over the last twenty years, the concept of keystroke analysis has been the focus of considerable research as a means of user authentication. The potential for profiling of keypresses was first identified by Gaines et al (1980). Since then, a number of research projects have been conducted to evaluate different methods of data gathering (using a range of operating systems and considering a variety of metrics) and post-processing techniques (ranging from purely statistical to AI/neural network approaches).

To date, however, virtually all published studies have focussed upon looking at the application of static strings, such as username and password pairs using the inter-keystroke digraph latency timing method. From the earliest studies in 1980 (Card et al & Gaines et al), the focus has been on the analysis of digraph latencies. Later studies, such as those by Joyce & Gupta (1990) and Mahar et al (1995) further enhanced the work, identifying additional statistical analysis methods that provided more reliable results.

In Legget et al. (1991), the concept of dynamic keystroke analysis was first proposed, with the introduction of a reference profile that could be used to monitor a live user session. Brown and Rogers (1993) also explored the idea of dynamic analysis, presenting preliminary results.

The authors' previous research (Dowland et al., 2002) described an experiment evaluating keystroke analysis based on inter-keystroke digraph latencies under Windows. This earlier trial concentrated upon the capture and subsequent analysis of digraph latencies using inter-keystroke timings. The trial results demonstrated the viability of this method, but suggested that, to be a reliable authentication measure, user profiles would need to be based upon much larger sample sizes. The previous trial was also based on a limited number of users in order to quickly evaluate the viability of the technique.

This paper presents the results of a long-term trial that was aimed at evaluating a range of techniques using a larger number of participants. This trial captured and evaluated trigraph and keyword latencies, in addition to digraph timings, under the Windows operating system. The paper begins by introducing the technical aspects of the trial conducted over a three month period before considering the statistical approach taken with the data analysis stage. The results are presented and discussed, leading to some overall conclusions, and proposals for future work.

2. CAPTURING KEYSTROKE DATA IN WINDOWS

While keystroke analysis has been investigated (and hence implemented) in previous studies, a GUI environment (e.g. Microsoft Windows) introduces new challenges. In previous published studies, the user has been required to type and interact with a specific application (typing either pre-defined or free-form text). While this approach makes the development of the keystroke monitoring software simple, and maintains the consistency of the

test environment, it is not representative of normal typing behaviour as the user becomes focussed upon the task of typing, rather than focussed upon a task that *involves* typing. If the aim is to produce static keystroke analysis for occasional authentication judgements (e.g. supplementing login authentication) then this approach will work well. However, to implement continuous supervision using dynamic keystroke analysis it is necessary to monitor the users' normal behaviour when interacting with their normal applications and operating system environment. Even providing a simulation of these environments may not be sufficient to obtain valid sample data upon which to base a profile.

In order to address this problem, software was developed that would transparently monitor and log all typing activity. The system was designed to allow keystroke data to be collected under the Microsoft Windows XP environment (although the technique is equally applicable in all Windows operating systems). In order to collect the required data, it was necessary to implement a mechanism for acquiring user typing patterns across all applications running within a users' active session. This is important as the experiment was designed to create a profile for each user based upon their typical typing patterns when using their computer (not constrained to a specific application or task). The implementation of the keylogger utilised several key features of the Windows operating system and the underlying chains of messages on which the operating system is built (these are briefly discussed in the following section). The authors have not investigated the applicability of these techniques under other operating systems but it is likely that the same system could be developed under other systems providing access is given to the keypress events at an appropriate level.

Figure 1 illustrates the software architecture used to capture and log keystroke activity under Windows. As keys are pressed, messages are generated by Windows for both key up and down events. These messages are captured through the use of a hook function that redirects messages to a nominated program. The messages are passed from the hook (implemented as a system-wide DLL written in C) to the keylogger (implemented in Visual Basic and deployed as a system tray application). The keylogger functioned completely transparently to the user, requiring no user action to start or stop the logging process. The application was automatically started when the operating system (O/S) booted (run from the Startup program group on the start menu) and shut down automatically when the O/S closed. Gathered data was automatically saved after every 500 digraphs pairs and when the application was closed. To reassure users, an option was included to

suspend logging of keystrokes. This was included due to concerns expressed by some users about monitoring of specific inputs – e.g. the typing of on-line banking login details.

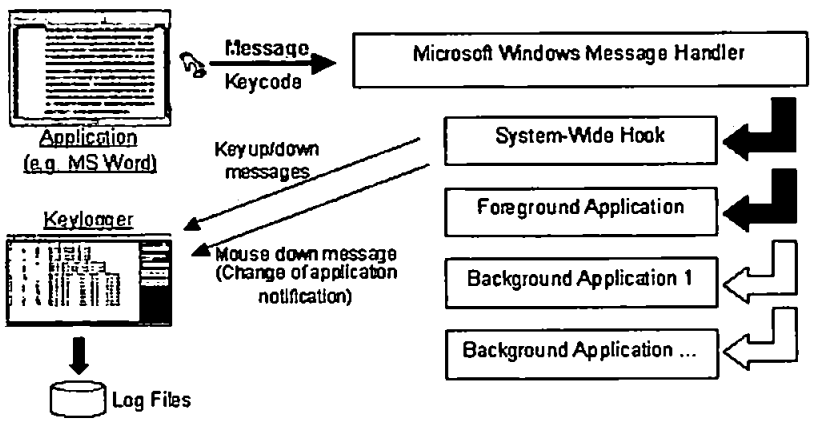


Figure 1. Implementation of keylogger

For each digraph pair logged, the application stored five items of information (Table 1) – these being written to an Access database after every 500 digraphs. This process was also repeated for each trigraph and keyword latency (i.e. trigraphs were stored as three consecutive characters and keywords as a string).

Table 1. Keylogger attributes logged per digraph

Item	Data types
AutoID	Auto-incrementing record number. This is used to maintain the order of the keystrokes typed as the timestamp is only accurate to 1 second.
Left character (C1)	ASCII code representing character
Right character (C2)	ASCII code representing character
Latency	Integer representing inter-keystroke latency in milliseconds
Application	String containing the window title from the foreground application.
Timestamp	A timestamp is added to every keystroke logged for later use.

While digraph and trigraph logging were based upon all keystrokes entered, keyword logging was based on a look up list. The top 200 commonly occurring words in the English language were monitored, and as each word was entered, its latency was recorded.

3. EXPERIMENTAL PROCEDURE

For this experiment a total of 35 users were profiled over a period of

three months. Unfortunately several users disabled the keylogger when entering sensitive information and consequently forgot to re-enable it. Despite this, the key-logging trial collected considerable volumes of data with nearly six million samples collected across digraphs, trigraphs and keywords (Table 2).

Table 2. User profile results

User	Mean Digraph Latency (ms)	Typing Skill Classification	Digraphs	Trigraphs	Words
User 1	91	Best	34352	23352	1403
User 2	156	Average (skilled)	53306	36912	2599
User 3	99	Best	156718	107107	6154
User 4	251	Average (non-skilled)	27324	18688	1310
User 5	112	Good	50822	36713	1465
User 6	154	Average (skilled)	50167	34484	1885
User 7	106	Good	78579	54959	4349
User 8	130	Good	50102	35102	2932
User 9	97	Best	37618	24755	1741
User 10	145	Average (skilled)	70337	48942	4643
User 11	147	Average (skilled)	227660	145846	10617
User 12	102	Good	20216	14142	1032
User 13	157	Average (skilled)	65312	43015	1730
User 14	141	Average (skilled)	33639	23090	1784
User 15	139	Good	15951	11159	1068
User 16	150	Average (skilled)	42839	30299	2037
User 17	106	Good	105543	68068	3173
User 18	177	Average (skilled)	89730	59292	3121
User 19	117	Good	103876	71635	4617
User 20	121	Good	78597	53495	4479
User 21	141	Average (skilled)	80626	55881	2807
User 22	110	Good	117365	79534	6557
User 23	131	Good	118805	77013	5682
User 24	89	Best	201260	131954	8517
User 25	203	Average (skilled)	38944	26655	2266
User 26	192	Average (skilled)	48469	33907	2555
User 27	125	Good	33068	23115	1679
User 28	91	Best	70217	47033	2128
User 29	104	Good	88059	55707	3815
User 30	202	Average (skilled)	40741	28789	1007
User 31	86	Best	310823	211419	19726
User 32	93	Best	353867	237274	18056
User 33	144	Average (skilled)	276669	183455	6057
User 34	143	Average (skilled)	124409	87079	953
User 35	130	Good	140044	85413	6240
		Totals	**3,436,054**	**2,305,283**	**150,184**

Before considering the data from each user, the typing skill for each participant was evaluated based on the categorisations proposed by Card et al. (1980) where typists are broadly categorised into one of six categories. The results are presented in Table 2 together with the quantity of samples for each user (shown separately for digraph, trigraph and keywords). The results are weighted towards typists with above average skills due to the nature of the test subjects (i.e. all subjects were regular computer users who spent prolonged periods typing). This was considered acceptable as the likely use for a fully implemented system would be in environments with semi-skilled users (i.e. relatively few unskilled/poor typists).

4. STATISTICAL ANALYSIS

To eliminate extreme short/long digraph latencies that may adversely affect the distribution of digraph times, any digraph pair whose latency fell outside a nominal range was excluded from the log files. For the purpose of this experiment the range was restricted to times above 10ms and below 750ms. In an earlier trial the range was restricted to 40ms – 750ms, with these thresholds based on previous work conducted by Furnell (1995), and were designed to eliminate samples where two keys may have been accidentally struck together (thus, producing an infeasibly small latency) or, where the user may have made a pause in their typing and thus introduced an unnaturally large inter-keystroke latency. Unfortunately, the low pass filter was responsible for substantial quantities of data being removed from the user profiles and, as such, was reduced to 10ms for the purposes of this trial. If a digraph was removed due to the filtering, this also reset the trigraph and keyword logging so no further thresholds were needed for these two measures.

Following the initial filtering, the experimental data for each user was processed off-line to calculate the mean and standard deviation values for each unique digraph, trigraph or keyword. In the event that any profiled sample had a standard deviation greater than its mean value, the samples were sorted and the top/bottom 10% was then removed, followed by subsequent re-calculation of the mean and standard deviation values. The reason for this additional step was to remove samples where the latencies would have an adverse affect on the standard deviation (i.e. the distribution of samples was tightened).

Once all the user profiles were calculated, another application (the data comparator) was used to generate tables of results for each of the methods.

The data comparator (Figure 2) was based on the original analyser developed in the previous trial (Dowland et al., 2002). A small number of additional features were introduced to the comparator to cater for the inclusion of trigraphs and keyword profiles.

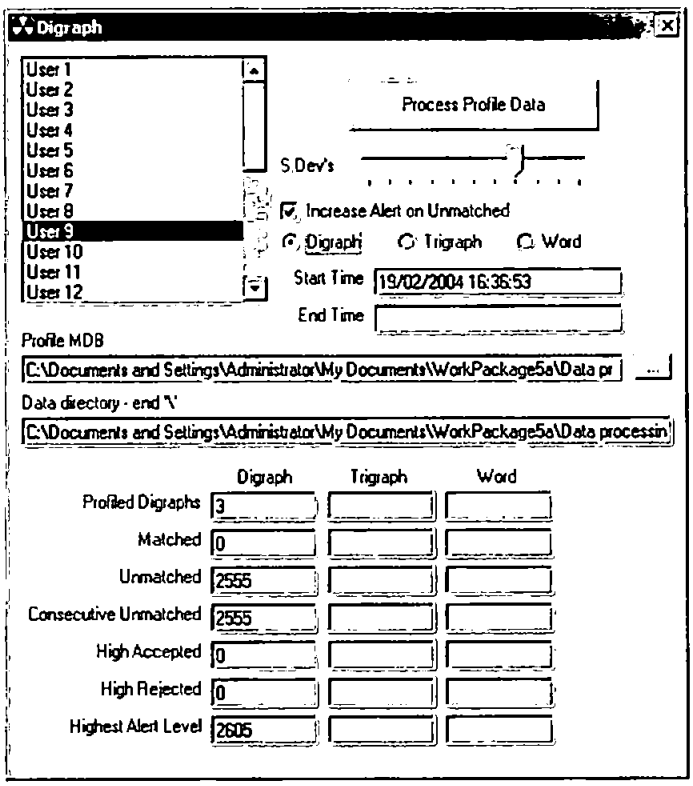


Figure 2. Data comparator (running)

In the previous trial, when a digraph was processed that did not exist in the reference profile, the alert level remained static (simply increasing the count of unmatched digraphs). This trial considered the role of unmatched samples as they are a potential indicator of impostor activity. I.e. if a user types a specific sample infrequently (to the extent that there is insufficient data on which to base a profile), it is reasonable to assume that these occurrences are un-representative of that user's normal typing behaviour. By default, in this trial, an unmatched sample increased the alert level by one, whilst a matched accepted/rejected sample varied the alert level by two accordingly. This behaviour can be adjusted by selecting the checkbox in

the comparator – once unchecked; the alert level was not affected by unmatched samples.

Before starting the full profile comparisons a trial comparison was conducted based on a random selection of five users in order to determine the optimum settings for the deviation threshold. In the previous study the deviation settings were chosen from a range of 0.5, 1.0, 1.5 and 2.0 standard deviations with the best results obtained at 0.5. In order to determine an optimum setting, profile comparisons were made between 0.5 and 1.0 standard deviations (values below 0.5 had already been assessed in earlier trials). For the randomly selected users the best results were obtained at 0.7 with an increase in alert level above and below this threshold. As such, the later comparisons were performed with standard deviations settings of 0.6, 0.7 and 0.8. The permitted deviation was determined by the slider control that selects the number of standard deviations from the mean.

digraph mean ± (digraph standard deviation * permitted deviation)

Once the profile comparison was started each users' reference profile was loaded and then compared against the raw keylogger data files for all 35 users. This resulted in a table of 35 sets of statistics for each user. This process was repeated for trigraphs and keywords with three different profile deviation settings (0.6, 0.7 and 0.8 standard deviations from the mean). NB a setting of 0.5 standard deviations was introduced to the trigraph comparisons due to poor performance at 0.6 and 0.7 and unmatched alert increases were optionally applied to digraphs and trigraphs (hence doubling the number of comparisons for these metrics). With an average of nearly 100,000 samples per data file, each data comparison took approximately two hours with a total of 17 comparisons conducted – six for digraph, eight for trigraph and three for keywords (see Table 3).

Table 3. Profile comparison settings

Metric	Standard Deviations (S.D.)
Digraphs	0.6, 0.7, 0.8 S.D.
Trigraphs	0.5, 0.6, 0.7, 0.8 S.D. *0.5 added due to poo performance at 0.6 and 0.7*
Keyword	0.6, 0.7, 0.8 S.D.

Once the profile comparison was completed, the results were exported and a number of functions were used to derive 2-dimensional tables of data from the raw results from the comparator from which the FAR/FRR figures could be derived.

Following the basic analysis described in this section, a further modification was made to the comparator to determine how many keystrokes were needed before either the valid user was challenged or an impostor detected. The threshold for this challenge was based upon the best performance thresholds from the earlier trials and was initially set at an alert level of 70. The results from this trial using the digraph keylogger files at a threshold of 0.7 standard deviations is presented in Table 4. The results from this trial were somewhat variable, while some users had good results (e.g. user 7, 10 and 26), most user profiles had only moderately successful results. If we consider user 2, while 29/34 (85%) impostors were challenged in less than 100 digraphs, user 16 (when acting as an impostor against user 2's reference profile) was able to type over 40,000 digraphs before being challenged.

The results in Table 4 can also be considered in terms of the average number of keystrokes required before a challenge is issued. The results show that an average of 6,390 digraphs were accepted before an impostor was challenged compared with an average of 68,755 digraphs before the valid user was challenged. While these results seem to provide the appropriate differentiation between impostor and valid user, giving an impostor the opportunity to type over 6,000 digraphs presents a major security risk.

Table 4. Number of keypresses before a challenge

For this trial the False Rejection Rate (FRR) was fixed at 0% (i.e. the valid user would not be rejected by the system). The False Acceptance Rates (FAR's) were then calculated for each user at the deviation thresholds specified in Table 3, and are shown in Table 5.

Table 5. Results from single-metric measures

Standard Deviation	Digraph FAR								Trigraph FAR								Keyword FAR							
	0.6		0.7		0.8		0.5		0.6		0.7		0.8		0.5		0.6		0.7		0.8	
Unmatched Alert	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
User 1	2.9	2.9	2.9	8.6	2.9	8.6	8.6	17.1	8.6	8.6	8.6	25.7	8.6	31.4	97.1	97.1	97.1	91.4	91.4	91.4	91.4	91.4
User 2	0.0	2.9	2.9	5.7	2.9	11.4	34.3	0.0	34.3	34.3	34.3	5.7	31.4	8.6	2.9	2.9	2.9	2.9	2.9	5.7	5.7	5.7
User 3	0.0	0.0	2.9	5.7	5.7	11.4	74.3	5.7	62.9	62.9	54.3	20.0	40.0	25.7	0.0	8.6	8.6	20.0	28.6	28.6	28.6	28.6
User 4	0.0	0.0	0.0	0.0	0.0	2.9	5.7	0.0	5.7	5.7	5.7	0.0	5.7	0.0	5.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
User 5	0.0	0.0	0.0	8.6	0.0	11.4	28.6	2.9	28.6	28.6	28.6	14.3	28.6	22.9	8.6	11.4	11.4	14.3	14.3	20.0	20.0	20.0
User 6	0.0	2.9	0.0	8.6	0.0	22.9	28.6	2.9	28.6	28.6	28.6	5.7	25.7	11.4	34.3	11.4	5.7	2.9	8.6	8.6	8.6	8.6
User 7	0.0	0.0	0.0	0.0	0.0	2.9	34.3	0.0	28.6	28.6	25.7	0.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
User 8	11.4	5.7	2.9	5.7	2.9	40.0	28.6	17.1	28.6	28.6	25.7	28.6	25.7	37.1	0.0	0.0	0.0	2.9	2.9	8.6	8.6	8.6
User 9	2.9	2.9	2.9	0.0	0.0	14.3	17.1	5.7	17.1	17.1	17.1	11.4	14.3	20.0	22.9	25.7	25.7	34.3	34.3	34.3	34.3	34.3
User 10	0.0	0.0	0.0	8.6	2.9	14.3	37.1	0.0	31.4	31.4	28.6	2.9	22.9	8.6	2.9	0.0	0.0	2.9	0.0	0.0	0.0	0.0
User 11	0.0	0.0	2.9	14.3	17.1	22.9	82.9	2.9	65.7	65.7	51.4	0.0	11.4	11.4	37.1	5.7	2.9	2.9	2.9	28.6	28.6	28.6
User 12	0.0	5.7	0.0	5.7	0.0	14.3	2.9	14.3	2.9	2.9	2.9	22.9	2.9	28.6	22.9	25.7	31.4	40.0	40.0	40.0	40.0	40.0
User 13	8.6	11.4	5.7	11.4	14.3	11.4	40.0	2.9	40.0	40.0	37.1	11.4	37.1	11.4	8.6	8.6	11.4	14.3	14.3	14.3	14.3	14.3
User 14	2.9	2.9	2.9	8.6	0.0	34.3	8.6	8.6	8.6	8.6	8.6	22.9	5.7	28.6	0.0	2.9	5.7	5.7	5.7	5.7	5.7	5.7
User 15	0.0	0.0	0.0	2.9	0.0	8.6	0.0	14.3	0.0	0.0	0.0	17.1	0.0	20.0	22.9	17.1	25.7	25.7	25.7	25.7	25.7	25.7
User 16	0.0	5.7	0.0	2.9	0.0	5.7	22.9	0.0	20.0	20.0	14.3	2.9	14.3	5.7	0.0	0.0	2.9	2.9	2.9	14.3	14.3	14.3
User 17	0.0	0.0	2.9	11.4	14.3	25.7	48.6	2.9	48.6	48.6	45.7	8.6	40.0	11.4	5.7	5.7	5.7	5.7	5.7	11.4	11.4	11.4
User 18	8.6	11.4	17.1	8.6	17.1	25.7	48.6	17.1	42.9	42.9	40.0	0.0	37.1	5.7	28.6	8.6	11.4	14.3	20.0	17.1	17.1	17.1
User 19	0.0	2.9	2.9	5.7	2.9	8.6	65.7	0.0	62.9	62.9	62.9	8.6	62.9	11.4	5.7	14.3	20.0	2.9	5.7	22.9	22.9	22.9
User 20	0.0	0.0	0.0	5.7	5.7	22.9	42.9	0.0	40.0	40.0	31.4	5.7	28.6	8.6	0.0	2.9	2.9	2.9	5.7	5.7	5.7	5.7
User 21	0.0	0.0	0.0	0.0	2.9	2.9	40.0	0.0	28.6	28.6	17.1	2.9	5.7	2.9	0.0	5.7	5.7	5.7	5.7	8.6	8.6	8.6
User 22	2.9	8.6	5.7	8.6	8.6	25.7	65.7	0.0	57.1	57.1	45.7	5.7	31.4	17.1	0.0	5.7	5.7	5.7	5.7	11.4	11.4	11.4
User 23	68.6	54.3	51.4	62.9	60.0	80.0	71.4	45.7	71.4	71.4	68.6	34.3	68.6	45.7	37.1	37.1	37.1	45.7	45.7	45.7	45.7	45.7
User 24	2.9	0.0	2.9	5.7	5.7	8.6	62.9	2.9	54.3	54.3	25.7	14.3	14.3	28.6	0.0	2.9	2.9	8.6	8.6	14.3	14.3	14.3
User 25	0.0	2.9	0.0	2.9	0.0	14.3	14.3	0.0	8.6	8.6	8.6	2.9	8.6	0.0	2.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
User 26	0.0	0.0	0.0	0.0	0.0	5.7	22.9	0.0	22.9	22.9	20.0	0.0	17.1	0.0	5.7	0.0	0.0	5.7	5.7	5.7	5.7	5.7
User 27	2.9	2.9	2.9	17.1	2.9	48.6	11.4	11.4	11.4	11.4	11.4	28.6	11.4	45.7	8.6	2.9	2.9	5.7	5.7	5.7	5.7	5.7
User 28	2.9	2.9	2.9	5.7	2.9	17.1	42.9	8.6	42.9	42.9	42.9	17.1	42.9	14.3	8.6	17.1	17.1	28.6	28.6	37.1	37.1	37.1
User 29	5.7	5.7	2.9	22.9	20.0	48.6	48.6	11.4	42.9	42.9	42.9	20.0	42.9	22.9	11.4	8.6	8.6	14.3	14.3	31.4	31.4	31.4
User 30	14.3	8.6	14.3	14.3	11.4	20.0	22.9	20.0	22.9	22.9	22.9	22.9	20.0	37.1	45.7	28.6	28.6	28.6	28.6	8.6	8.6	8.6
User 31	0.0	0.0	2.9	2.9	5.7	8.6	0.0	0.0	0.0	0.0	0.0	2.9	0.0	5.7	0.0	0.0	0.0	2.9	2.9	5.7	5.7	5.7
User 32	0.0	0.0	2.9	8.6	11.4	20.0	54.3	0.0	0.0	0.0	0.0	8.6	0.0	14.3	2.9	5.7	5.7	14.3	14.3	14.3	14.3	14.3
User 33	17.1	14.3	17.1	14.3	22.9	28.6	88.6	77.1	82.9	82.9	65.7	54.3	57.1	71.4	80.0	68.6	68.6	20.0	20.0	28.6	28.6	28.6
User 34	5.7	2.9	11.4	14.3	14.3	14.3	60.0	0.0	48.6	48.6	42.9	0.0	37.1	2.9	97.1	97.1	97.1	97.1	97.1	97.1	97.1	97.1
User 35	22.9	14.3	5.7	34.3	25.7	54.3	71.4	25.7	65.7	65.7	65.7	25.7	62.9	45.7	34.3	5.7	11.4	11.4	11.4	14.3	14.3	14.3
Average	5.2	5.0	4.9	9.8	8.1	20.5	38.2	9.1	33.3	33.3	29.5	13.0	25.2	18.9	18.3	15.2	15.2	16.5	16.5	20.2	20.2	20.2

When the results were calculated, the False Acceptance Rates per user were averaged across all users to provide an average FAR for each metric. The averaged results for the statistical approach are shown in Table 6. It

should be noted that the keyword latencies did not use the unmatched alert increase due to the use of a word list/dictionary – i.e. many words would not be matched in the users' profile.

Table 6. Final statistical results (best results highlighted)

Metric	S.D.	Unmatched Alert	FAR
Digraphs	0.6	No	5.2%
	0.6	Yes	5.0%
	0.7	No	4.9%
	0.7	Yes	9.8%
	0.8	No	8.1%
	0.8	Yes	20.5%
Trigraphs	0.5	No	38.2%
	0.5	Yes	9.1%
	0.6	No	33.3%
	0.6	Yes	33.3%
	0.7	No	29.5%
	0.7	Yes	13.0%
	0.8	No	25.2%
	0.8	Yes	18.9%
Words	0.5	No	18.3%
	0.6	No	15.2%
	0.7	No	16.5%
	0.8	No	20.2%

5. DISCUSSION

While the results shown in Table 6 show some encouraging FAR levels there is still significant variation with the best results obtained at 0.7 standard deviations for digraphs, 0.5 standard deviations for trigraphs (with increased alert levels for unmatched digraphs) and 0.6 for keywords. However, when the full results are considered (as shown in Table 5), even at the optimum settings, certain users show high FAR levels (e.g. user 23's profile returned FAR levels of 51.4%, 45.7% and 37.1% respectively for digraph, trigraph and keywords at the average optimum settings). It can also be clearly observed that the results for trigraphs and keywords are significantly worse when compared with those for digraphs – this is most likely to be related to the number of underlying samples used for these techniques (i.e. the number of sampled digraphs were significantly higher than that for trigraphs and keywords, with a corresponding increase of samples per digraph). It is probable that over a longer period of time, the profiles could be refined for trigraphs and keywords to produce a more distinct user profile with a corresponding reduction in the FAR.

These results also demonstrate that the techniques can be very effective for some users while very ineffective for others. For example, when considering digraph FAR's at 0.6 standard deviations (where 0% FAR was actually experienced for 19 out of the 35 users – 54.3%) the average FAR (5.2%) has been heavily influenced by a single user (user 23) whose 68.6% FAR dramatically increases the average. In a full implementation, the authors propose that the use of keystroke analysis should only form a part of a comprehensive user monitoring system. As such, a users' typing would only be monitored if the method was shown to be a discriminating authentication technique for that user. The removal of user 23 from the results in Table 5 significantly affects the average FAR's presented in Table 6, reducing the best digraph results from 4.9% to 3.5% , trigraph results from 9.1% to 8.0% and keywords from 15.2% to 14.5%.

Further optimisation can be achieved by removing the worst 5 participants (15%) from the trial results. This provides a significant improvement in the results of the technique with average FAR's as low as 1.7% for digraphs, 4.4% for trigraphs and 12.8% for keywords (Table 7). While the keyword FAR in particular remains unacceptably high, a reference back to Table 5 reveals that there were still almost a third of users for whom 0% FAR was observed at the 0.5 standard deviation threshold. This suggests a clear potential for using the technique in a subset of cases – which could also increase if additional keyword typing samples were obtained to support the profiling.

Table 7. Optimised results

Metric	S.D.	Unmatched Alert	FAR
Digraphs	0.6	No	1.7%
	0.6	Yes	2.4%
	0.7	No	2.2%
	0.7	Yes	7.0%
	0.8	No	4.9%
	0.8	Yes	17.0%
Trigraphs	0.5	No	34.5%
	0.5	Yes	4.4%
	0.6	No	29.3%
	0.6	Yes	29.3%
	0.7	No	25.6%
	0.7	Yes	10.6%
	0.8	No	21.2%
	0.8	Yes	15.2%
Words	0.5	No	13.8%
	0.6	No	12.8%
	0.7	No	15.3%
	0.8	No	19.7%

The removal of a number of specific user accounts from the keystroke monitoring process is not an ideal solution to the problem of poor user authentication. Keystroke analysis is unlikely to be used as a sole-method of user authentication, instead, it is envisaged that the methods described in this paper would form a part of a larger authentication system and would be only one of a range of authentication metrics that each user could be monitored with. With a larger number of users (and hence a wider range of user typing abilities and corresponding authentication rates) there is likely to be a proportional increase in the number of users for whom keystroke analysis does not produce appropriate FAR/FRR rates – in these cases other, more appropriate, techniques would have to be used. Identifying the cause of poor user performance when using keystroke analysis is vital; on-going work within the authors' research group will conduct further analysis on the gathered data sets to try to determine the cause of the variation between users and identify common factors (e.g. users' typing abilities, differences between application usage etc.).

6. CONCLUSIONS

It is clear from the results presented in this paper that there is considerable potential for continuous user authentication based on keystroke analysis. The long-term sampling of digraph keystrokes has served to reinforce the validity of the technique, while the introduction of trigraph and keyword monitoring has provided additional metrics that can be used as alternative (or complimentary) techniques. In particular, the use of keyword monitoring has considerable potential when used to monitor for specific, high-risk typed words (e.g. delete, format etc.).

It is also clear that the simple statistical approach does not provide sufficient distinction for all users and a live implementation would have to consider which metric (if any) is most appropriate for each user. It is envisaged that keystroke analysis would become only one of a number of monitoring characteristics used by a more comprehensive system with other authentication and supervision techniques.

Future work will also consider how the individual keystroke metrics can be combined together. For example, by combining the confidence measures of multiple metrics (e.g. monitoring digraphs and trigraphs), coupled with monitoring specific keywords (e.g. the typing patterns for high-risk words – format, delete etc.), it may be possible to provide a higher level of

confidence in the authentication of the user. The potential for this method will be considered in a later paper.

7. ACKNOWLEDGMENTS

The authors would like to thank the trial participants who assisted in the collection of the keystroke data. In particular, we would like to acknowledge the assistance of the staff and researchers of the Network Research Group, the Department of Psychology and the staff of TMA Global and John Nicholls Builders.

8. REFERENCES

Brown M. and Rogers S.J., 1993, "User identification via keystroke characteristics of typed names using neural networks", *International Journal of Man-Machine Studies*, vol. 39, pp999-1014.

Card S.K., Moran T.P. & Newell A., 1980. "Computer text-editing: An information-processing analysis of a routine cognitive skill", *Cognitive Psychology*, vol. 12, pp32-74.

Dowland P.S., Furnell S.M. & Papadaki M., 2002, "Keystroke Analysis as a Method of Advanced User Authentication and Response", *Proceedings of IFIP/SEC 2002 - 17th International Conference on Information Security*, Cairo, Egypt, 7-9 May, pp215-226.

Furnell, S.M., 1995, "Data Security in European Healthcare Information Systems", PhD Thesis, University of Plymouth, UK.

Gaines R., Lisowski W., Press S. and Shapiro N., 1980, "Authentication by Keystroke Timing: some preliminary results", *Rand Report R-256-NSF*, Rand Corporation.

Joyce R. and Gupta G. 1990. "Identity authentication based on keystroke latencies", *Communications of the ACM*, vol. 33, no. 2, pp168-176.

Legett J., Williams G., Usnick M. and Longnecker M., 1991, "Dynamic identity verification via keystroke characteristics", *International Journal of Man-machine Studies*, vol. 35, pp859-870.

Mahar D., Napier R., Wagner M., Laverty W., Henderson R.D. and Hiron M., 1995, "Optimizing digraph-latency based biometric typist verification systems: inter and intra typist differences in digraph latency distributions", *International Journal of Human-Computer Studies*, vol. 43, pp579-592.

Keystroke Analysis as a Method of Advanced User Authentication and Response

P.S.DOWLAND, S.M.FURNELL and M.PAPADAKI

nrg@plymouth.ac.uk
Network Research Group
Department of Communication and Electronic Engineering
University of Plymouth
Drake Circus
PLYMOUTH
PL4 8AA
United Kingdom
Tel: +44 1752-233521 Fax: +44 1752-233520

Key words: Keystroke Analysis, User Authentication, Biometrics, Intrusion Response.

Abstract: There has been significant interest in the area of keystroke analysis to support the authentication of users, and previous research has identified three discrete methods of application; static, periodic dynamic and continuous dynamic analysis. This paper summarises the approaches and metrics arising from previous work, and then proceeds to introduce a new variation, based upon application-specific keystroke analysis. The discussion also considers the use of keystroke analysis as a progressive, escalating response measure in the context of a comprehensive user authentication and supervision system, presenting an example of how this could be realised in practice.

1. INTRODUCTION

The issue of user authentication in IT systems has long been recognised as a potential vulnerability, with the majority of current systems relying upon password methods. Such methods have been repeatedly proven to be open to compromise, and can also be considered problematic in the sense that they typically only serve to facilitate a one-off authentication judgement at the start of a session. A number of previous works [1, 2, 3] have

consequently discussed the need for some form of monitoring to continuously (or periodically) authenticate the user in a non-intrusive manner. Although such monitoring is technically feasible, there are significant issues to be considered in selecting appropriate attributes to assess. This is particularly important, as continuous monitoring must be transparent to the end user in order to minimise any perceived inconvenience (with the exception of appropriate challenges in the event of suspected impostor activity).

A number of studies have considered the application of keystroke analysis to the problem of inadequate user authentication in modern IT system using static [4, 5, 6] and dynamic [7, 8] implementations. While these studies have evaluated the effectiveness of the proposed solutions, none have considered the implementation and necessary supporting application framework to effectively use keystroke analysis as a viable authentication and supervision mechanism.

This paper summarises the potential approaches to keystroke analysis, and presents details of a new method based on application-specific user profiling. It then proceeds to consider how keystroke analysis may be utilised as part of an intrusion response framework.

2. KEYSTROKE ANALYSIS OVERVIEW

Previous studies have identified a selection of data acquisition techniques and typing metrics upon which keystroke analysis can be based. The following section summarises the basic methods and metrics that can be used.

- **Static at login** - Static keystroke analysis authenticates a typing pattern based on a known keyword, phrase or some other pre-determined text. The captured typing pattern is then compared against a profile previously recorded during system enrolment. Static keystroke analysis is generally considered to be an initial login enhancement as it can supplement the traditional username/password login prompt, by checking the digraph latencies of the username and/or password components (i.e. authenticating the user on the basis of both *what* they typed and *how* they typed it).

- **Periodic dynamic** - Dynamic keystroke analysis authenticates a user on the basis of their typing during a logged in session. The captured session data is compared to an archived user profile to determine deviations. In a periodic configuration, the authentication

judgement can be intermittent; either as part of a timed supervision, or, in response to a suspicious event or trigger. This method provides distinct advantages over the static approach. Firstly, it is not dependent on the entry of specific text, and is able to perform authentication on the basis of any input. Another factor is the availability of data; in static keystroke analysis, the range of digraphs and frequency of their occurrence is likely to be significantly limited compared with a dynamic approach. Even an inexperienced typist is likely to produce sufficient digraph pairs to allow an authentication judgement to be derived. This is an important factor as it is necessary to have a statistically significant volume of keystroke data in order to generate a user profile.

- **Continuous dynamic** - Continuous keystroke analysis extends the data capturing to the entire duration of the logged in session. The continuous nature of the user monitoring offers significantly more data upon which to base the authentication judgement. With this method it is possible that an impostor may be detected earlier in the session than under a periodically monitored implementation. On the downside, however, the additional processing required will add to the computational overhead of the supervision system.

- **Keyword-specific** - Keyword-specific keystroke analysis extends the continuous or periodic monitoring to consider the metrics related to specific keywords. This could be an extra measure incorporated into a monitoring system to detect potential misuse of sensitive commands. For example, under a DOS/Windows environment it may be appropriate to monitor the keystroke metrics of a user attempting to execute the FORMAT or DELETE commands. This could represent a significant enhancement, as a command with a high misuse consequence (e.g. DEL *.*) is unlikely to cause sufficient profile deviation when observed from a system-wide context, due to the limited selection of digraphs. By contrast, static analysis could be applied to specific keywords to obtain a higher confidence judgement.

- **Application-specific** - Application-specific keystroke analysis further extends the continuous or periodic monitoring. Using this technique, it may be possible to develop separate keystroke profiles for distinct applications. For example, a user may be profiled separately for their word processing application and their email client. The potential of this new technique is discussed in more detail in section 3.

In addition to a range of implementation scenarios, there are also a variety of possible keystroke metrics that can be profiled as the basis for subsequent comparison:

- **Digraph latency** - Digraph latency is the metric that has traditionally been used for previous studies, and typically measures the delay between the key-up and the subsequent key-down events, which are produced during normal typing (e.g. T-H). In most cases, some form of low and high pass filter is applied to remove extraneous data from the session data.

- **Trigraph latency** - Trigraph latency extends the previous metric to consider the timing for three successive keystrokes (e.g. T-H-E).

- **Keyword latency** - Keyword latencies consider the overall latency for a complete word or may consider the unique combinations of digraph/trigraphs in a word-specific context.

- **Mean error rate** - The mean error rate can be used to provide an indication of the competence of the user during normal typing. Whilst this may not be user specific, it may be possible to classify users into a generic category, according to their typing ability, which can then be used as an additional authentication method.

- **Mean typing rate** - A final metric is that of the mean typing rate. As with the mean error rate, individuals can be classified according to their typing ability and hence evaluated based on their average typing speed.

While the final two metrics indicated above are unlikely to provide a suitably fine-grained classification of users for direct authentication judgements, they may be used to provide a more generic set of user categories that can contribute to a combined measure.

It should be noted that all of the above techniques and metrics can be implemented on a standard PC platform, without the need for special hardware.

3. EXPERIMENTAL DYNAMIC KEYSTROKE ANALYSIS

The idea of using keyboard characteristics for authentication is not unique, and there have been a number of previous published studies in the area. To date, however, virtually all published studies have focussed upon static or context-independent dynamic analysis, using the inter-keystroke latency timing method. From the earliest studies in 1980 [9], the focus has been on the analysis of digraph latencies. Later studies [6, 8] further enhanced the work, identifying additional statistical analysis methods that provided more reliable results.

In [7], the concept of dynamic keystroke analysis was first proposed, with the introduction of a reference profile that could be used to monitor a live user session. Brown and Rogers [5] also explored the idea of dynamic analysis, presenting preliminary results.

A summary of some of the main results from studies to date is presented in Table 1 below, which illustrates the effectiveness observed (in terms of false acceptance and false rejection errors), as well as the type of keystroke analysis technique employed (digraph/trigraph etc.) and the analysis approach taken (statistical/neural network etc.).

Authors	Method	%FAR	% FRR
Umphress & Williams (1985) [10]	Digraph Statistical	6%	12%
Legget & Williams (1988) [11]	Digraph Statistical	5%	5.5%
Joyce & Gupta (1990) [6]	Digraph Statistical	0.25%	16.67%
Bleha et al. (1990) [12]	Digraph Statistical	2.8%	8.1%
Legget et al. (1991) [7] [1]Static, [2]Dynamic	Digraph Statistical	5% [1] 12.8% [2]	5.5% [1] 11.1% [2]
Brown & Rogers (1993) [5] [1]Group 1, [2]Group 2	Digraph Combined Neural Network & Statistical	0%	4.2% [1] 11.5% [2]
Napier et al. (1995) [13]	Digraph Statistical	29.5% / 3.8%	
Mahar et al. (1995) [8]	Digraph Statistical	35% / 17.6%	
Furnell et al. (1996) [14] [1]Static, [2]Dynamic	Digraph Neural Network [1], Statistical [2]	8% [1] 15% [2]	7% [1] 0% [2]

Table 1: Previous keystroke analysis studies

A further variation in the data analysis can be introduced through the consideration of application specific keystroke profiles. If we accept from previous work that individual users have a distinct typing pattern, it can be hypothesised that an individual's typing pattern may also vary depending upon the application in use. For example, a user participating in a chat session may type in a fairly relaxed style, while the same user may type in an significantly different way when producing a document. It should also be noted that certain categories of user might use the numeric keypad for large quantities of data entry. Under these circumstances the volume and diversity of the keystroke digraphs will vary tremendously when compared to the more usual alphanumeric typing encountered with most user profiles. Previous research has been carried out in this area [15], which has shown that analysis of numeric keystrokes can provide a viable authentication measure. This is an area receiving on-going attention through a separate research project at the authors' institution.

In [16] the authors described a trial in which keystroke data, obtained within Microsoft Windows NT, was evaluated across all applications. While the results from this trial were encouraging, the quantity of data collected was insufficient to make a true, statistically valid, conclusion. Instead it was determined that further trials were necessary. Following the first trial, the authors conducted a second round of monitoring in which eight test subjects were profiled. Over a period of 3 months, a total of 760,000 digraph samples were captured and stored for analysis. In this case, however, the analysis was conducted with a view to determining viability of application-specific keystroke profiling. To this end, it was necessary to identify a series of applications for profiling, with the selection criteria being those for which sufficient keystroke data had been logged during the sampling period. A review of the keystroke data revealed that the applications satisfying this requirement were Microsoft MSN Messenger, Internet Explorer, Word and PowerPoint. While the authors considered that a numerically intensive application such as Excel would have provided an interesting candidate, insufficient keystrokes were captured to enable the creation of a profile. Additionally, of the eight users sampled during the trial, only five produced sufficient data to analyse from all of the aforementioned applications. Although the resulting sample group was very small, it was sufficient to yield interesting results in relation to an initial assessment of application-specific profiling.

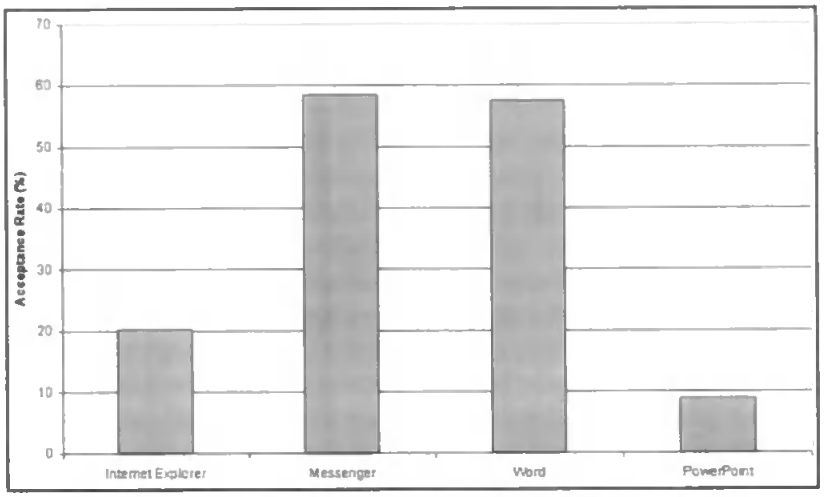


Figure 1: Acceptance Rate for application specific keystroke data compared against a system-wide context user profile

In Figure 1 above, a single user's application-specific keystroke data is compared against the reference profile from the same user. The reference profile was based on all keystroke data acquired from all applications. Although the figure does not show distinct differences in all cases, there is a clear distinction between all applications apart from Messenger and Word. This can be explained when the nature of these applications is considered. Messenger and Word are both significantly textual in their usage, and users will typically type within Messenger and/or Word for considerable periods of time. In contrast, while Internet Explorer and PowerPoint sessions may both involve significant elements of keyboard activity, the typing is more likely to occur in sporadic bursts. As such, any dynamic that emerges is likely to be markedly different to that which would emerge in applications where more sustained typing is the norm. Considering the information portrayed above, the creation of application specific profiles would be likely to increase the acceptance rates observed.

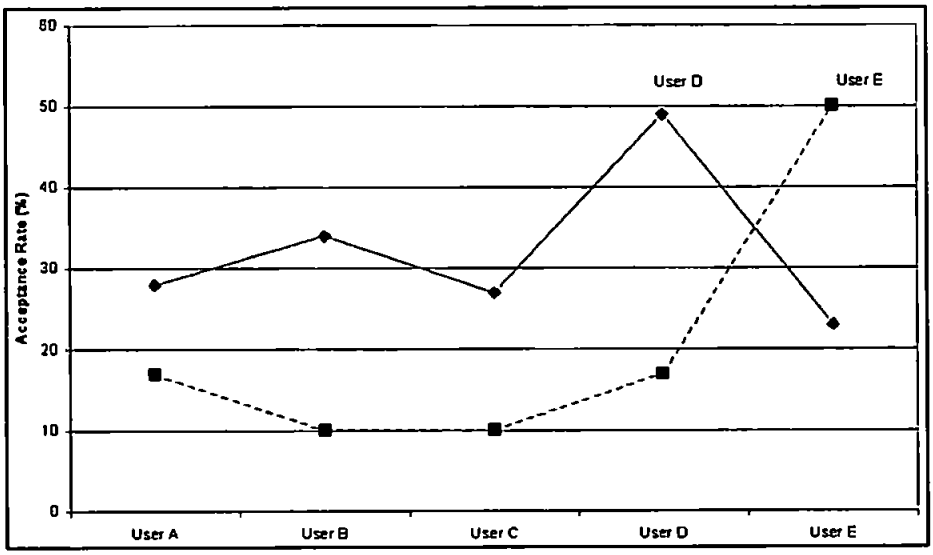


Figure 2: Acceptance Rate for two user profiles

In Figure 2 above, a specific users' profile (users D and E when using Internet Explorer) is examined, showing there is a clear difference between other users' keystroke data (impostors) with appropriate peaks in acceptance rate for the valid users.

While the results shown do not indicate a suitably discriminative metric upon which to base a satisfactory authentication judgement, they do show a level of correlation between a user's typing pattern in an application-specific context. These preliminary results show that further work is needed to investigate the use of application-specific keystroke analysis.

4. AN ESCALATING RESPONSE FRAMEWORK USING KEYSTROKE ANALYSIS

The earlier discussion summarised the different potential implementations of keystroke analysis, and explained the operational differences between the approaches. It is possible to integrate these analysis approaches into an overall user authentication and supervision framework, with the varying techniques being invoked as responses to anomalies detected at earlier stages. A possible example of this is illustrated in Figure 3, which shows how the five variations discussed earlier can be incorporated within a four-level response framework. It should be noted that this is by no means the only method by which the techniques could be combined, and

specific implementations could vary depending upon rule sets for a particular user, class of users, or general organisational security policy.

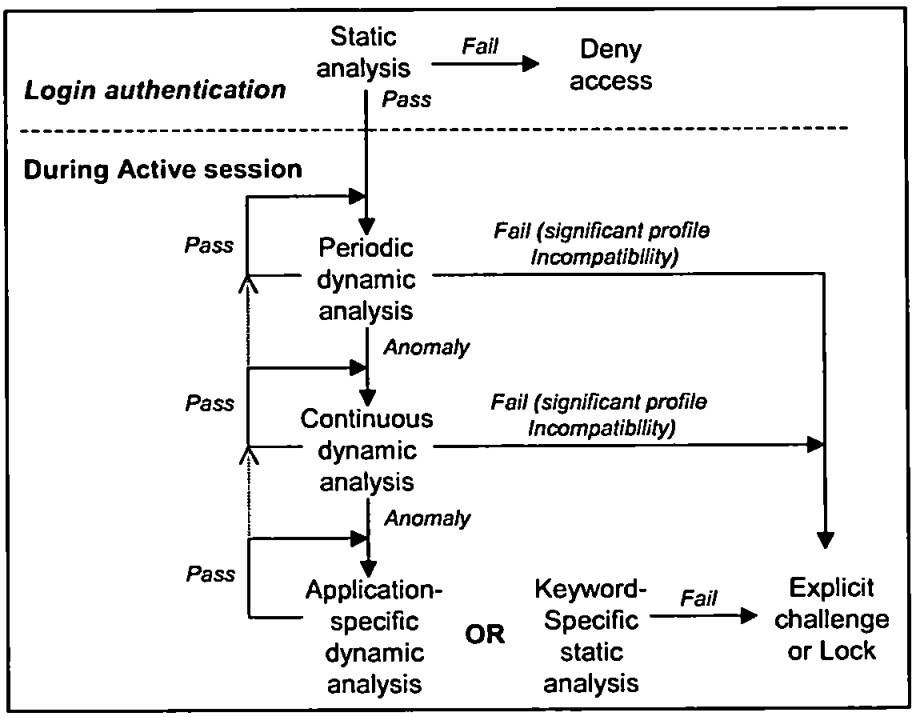


Figure 3: Response framework using keystroke analysis

A suitable architecture for achieving such an approach is offered by the Intrusion Monitoring System (IMS) [17]. This proposes an architecture for real-time user authentication and misuse detection, based upon a monitoring *Host* that has the responsibility for supervising a number of *Client* systems (e.g. in the form of end user PCs or workstations). Key elements of the architecture, from the perspective discussed in this paper are the *collector* (which obtains the keystroke data from the individual client systems), the *anomaly detector* (which performs the actual keystroke analysis and profile comparison, maintaining a consequent alert status metric), and the *responder* (which is responsible for initiating the different keystroke analysis approaches in response to increases in the alert status and other contextual factors). Assuming such a monitoring context, the text below describes how the response process in Figure 3 would proceed.

Initial authentication may occur using a standard username/password pair, but supplemented by the use of static keystroke analysis to assess how the information is entered. If the user fails to authenticate at this stage (e.g. after being permitted three attempts to enter the details), then the most appropriate

response is to deny access (if the correct password is provided, but the keystroke analysis aspect fails, then an alternative option could be to allow the login to proceed, but to begin the session with a higher level of subsequent monitoring – e.g. continuous rather than periodic assessment). If this login authentication is successful, the user will proceed to a logged in session, during which dynamic keystroke analysis could be applied on a periodic basis (in order to minimise the associated processing overhead in the initial instance). Assuming no anomalies, this could simply continue throughout a logged in session. If a departure from the typing profile is noted during the monitoring period, however, there would be two options for response. If the keystroke data exhibits a significant incompatibility, then a high confidence of impostor action could be assumed and the responder could proceed directly to some form of explicit action (e.g. interrupting the user session by issuing a challenge or suspending their activity pending an administrator intervention). In cases where the profile incompatibility is not conclusive, the responder could initiate an increase in the monitoring resolution – firstly to invoke continuous dynamic analysis, and then beyond this to invoke either application or keyword-specific methods. The choice in the latter case would depend upon the context of the current user's activity. For example, if they were word-processing, then application-specific dynamic analysis would potentially give a more accurate assessment of identity. If, by contrast, they were operating at a command line level, then it could be considered more appropriate to invoke keyword-specific static analysis, looking for instances of particularly sensitive commands such as 'format' or 'erase'. Profile incompatibility at this final stage would automatically result in more explicit response action.

In cases where the responder agent has initiated a more detailed level (e.g. from periodic to continuous, or from continuous to application-specific), then the monitoring would continue at this level for a period of time, in order to ensure that profile incompatibilities were no longer observed. A suitable trigger (e.g. the entry of a certain number of further keystrokes without significant profile departure) would be used to reduce the alert status of the monitoring system, and thereby allow the responder agent to re-invoke a lesser level of analysis (this is indicated by the dotted arrow lines in the figure).

The combination of mechanisms in this manner allows a system to provide a standard, and hence acceptable, user login for the initial authentication, while also providing enhanced user supervision for the duration of the users' session. Such a system should, in theory, ensure transparent operation to legitimate users. It should also be noted that, in a practical context, keystroke analysis may not be the only technique involved, and other metrics relating to user activity and behaviour might also be

considered by the *anomaly detector*, and thereby used to inform the *responder* agent.

5. CONCLUSIONS

This paper has considered the significant variety of implementation methods and metrics that can be associated with keystroke analysis. The new concept of application-specific analysis has been introduced, along with initial experimental findings that support the feasibility of the approach. The preliminary results suggest that the technique is worthy of further investigation.

The discussion has also considered the application of keystroke analysis as a response mechanism within an intrusion detection system. The combination of analysis techniques, placed within such an authentication/supervision framework has the potential to provide a significant improvement in system-wide security against impostor attacks, as well as ensuring transparency to legitimate end users.

6. REFERENCES

[1] Morrissey J.P.; Sanders P.W. & Stockel C.T. 1996. "Increased domain security through application of local security and monitoring"; Expert Systems; vol. 13; no. 4; pp296-305.

[2] Lunt T.F. 1990. "IDES: an intelligent system for detecting intruders"; Proceedings of the Symposium on Computer Security: Threat and Counter Measures"; Rome.

[3] Mukherjee B. & Heberlein L.T. 1994. "Network intrusion detection"; IEEE Networks; vol. 8; no. 3; pp26-45.

[4] Jobusch D.L. & Oldehoeft A.E. 1989. "A survey of password mechanisms: Weaknesses and potential improvements. Part 1"; Computers & Security; vol. 8; no. 7; pp587-603.

[5] Brown M. & Rogers S.J. 1993. "User identification via keystroke characteristics of typed names using neural networks"; International Journal of Man-Machine Studies; vol. 39; pp999-1014.

[6] Joyce R. & Gupta G. 1990. "Identity authentication based on keystroke latencies"; Communications of the ACM; vol. 33; no. 2; pp168-176.

[7] Legett J.; Williams G.; Usnick M. & Longnecker M. 1991. "Dynamic identity verification via keystroke characteristics"; International Journal of Man-machine Studies; vol. 35; pp859-870.

[8] Mahar D.; Napier R.; Wagner M.; Laverty W.; Henderson R.D. & Hiron M. 1995. "Optimizing digraph-latency based biometric typist verification systems: inter and intra typist differences in digraph latency distributions"; International Journal of Human-Computer Studies; vol. 43; pp579-592.

[9] Card S.K.; Moran T.P. & Newell A. 1980. "Computer text-editing: An information-processing analysis of a routine cognitive skill"; Cognitive Psychology; vol. 12; pp32-74.

[10] Umphress D. & Williams G. 1985. "Identity verification through keyboard characteristics"; International Journal of Man-Machine Studies; vol. 23; pp263-273.

[11] Legett J. & Williams G. 1988. "Verifying user identity via keystroke characteristics";International Journal of Man-Machine Studies; vol. 28; pp67-76.

[12] Bleha S.; Slivinsky C. & Hussein B. 1990. "Computer-access security systems using keystroke dynamics"; Actions on pattern analysis and machine intelligence; vol. 12; no. 12; pp1217-1222.

[13] Napier R.; Laverty W.; Mahar D.; Henderson R.; Hiron M. & Wagner M. 1995. "Keyboard user verification: towards an accurate, efficient, and ecologically valid algorithm"; International Journal of Human-Computer Studies; vol. 43; pp213-222.

[14] Furnell S.M.; Morrissey J.P.; Sanders P.W. & Stockel C.T. 1996. "Applications of keystroke analysis for improved login security and continous user authentication"; Proceedings of the 12th International Conference on Information Security (IFIP SEC '96), Island of Samos, Greece; 22-24 May, pp283-294.

[15] Ord T. & Furnell S.M. 2000. "User authentication for keypad-based devices using keystroke analysis"; Proceedings of the Second International Network Conference (INC 2000), Plymouth, UK, 3-6 July; pp263-272.

[16] Dowland P.S.; Singh H. & Furnell S.M. 2001. "A preliminary investigation of user authentication using continuous keystroke analysis"; Proceedings of the IFIP 8th Annual Working Conference on Information Security Management & Small Systems Security, Las Vegas; 27-28 September.

[17] Furnell S.M. & Dowland P.S. 2000. "A conceptual architecture for real-time intrusion monitoring"; Information Management & Computer Security; vol. 8; no. 2; pp65-74.

A Preliminary Investigation of User Authentication Using Continuous Keystroke Analysis

P. S. DOWLAND[1], H. SINGH[2], S. M. FURNELL[3]

[1]*pdowland@plymouth.ac.uk*
[2]*hsingh@jack.see.plym.ac.uk*
[3]*sfurnell@ plymouth.ac.uk*
Network Research Group
Department of Communication and Electronic Engineering
University of Plymouth
Drake Circus
PLYMOUTH
PL4 8AA
United Kingdom
Tel: +44 1752-233521 Fax: +44 1752-233520

Key words: Keystroke Analysis, Authentication, Biometrics, Data Mining.

Abstract: There has been significant research in to the provision of reliable initial-login user authentication, however there is still a need for continuous authentication during a user session. This paper presents a series of results from the preliminary statistical analysis of multi-application keystroke data. This has been contrasted with a Data Mining approach to the production of a unique user profile. This paper aims to determine which approach provides the best basis for further research. It is determined that the technique offers promise as a discriminator between individuals in an operational context, but further investigation with larger data sets is required with a combination of approaches being considered in order to improve the accuracy.

1. INTRODUCTION

There have been a number of previous studies that have considered the security weaknesses in modern IT system and, whilst various recommendations and technical solutions have been proposed, many still

rely on enhancing the initial login-stage mechanism (e.g. via biometric identification, smart cards etc.) [COPE 90, SHER 92, MILL 94]. Whilst this improves the initial authentication judgement, there is still a need for user authentication throughout a session. In most systems there is no further check on a users' identity beyond the initial username/password. Once a user gains legitimate access to IT resources, it is feasible for there to be no further challenge, with the only possibility for detection of a masquerader being the post-event detection of a major incident (i.e. an impostor can masquerade as the valid user without detection or challenge).

To counter this risk, it is suggested that some form of user monitoring is desirable to continuously (or periodically) authenticate the user in a transparent manner. Whilst such monitoring is technically feasible, there are significant issues to be considered in selecting appropriate attributes to assess. This is particularly important, as continuous monitoring must be transparent to the end user in order to minimise any perceived inconvenience (with the exception of appropriate challenges in the event of a significant profile deviation).

This paper specifically considers the problems of continuous user authentication using keystroke digraph latencies. This area has not received much attention and as such, most of the background research is based upon static keystroke analysis [JOBU 89, BROW 93, JOYC 90] (i.e. where the users' typing was constrained). Keystroke analysis is, however, considered by end users as the most acceptable form of continuous authentication [FURN 00]. A GUI environment produces a new challenge, as there is no option to control the users' typing. This can cause problems, as it is difficult to determine in which application individual digraph pairs were entered. This paper will introduce a statistical approach for detecting deviation from a user's historical keystroke profile captured under a multi-tasking windowed environment. Following this initial analysis, a Data Mining (DM) approach will be considered in order to determine the potential for improving user classification. It should be noted that the aim of this paper is to determine which approach provides the best basis for further research and is not intended as a thorough analysis of keystroke latencies for user authentication. Finally, some thoughts on future work are introduced which will be developed further.

2. EXPERIMENT OVERVIEW

Although there have been a series of papers describing the mechanisms for keystroke analysis, the authors have been unable to identify any research specifically focussed on continuous keystroke analysis in which the

collection of users typing samples was not artificially constrained in some way through a custom interface (e.g. asking the user to type known strings).

The experiment was designed to allow keystroke data to be collected under the Microsoft Windows NT environment. In order to collect the required data, it was necessary to implement a mechanism for acquiring keystroke notifications across all applications running within a users' active session. As the client systems were running Microsoft Windows NT v4.0, it was necessary to implement a system-wide hook function that would receive keyboard events through the Windows message chain. System-wide hooks allow a specified code block (hook-function) to receive the appropriate Windows messages (e.g. WM_KEYUP for the key-up event) irrespective of the target application (i.e. it is possible for a hook function residing in a system DLL to receive keystroke notifications for all currently running applications). This effectively allowed application keystroke data to be duplicated and directed towards the data logger on the client workstation. Technical details of the implementation of the hook function and its associated support files are beyond the scope of this paper and, as such, have been omitted. There are a number of resources available that provide further information for interested readers [DOWL 00, MICR 00]. In order to determine accurate digraph latencies, it was also necessary to implement a high-accuracy timer (as the default timers available do not offer adequate accuracy for the millisecond latencies expected).

To eliminate extreme short/long digraph latencies that may adversely affect the distribution of digraph times, any digraph pair whose latency fell outside a nominal range was excluded from the archived data. For the purposed of this experiment the range was restricted to times above 40ms and below 750ms. These thresholds are based on the original experiments carried out by the authors [FURN 95] and are designed to eliminate samples where two keys may have been accidentally struck together (thus, producing an infeasibly small latency) or, where the user may have made a pause in their typing and thus introduced an unnaturally large inter-keystroke latency. The output of this pre-processing was a data file containing the following structure:

first_char *second_char* *digraph_latency*

For this experiment a total of ten users were profiled. As the intention was to evaluate the analysis mechanisms without implementing a large-scale trial, tests were carried out using a small set of test subjects. The main limiting factor was the need to collect data over a prolonged period (weeks rather than hours). Despite the small scale of the trial, it still proved difficult to collect sufficient data in order to provide a valid comparison between users. Due to this limited set of data, analysis has focussed on the 4 main

users who provided the largest profiled data sets in order to best illustrate the trends observed.

3. STATISTICAL ANALYSIS

Following the pre-processing described in the previous section, the experimental data for each user was then processed off-line to calculate the mean and standard deviation values for each unique digraph pair. In the event that any digraph pair had a standard deviation greater than its mean value, the digraph samples were sorted and the top/bottom 10% were then removed with subsequent re-calculation of the mean and standard deviation values – this was only attempted where at least ten samples were available for the digraph pair. The reason for this additional step was to remove digraph samples where the latencies would have an adverse affect on the standard deviation (i.e. the distribution of samples was tightened).

Once a set of digraph pairs was produced (with corresponding mean/standard deviation digraph latency values), the user's profile was further constrained by filtering out digraph pairs where the sample count fell below a nominal threshold value. Our experiments fixed this value at fifty samples; however, the software used for analysis allows a variable threshold that will be investigated further in the future work described in a later section. A summary of the profiles generated by this method is shown in *Table 1*.

Table 1: Summary of user profile statistics

User	Unique Digraph Pairs	Filtered Digraph Pairs	Average Typing Speed
User A	466	122	151ms
User B	405	51	145ms
User C	412	89	206ms
User D	461	127	162ms

Once a user profile was generated, the profile was evaluated by comparison with the users' raw keystroke data. This allowed the test profile to be evaluated using the users' own data (to test the False Rejection Rate – FRR) and against other users' keystroke data (to test the False Acceptance Rate – FAR).

As there is likely to be significant variation in a users' own session data, a compensatory factor was applied to the standard deviation that could be varied in a "live" environment according to the security needs of the organisation. This factor allowed the number of standard deviations from the

mean to be adjusted. For the purposes of this experiment, four weightings were considered, namely 0.5, 1, 1.5 and 2. This produced an acceptable digraph range:

digraph range = mean ± (standard deviation * weighting factor)

When viewing the preliminary results (*Figure 1*), if we consider the four users A, B, C and D and follow the vertical columns of data, we can see a clear peak for each users data when compared with their own profile. This is most noticeable for user C where a significant peak is observed (50% of all digraphs accepted) compared with 35% when user B's digraph data was tested against the same profile.

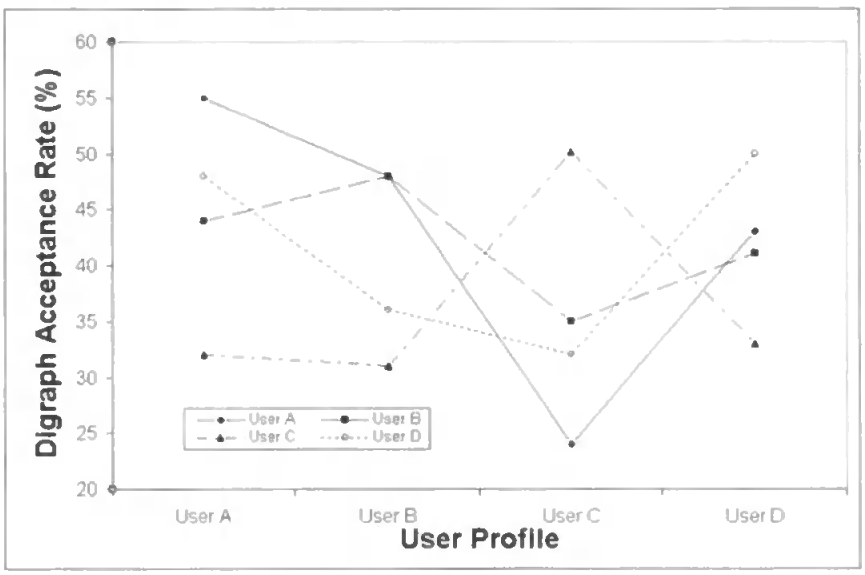


Figure 1: User profile comparisons

Although there was a clear correlation between user C's profile and data, if we consider user A, there was a high FAR for data from users D and B (impostors) when compared with user A's profile. We can also see that in user B's profile the impostor "user A" achieved the same acceptance rate (48%). It is clear from these results that an additional measure of acceptance/rejection is required. To further test the FAR/FRR of the test system, the analysis software monitored the number of consecutively rejected digraph pairs – representing the highest alert level of the system (*Figure 2*).

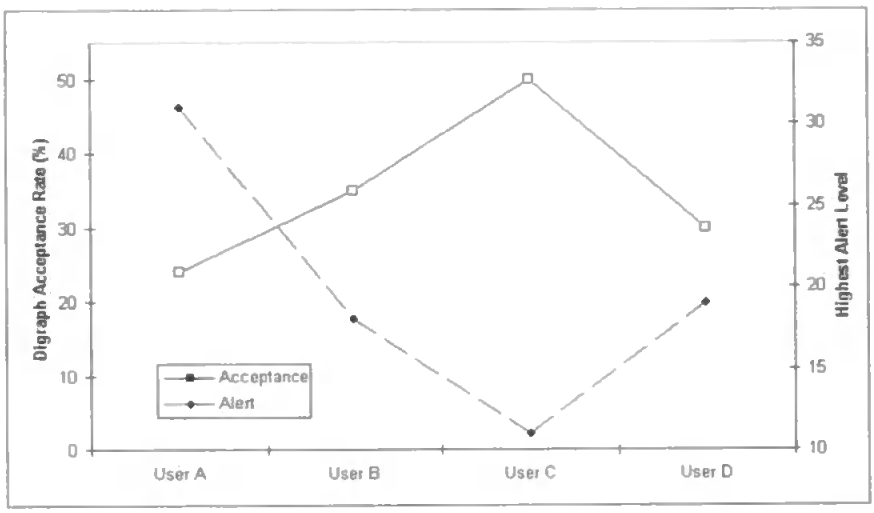


Figure 2: Single user profile comparison

When considering (*Figure 2*) we can identify two distinct trends. Firstly, the top line plots the digraph acceptance rate for all user data sets against user C's profile. Here we can see a clear peak correlating to user C's own data and corresponding reductions in the acceptance rates for the other users' data. Secondly, the lower line indicates the highest alert level detected by the analysis software. This is simply a record of the highest count of consecutively rejected digraph times (excluding non-profiled digraph pairs). Again, we can see a correlation between user C's own data when compared with their profile and corresponding increases in the alert level as impostor data sets are compared with the target profile.

4. DATA MINING ANALYSIS

The methodology described in the previous sections, using traditional statistical approaches, requires a significant level of manual intervention in the data analysis stages. Further, it is time consuming when considering the amount of data generated from a single session or multiple sessions and the number of users on a system. From this we can determine there is a need to automate some of the data analysis pre-processing stages. These stages offer the opportunity to investigate Data Mining (DM) methodology and algorithms, a previously untried approach in this field, in order to eliminate the manual approaches adopted and also to compare the FAR/FRR percentage accuracy with the statistical approach. Data Mining can be described as a collection of techniques and methodologies used to explore

vast amounts of data in order to find potentially useful, ultimately understandable patterns [FAYY 96] and to discover relationships. The methodology used to analyse the raw keystroke data is derived from the four main activities of DM; selection, pre-processing, data mining and interpretation [FAYY 96]. DM is an iterative and interactive process, involving numerous steps with many decisions being made by the user. Different algorithms are optimised based on the predefined DM task. This involves deciding whether the goals of the DM process are classification, association, or sequential [MICH 94].

For the purpose of this work, the data sets were split into a ratio of 9:1 hence into two parts; a training set and a testing set, which is a commonly used technique known as train and test. The Intelligent Data Analysis (IDA) Data Mining Tool [SING 99] is used to analyse the sample data sets which incorporates algorithms from the fields of Statistical, Machine Learning and Neural Networks. Six algorithms, k-NN, COG, C4.5, CN2, OC1 and RBF were chosen for this investigative work. The algorithm or classifier is subjected initially with the training set and then the classification accuracy is tested using the unseen data set or testing set. The results give an indication of the error rate (or FAR) and the overall classification accuracy of the trained algorithms.

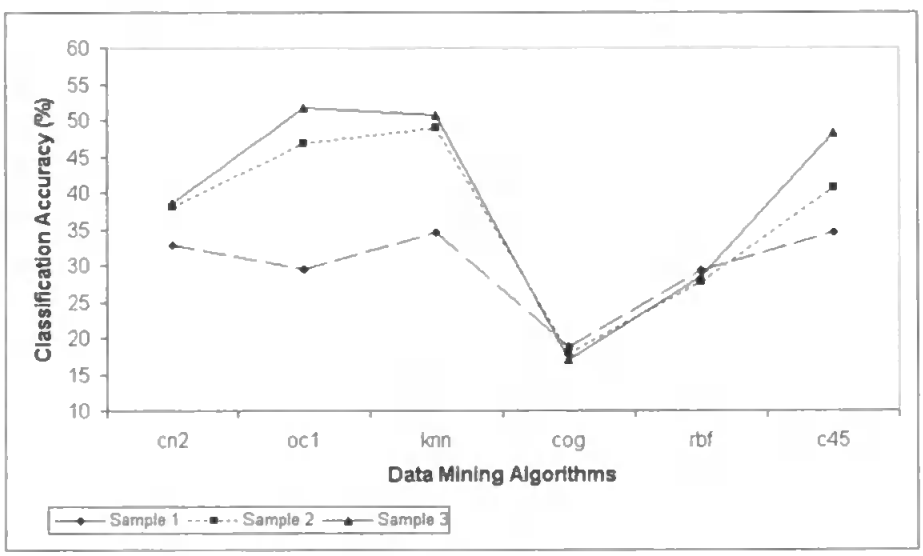


Figure 3: Varying sample sizes with fixed number of classes and attributes

The percentage classification accuracy obtained is encouraging as depicted in (*Figure 3*), which shows that when the sample size is increased, the classification accuracy obtained increases proportionally, except for the

COG a statistical based algorithm and RBF a Neural Network based algorithm. This is important when considering the size of data being analysed and hence eliminates the ad-hoc approaches adopted using traditional statistical methods.

The initial results suggest that Machine Learning (OC1 and C4.5) and Statistical (k-NN) based algorithms are suitable for these types of data sets. Despite the results, more work needs to be carried out in order to correlate the results to a specific or group of algorithm(s), in order to obtain a higher percentage of classification accuracy.

5. CONCLUSIONS

It is clear from the results presented in this paper that there is some potential for continuous user authentication based on keystroke analysis. However, it is also clear that a simple statistical approach does not provide sufficient distinction between users. The DM approach is limited due to the nature of the data gathered and will also require further research. It is proposed that further work will investigate the usefulness of trigraph keystroke combinations (timings for three consecutive keystrokes) and the possible use of word-graph timings (timings for frequently occurring words). Further analysis will be carried out on much larger data sets in order to give a higher statistical reliability and will also incorporate high-level characteristics (average typing speed and typing error rates) which will provide additional information to the system-characteristic based DM approach being developed in parallel with this research [SING 01]. Other approaches that will be investigated include, consideration of various standard deviation weightings, varying the minimum number of samples for profiled digraphs and varying the inclusion threshold for each sampled digraph. A further possibility for research may be an investigation into a correlation between digraph latencies and the applications in which they were generated (i.e. application specific keystroke profiles).

This paper has presented a series of results from the preliminary statistical analysis of multi-application keystroke data. This has been contrasted with a DM approach to the production of a unique user profile. Whilst the results from this stage of the research are not as encouraging as we had hoped for, they have shown a potential for the use of continuous user authentication. The next phase will concentrate on a combination of techniques to improve the digraph acceptance rate seen in these results.

6. REFERENCES

[COPE 90] Cope J.B.; "Biometric systems of access control"; Electrotechnology; pp71-74; April/May 1990.

[BROW 93] Brown M. & Rogers S.J.; "User identification via keystroke characteristics of typed names using neural networks"; International Journal of Man-Machine Studies; pp999-1014; 1993.

[DOWL 00] Dowland P.S. & Furnell S.M.; "Enhancing Operating System Authentication Techniques"; Proceedings of the International Network Conference 2000 (INC2000); pp253-261; July 2000.

[FAYY 96] Fayyad U.M.; "Data Mining and Knowledge Discovery: making sense out of data"; IEEE Expert; vol. 11; no. 6; pp20-25; 1996.

[FURN 95] Furnell S.M.; "Data security in European healthcare information systems"; PhD Thesis; University of Plymouth, UK; 1995.

[FURN 00] Furnell S.M., Dowland P.S., Illingworth H.M. & Reynolds P.L.; "Authentication and Supervision: A survey of user attitudes"; Computers & Security; vol. 19; no. 6; pp519-539; 2000.

[JOBU 89] Jobusch D.L. & Oldehoeft A.E.; "A survey of password mechanisms: Weaknesses and potential improvements. Part 1"; Computers & Security; p587-603; 1989.

[JOYC 90] Joyce R. & Gupta G.; "Identity Authentication Based on Keystroke Latencies"; Communications of the ACM; vol. 33; no. 2; pp168-176.

[MICH 94] Michie D., Spiegelhalter D.J. & Taylor C.C.; "Machine Learning, Neural and Statistical Classification"; Ellis Horwood; ISBN 0-13-106360-X; pp136-141; 1994.

[MICR 00] Microsoft Corporation; "Monitoring System Events"; 2000; http://msdn.microsoft.com/library/psdk/winbase/hooks_9rg3.htm

[MILL 94] Miller B.; "Vital Signs of Identity"; IEEE Spectrum; February; 1994.

[SHER 92] Sherman R.L.; "Biometrics Futures"; Computers and Security; vol. 11; no. 2; pp128-133; 1992.

[SING 99] Singh H., Burn-Thornton K.E. & Bull P.D.; "Classification of Network State Using Data Mining"; Proceedings of the 4th IEEE MICC & ISCE '99; Malacca, Malaysia; vol. 1; pp183-187; 1999.

[SING 01] Singh H., Furnell S.M., Lines B.L. & Dowland P.S.; "Investigating and Evaluating Behavioural Profiling and Intrusion Detection Using Data Mining"; Proceedings of the International Workshop on Mathematical Methods, Models and Architectures for Computer Networks Security (MMM) 2001; St Petersburg, Russia; 21-23 May 2001.

Authentication and supervision: A survey of user attitudes

S.M. Furnell, P.S. Dowland, H.M. Illingworth and P.L. Reynolds

Abstract

User authentication is a vital element in ensuring the secure operation of IT systems. In the vast majority of cases, this role is fulfilled by the password, but evidence suggests that this approach is easily compromised. Whilst many alternatives exist, particularly in the form of biometric methods, questions remain over the likely user acceptance. This paper presents the results of a survey that examines user attitudes towards a range of authentication and supervision techniques. It is concluded that whilst there is still an element of reluctance amongst users to depart from the familiar password based mechanisms, many are convinced of the need for improved authentication controls. The acceptability to users of various new techniques is variable, but many seem willing to consider a range of alternative methods.

Keywords

Authentication, Biometrics, User supervision, Survey.

Introduction

User authentication is widely accepted to represent an essential first line of defence in the security of Information Technology (IT) systems. All but the most trivial systems, therefore, require some form of authentication in order to verify that a claimed user identity is indeed correct. There are three main approaches to user authentication: something the user knows (e.g. password or PIN), something the user has (e.g. a card or other token) and something the user is (e.g. a biometric characteristic) [1]. By far the most commonly used means of authentication in IT systems is the password. Passwords are conceptually simple for both system designers and end users, and can provide effective protection if they are used correctly. However, the protection provided is often compromised by users themselves. Typical problems include forgetting passwords, writing them down, sharing them with other people and selecting easily guessed words.

If the password approach is to be replaced or supplemented, then alternative means of authentication are clearly required. However, when considering such alternatives, a number of factors can be cited that may complicate their adoption:

- effectiveness (i.e. the ability to detect impostors, whilst allowing legitimate access);
- cost (i.e. financial overheads of deployment);
- user acceptance (i.e. the friendliness and transparency of the measure).

Of these, the issue of user acceptance is possibly the most difficult to assess, as it represents a highly subjective measure. This paper presents the results from a survey that set out to assess public attitudes to various forms of user authentication and, thereby, determine whether acceptable alternatives to the password could be identified. The discussion begins by summarising the potential problems with existing password approaches and then proceeds to consider the alternatives that are offered by various classes of biometric method. Details of the survey itself are then presented, leading into an analysis of the results obtained.

The problems with passwords

The password approach has a number of shortcomings, which can undermine the effectiveness of the approach [2]. Indeed, passwords can often be considered a mere hindrance to a determined hacker and can easily be bypassed by relatively inexperienced individuals using tools freely available on the Internet.

Several studies have been carried out over the last 20 years looking at the ease with which passwords can be determined. In 1979, 86% of the 3829 passwords gathered, could be guessed by a PC in less than one week [3]. This was later repeated by Klein in 1990 [4] and Spafford in 1992 [5]. Whilst the results from these subsequent experiments showed that password selection had improved (only 21% could be guessed in a week), so have the tools that can be used to *guess* them. In 1998, L0pht Heavy Industries released L0phtCrack [6], a utility which allows Windows NT Server Message Block (SMB) password packets to be captured during network authentication sessions. This utility not only allows the encrypted passwords to be captured directly off the network, it can also perform a dictionary and brute force attack against the encrypted passwords. Similar utilities are also available for other operating systems - most notably CRACK which runs under a number of flavours of UNIX [7].

There are a number of measures that can be taken to improve password security. For example:

- *Non-Dictionary words.* Forcing users to select non-dictionary passwords prevents the use of dictionary based attacks. Such attacks can identify a password in less than 20 minutes even on dictionaries with up to one million words. The only way to identify non-dictionary passwords is using a brute-force approach (testing every combination of characters for every length of password).
- *Passwords with mixed case/symbols.* Including both upper/lower case and symbols (!£$% etc.) in passwords requires any attack to use a brute force method and increases the number of character permutations that must be tried.
- *Password ageing.* Should an intruder obtain a valid username/password combination, most systems will allow them to continue to access the system until the intrusion is noticed. If a password ageing policy is in place users can be forced to change their passwords regularly, thus forcing the intruder to identify the new password.

Although these sugestions will help to make a password-based system more resiliant to an intruder they are by no means secure. A determined intruder can utilise password cracking utilities to determine even the most random password in a matter of weeks. With the advent of more powerful processors, intruders can crack passwords in a more realistic time – a matter of days for some PCs. In addition, it can be argued that restrictions such as those above may compromise the simplicity (and, hence, user friendliness) of the password method – one of the previously cited advantages. To counter these problems with password based systems, it is necessary to consider alternative approaches to user authentication.

An overview of biometric authentication approaches

Whereas the password approach relies upon something the user *knows*, biometric authentication is based upon something the user *is*. This has the advantage that it is less straightforward for the user to be impersonated or to compromise protection themselves (e.g. they cannot share, write down or forget a biometric characteristic).

Methods of biometric authentication fall into two distinct categories, namely physiological and behavioural characteristics [8].

- Physiological biometrics represent those traits that describe who we are based on physical attributes, for example fingerprints, hand geometry, retinal and iris scanning. These characteristics usually require additional equipment to be connected externally to the computer to provide the necessary data capture.

- Behavioural biometrics encompass attributes such as typing style, voice pattern and signature recognition. Most behavioural characteristics can be acquired without the need for external equipment (e.g. keyboard & mouse), although some do require specialised hardware solutions (e.g. signature recognition).

Most biometric devices offer a compromise between high security/low user acceptance and low security/high user acceptance. This trade-off can be measured as the False Acceptance Rate (FAR) and False Rejection Rate (FRR) of the devices. It has so far proved impossible to achieve a system where the FAR and FRR are simultaneously reduced to zero, as they share a mutually exclusive relationship [9]. Most systems select an appropriate level at which inconvenience to the user, through denial of access (false rejections), is acceptable, without allowing too many intruders unauthorised access (false acceptances). All systems have an Equal Error Rate, the point at which the FAR and FRR rates are equal. Whilst this rate represents the theoretical "best-fit" for security measures, it is rarely ideal in a secure environment where a preference for either high FAR or FRR exists.

In recent years, biometric techniques have progressed from the research environment to consumer products. Indeed, Microsoft Windows now incorporates a biometric application programming interface to enable easy integration and utilisation of such approaches within the operating system [10]. Some biometrics are, however, more mature and well-known than others. The table below presents a list of biometric

techniques and accompanying descriptions (these descriptions are worded as presented to the respondents in the survey that is described in the next section).

Method	Description
Keystroke analysis	Research has shown that users have different typing styles and that they can be identified by measuring the times between keystrokes [11].
Face recognition	A snapshot of the user, taken by a camera positioned on the monitor, is compared with a previously stored 'faceprint'.
Mouse dynamics	Similar to keystroke analysis, users can be identified by the way in which they use the mouse.
Voice verification	A user's voice, when speaking a word or phrase into the computer's microphone, is compared with a previously stored 'voiceprint'.
Signature analysis	A user signs their name using a special pen and pad, the signature is digitised and compared with a previously stored version.
Iris scanning	A snapshot of the user's iris, taken by a camera, is compared with a previously stored image.
Hand geometry	This technique measures the physical dimensions of the hand using a small camera and compares these with previously stored values.
Fingerprint analysis	An automated version of the fingerprint identification system similar to that traditionally used in criminology.

Table 1 : Biometric methods, as presented to survey respondents

Many organisations are already testing such alternative forms of user authentication. For example, trials of iris recognition systems have been conducted in the banking sector for use in automated teller machines [12].

A subset of the above biometrics (e.g. keystroke analysis, mouse dynamics) can be considered to represent aspects of the wider issue of behaviour monitoring. This recognises that everyone has characteristic ways of doing things and that, over time, it may be possible to establish individual profiles of behaviour. IT systems offer a number of factors that may be monitored in order to establish such a profile. Examples include:

- typical access time and location;
- operating system command usage;
- typical application and resource utilisation;
- methods of user interaction.

Techniques such as these have been incorporated into a variety of intrusion detection and monitoring systems, which can provide real-time supervision of user activity in order to detect potential impostor activity and other forms of misuse [13]. Although such an approach represents an increase in the level of security, there is also the potential to alienate legitimate users, who may be concerned about their activities being monitored to this level.

A significant body of work exists in relation to biometrics and behavioural monitoring systems and, as previously mentioned, many commercial products are now available as alternatives to simple passwords. It is, therefore, relevant to consider what the views of the potential users themselves are towards the technologies. This issue is explored in the sections that follow.

A Survey of attitudes towards authentication technologies

In order to determine the acceptability of user authentication and supervision techniques, a survey was conducted to assess the attitudes and awareness of the general public. The survey aimed to assess the following issues:

- public attitudes towards different forms of user authentication;
- the attitudes towards the concept of continuous monitoring.

The survey questionnaire consisted of 53 main questions, the majority of which were multiple choice, with the remainder requiring short written responses. Many of the questions contained multiple sections, resulting in a maximum of 130 possible answers per respondent. The survey was split into a number of categories, each focussing on a specific area of interest to the authors. Questions 1-7 gathered general details, to determine the gender, age, education, and level of computer use; these provided demographic information on the survey response base. Questions 8-14 considered the use of computers within the respondent's work environment, whilst questions 15-19 considered the use of computers at home. These helped to provide information on the spread of IT into the home and work contexts, as well as the likely IT awareness of the respondents. Questions 20-34 were intended to determine individual opinions and knowledge in the area of computer crime and abuse. The final section (encompassing questions 35-53) looked at the respondent's views on user authentication and supervision. This paper targets the issues of user authentication and supervision, whilst the findings relating to computer crime have been documented in a previous publication [14].

The survey was distributed to a wide range of individuals and organisations with the intention of gaining a diverse variety of opinions. The questionnaire was made available in two forms, a printed copy and an online version published on the authors' WWW site. Approximately 300 printed surveys were distributed with 148 completed responses being received, representing a response rate of 49%. A further 27 surveys were submitted via the web site resulting in a total of 175 responses. It should be noted that, whilst questionnaires were sent to companies, the focus required respondents to reply from an individual rather than organisational perspective. As

such, these responses were still representative of a public rather than business viewpoint on the issues.

Analysis of results

General

The vast majority (80%) of the survey respondents were male. In terms of age, 74% of the respondents were below 35, indicating that the vast majority of the responses were likely to be from people who had 'grown up' with IT to some extent. The overall breakdown of respondents by age group is given in table 2.

Age range	Respondents
16 to 24	42%
25 to 34	32%
35 to 49	18%
50 to 64	7%
65 and over	0%

Table 2 : Survey respondents by age

In terms of employment background, a high number of responses were received from the technology fields (with 103 out of the 175 responses claiming to be from the computing, communications or engineering domains). Academically over 70% of the respondents claimed to hold post-16 qualifications, with 44% having a degree level education. This represents a high level of academic achievement among the respondents and reflects the fact that the distribution of a large proportion of surveys occurred via academic channels.

The respondents had considerable familiarity with IT, with over 98% having used a computer for over one year, 88% using a computer at work and 84% using one at home. The respondents were also asked about the availability of Internet access. 129 respondents (88%) claimed to have access at work, while 69 respondents (48%) claimed to have access at home.

The information above indicates that the respondents were generally IT literate and had considerable experience using computers in both home and work environments. As later sections of the survey looked at views on user authentication and supervision in relation to such systems, it was felt that the respondents were suitably qualified to comment on these issues.

Password based authentication

Given that they represent the most common (and, hence, familiar) form of authentication, the survey began by assessing respondent attitudes towards passwords. The results indicated that over 91% of respondents relied on passwords for access

control to their computers, a figure that is generally compatible with the 1998 KPMG security survey, which showed 97% of organisations using them [15].

Due to the dominance of passwords, most users have multiple passwords for different systems and applications. When asked how many different systems or applications they use which require passwords, 26% of respondents claimed to have five or more, with 18 people claiming in excess of ten (see figure 1).

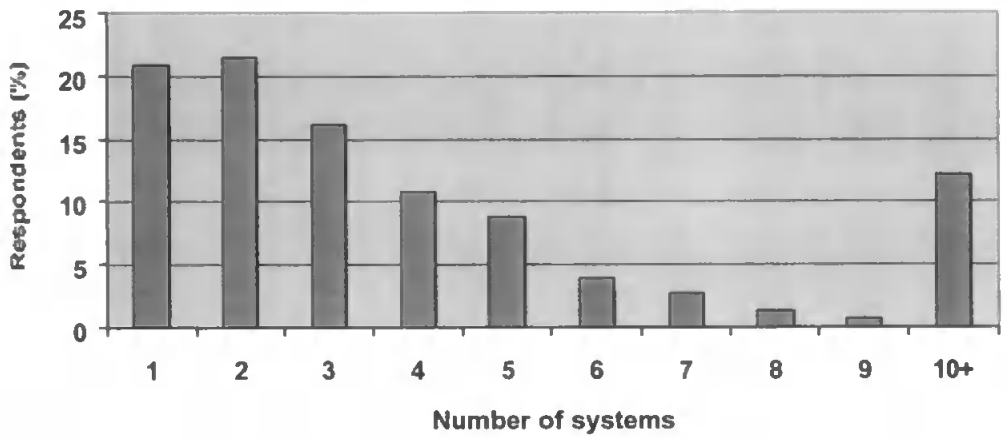


Figure 1 : Number of different systems/applications used requiring passwords

The requirement to remember such a large number of passwords can cause a major problem for users. It is, therefore, no surprise that users often select dictionary words or personal names as the basis for their passwords, as these are easier to remember. Having said this, only 15% of respondents felt that their passwords could be easily guessed. The phrasing of the question in this case gave examples of information that, if used as a basis for selection, could render the password more easily guessed (i.e. "is it part of your address, name, partner's name?"). Although the majority of users considered themselves to be safe on this basis, the question did not provide an exhaustive list of what might constitute obvious choices. As such, many respondents may still have been using insecure passwords, such as dictionary words (which the aforementioned L0phtCrack tool can determine in less than a minute).

Not only do users often choose insecure passwords, they also frequently select the same password for multiple accounts, with 40% of respondents re-using the same password. As such, should an intruder gain access to one protected account, it is quite likely that he/she will be able to reuse that same password for other machines and applications. A further issue is that of the password's lifetime. Once a password is illegitimately acquired then, without time limits, restricted logins or account monitoring, it is possible that the intruder would remain unnoticed until he/she committed an act that caused some form of disruption. The respondents were asked how frequently they changed their passwords and if they were forced to change their passwords by the system or the system administrators. As indicated in table 3, an alarming 34% of respondents claimed to never change their passwords. Furthermore, the responses to the subsequent question revealed that 51% were not forced to change their password by the system. The former represents bad practice on the part of the users, whereas the latter reflects poor system administration. From an administration

point of view, it is more encouraging to observe that 70% of users claimed to use systems in which a minimum password length is enforced. Having a minimum length of seven or more characters helps to ensure that passwords are more resilient to brute force attacks.

Frequency of password change	Respondents
Weekly	2%
Fortnightly	1%
Monthly	25%
Six-monthly	18%
Less frequently	20%
Never	34%

Table 3 : Frequency of password changes

Responses to subsequent questions revealed that, in many cases, the respondents themselves were compromising password protection, with 15% admitting to writing them down and 29% willingly sharing them with colleagues. In addition to this, 31 (21%) of the 151 respondents who used computers at work claimed to have used another person's password without their consent or knowledge.

These results serve to underline some of the known problems with passwords and provide the justification for the subsequent questions, which asked users about other forms of authentication.

Alternative authentication and supervision methods

One of the main objectives of the survey was to evaluate user's opinions regarding different authentication methods. In order to achieve this, the respondents were asked to rate the acceptability of a variety of initial login and continuous supervision techniques on a 5-point sliding scale from 'totally acceptable' to 'totally unacceptable'. A total of nine methods were cited, ranging from passwords to a variety of physiological and behavioural biometric methods. Each of the methods was briefly described on the questionnaire sheet to ensure that the respondents understood the context (using the text previously shown in table 1). Table 4 summarises the ranked results, which are also illustrated graphically in figure 2. The responses have been normalised to reflect the variable response rate to each question, as there was a higher response rate to questions on initial login authentication (probably reflecting a lack of understanding of the concept of continuous supervision amongst some respondents). The positive responses ('totally acceptable' and 'acceptable') were summed and then the total number of negative responses ('unacceptable' and 'totally unacceptable') were subtracted, thus producing a rank of user preference.

Method	Initial login authentication	Continuous supervision
Password	95.7%	-10.2%
Keystroke analysis	29.8%	25.5%
Face recognition	49.1%	3.2%
Mouse dynamics	21.3%	21.8%
Voice verification	53.4%	-0.6%
Signature analysis	40.1%	-35.9%
Iris scanning	47.2%	-16.8%
Hand geometry	44.4%	-19.9%
Fingerprint analysis	48.8%	-16.0%

Table 4 : Ranked user preference of security methods

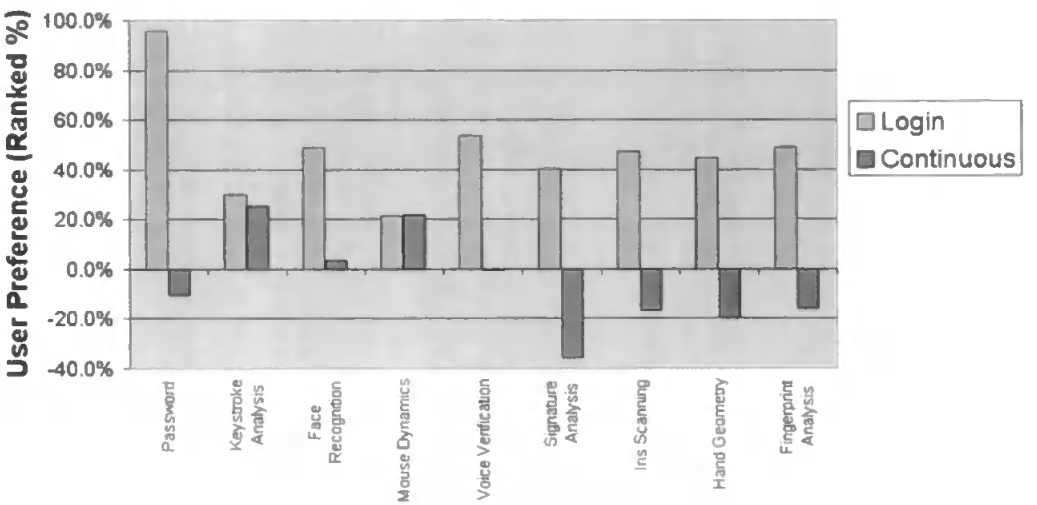


Figure 2 : User preference of authentication methods

As expected, the most popular form of initial login authentication was the password, with 90% of respondents rating it as 'totally acceptable' (scoring more than twice as many votes in this category than most other methods). However, this did not mean the outright rejection of alternative methods and many also achieved respectable scores. The authors were, however, surprised to see a general acceptance of mouse dynamics for initial login authentication. This was felt to be somewhat erroneous, as it is unlikely that moving the mouse for logging-on would provide sufficient data for a unique identification. It is expected that using a combination of methods, such as password and keystroke analysis, would provide a much more reliable method of initial login authentication.

It is clear that there is a high level of user acceptance for all the initial login authentication techniques suggested. Methods such as face recognition, voice verification, signature analysis, iris scanning, hand geometry and fingerprint analysis were all considered favourably. It is interesting to note that all of these techniques (with the exception of signature analysis) have had significant media coverage,

especially through film and television. It is possible that familiarity with these techniques influenced the respondents' choices. The acceptance of signature analysis cannot be readily explained by the familiarity with the technology through the media, however the concept of a signature as a means of identity verification is well established in our society.

After passwords, the most acceptable forms of login authentication were considered to be voice verification and fingerprint recognition, scoring raw overall acceptability ratings of 68% and 67% respectively. The latter result is somewhat surprising, in that conventional wisdom suggests that the association of fingerprints with criminal identification may represent a potential barrier to user acceptance. However, it is clear from these results that the majority of respondents are comfortable with the concept. It can, however, be noted that, in the normalised results (as presented in table 2), face recognition scored higher than fingerprints once negative responses had been taken into account

One of the significant questions posed in the survey was whether respondents would be comfortable with the concept of continuous supervision. This would provide a means for authentication to become an ongoing process within a logged in session, rather than being merely a one-time judgement at the beginning. This, in turn, would guard against situations such as an impostor replacing a legitimate user at the terminal or an impostor who may have been able to fool the initial login authentication system. In general, the respondents were positive towards the idea of monitoring, with 43% considering it acceptable, though 29% were unsure. However, the respondents considered only three techniques acceptable; namely keystroke analysis, mouse dynamics and face recognition (the latter being with a very low preference). Whilst the overall ranked results reflected sensible views, some of the individual responses in the underlying data did provide a few surprises. In particular, 34 respondents rated the use of signature analysis for continuous monitoring to be 'acceptable'. This is most likely to be a misunderstanding, as few computer users would be prepared to stop work and sign their name intermittently (a view borne out by the fact that 90 rated this as 'unacceptable').

Respondents were also asked to consider how long they would be prepared to spend creating a behaviour profile that the monitoring system would use to authenticate them. The responses are shown in table 5. It is clear that the majority of users would not be tolerant of explicit profiling activity for any long periods. Equally, the time that most of them would consider acceptable is 15 minutes or less – which would be unlikely to be adequate for some measures (e.g. whilst face and fingerprint recognition systems would allow adequate registration within this time, accurate measures relating to typing and more general system usage would require longer periods). As such, elements of profiling would need to occur as a transparent background task in order to ensure user acceptance.

User-profile set-up time	Respondents
No time	11%
Up to 5 mins	36%
Up to 15 mins	24%
Up to 30 mins	13%
Up to 1hr	12%
> 1hr	5%

Table 5 : Acceptable duration of profiling activity

Once a profile has been created, there is still the possibility that a monitoring system may falsely reject a legitimate user, believing them to be an impostor. The questionnaire made the respondents aware of this and asked them how frequently they would be willing to tolerate such errors. The results are presented in table 6 and clearly illustrate that any deployed system would need to have a very low error rate in order to avoid alienating the user population.

Frequency of false rejection	Respondents
Hourly	7%
Daily	27%
Weekly	36%
Never	29%

Table 6 : Perceived tolerable frequency of false rejection by monitoring system

It is recognised that the concept of continuous supervision also introduces ethical considerations. Indeed, 40% stated that they would consider monitoring as an invasion of their privacy, with a further 18% being unsure. It is clear that if continuous supervision of users is to be implemented, then certain safeguards should be considered. In particular, users should be aware of the intended uses of the information collected. 45% of respondents felt that they could not trust their organisation to use the supervision data for security-related purposes only and were concerned that it could be utilised for an ulterior motive, such as monitoring work productivity. 85% stated that users should be aware of any monitoring being used. The simplest way to ensure these requirements are met is to involve the users in the planning and implementation of these systems and provide clear policies on the uses for the gathered information.

Finally, the respondents were asked to indicate which fields/sectors would benefit most from supervision of users by computer, rating the benefit from 'great benefit' to 'no benefit at all'. These results were collated and ranked and are shown in figure 3.

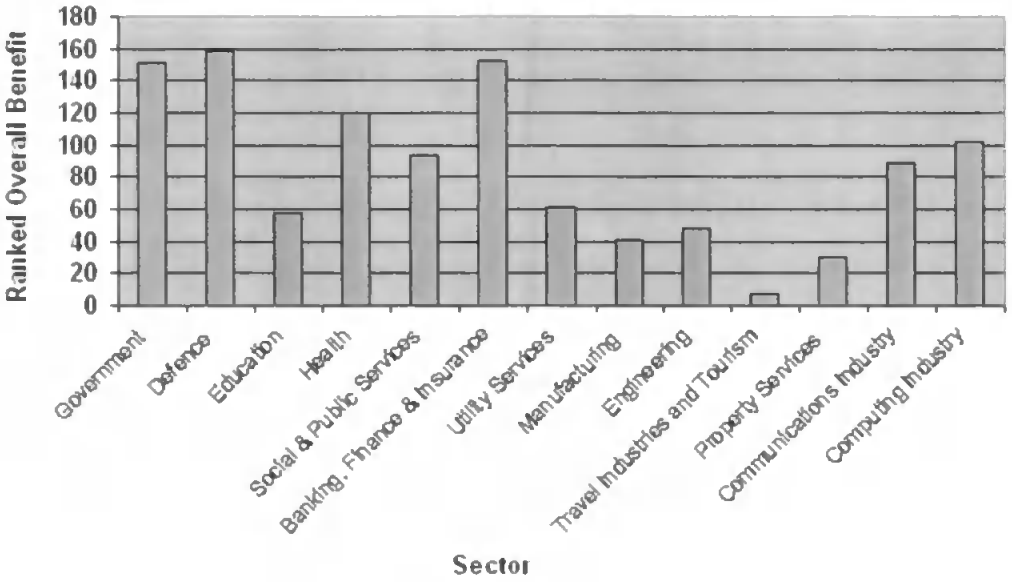


Figure 3 : Benefit from monitoring by sector

As expected, the majority of respondents considered the areas of government, defence, health and banking to benefit most from user supervision (these being the areas with the most obviously sensitive systems and data to protect). However, the respondents felt that all areas could benefit from improved supervision, showing that there is still considerable concern over the perceived computer security across all sectors.

Discussion

The results clearly demonstrate the shortcomings of password-based authentication, as well as the fact that, in spite of these, it remains the dominant form of user authentication. However, the fact that the respondents have shown a willingness to use alternative authentication techniques can be considered to be encouraging. It should be noted, however, that in the majority of cases, it is unlikely that the respondents had actually used the techniques that they were being asked to comment upon. As such, it is possible that their views may change if presented with the practical experience.

Given that a strong preference was expressed for passwords, consideration should be given to retaining them as the means of login authentication, whilst identifying means to compensate for their weaknesses. Suitable strategies in this respect could include:

- Utilising password login in conjunction with transparent keystroke analysis of the information entered. In this way, the user would be authenticated not only by *what* they type, but also *how* they type it. This should not have any significant influence on user acceptance, as the primary authentication mechanism will still appear to be the password.

- Retaining password-only authentication at login, but supplementing it with continuous supervision during the user session. The survey results suggest that techniques such as keystroke analysis and mouse dynamics would be acceptable to users in this regard.

The respondents preference for passwords is in agreement with the previously published results from the Australian TRUST project, which (from a survey of 76 participants) found users' principal preference to be for passwords, followed by physiological biometrics and, finally, behavioural measures [16]. The latter finding is, however, in contrast to the results from this study in that (for continuous monitoring) the behavioural techniques of keystroke and mouse dynamics were chosen in preference to the physiological technique of face recognition. Indeed, in the TRUST study, keystroke analysis and pointing device based verification scored the lowest of the seven biometrics assessed.

Although many considered the concept of continuous supervision to be acceptable for security purposes, the respondents showed concern over the potential wider use of such data. As such, it is important for organisations to establish agreed working practices to employees before proceeding with such methods (this may assist in reassuring those such as the 29% of respondents who were undecided over the acceptability of the monitoring concept). If such practices are not naturally adopted by organisations, it is possible (maybe even preferable in some cases) to legislate on acceptable supervision practices. This could be implemented in a similar way to that which restricts the rights of an employer to intercept and/or read an employee's email correspondence.

Overall, a significant factor in the acceptance of alternatives to the password will be that of education. If people can be shown that newer authentication techniques are safe, reliable and secure, then their acceptance is likely to be improved.

Conclusions

The survey has shown that, although demonstrably weak, the password remains the most popular form of authentication in the minds of users. However, a number of other methods emerged as possible contenders and it is possible that practical experience of using them, combined with improved awareness of the vulnerabilities of passwords, would increase their perceived acceptability as alternatives.

Another conclusion that can be drawn from the survey results is that the use of continuous supervision is, in general, acceptable. However the viability of such a scheme would be dictated by the methods chosen and subject to suitable assurances being given to the monitored population regarding the planned uses of the collected data.

The findings from the survey will be used to inform on-going work in relation to an architecture for real-time user supervision and monitoring [17]. This system will be based upon composite authentication techniques, rather than attempting to apply particular techniques in isolation.

References

[1] Wood, H.M. 1977. "The use of passwords for controlled access to computer resources", NBS Special Publications, U.S. Dept. of Commerce/NBS: 500-509.

[2] Jobusch, D.L. and Oldehoeft, A.E. 1989. "A Survey of Password Mechanisms : Part 1", Computers & Security, Vol. 8, No. 7: 587-604.

[3] Morris, R. and Thompson, K. 1979. "Password Security: A Case History", Communications of the ACM, Vol. 22, No. 11: 594-577.

[4] Klein, D. 1990. "A survey of, and improvements to, password security", Proceedings of the USENIX Second Security Workshop, Portland, Oregon, August 1990: 5-14.

[5] Spafford, E.H., 1992, "Opus: Preventing Weak Password Choices", Computers and Security, Vol. 11, No. 3: 273-278.

[6] Heskett, B. 1998. "A new windows password cracker", Cnet News.com, 13th February 1998, http://news.cnet.com/news/0-1003-200-326537.html

[7] Cherry, A., Henderson, M.W., Nickless, W.K., Olson, R. and Rackow, G. 1992. "Pass or Fail: A New Test for Password Legitimacy", Mathematics and Computer Science Division, Argonne National Laboratory, MCS-P328-1092, September 25th 1992.

[8] Sherman, R. 1992. "Biometrics Futures", Computers & Security, vol. 11, no. 2: 128-133.

[9] Cope, B.J.B. 1990. "Biometric Systems of Access Control", Electrotechnology, April/May: 71-74.

[10] Sapsford, J. 2000. "Biometrics to bolster Windows security", ZDNet News, 2 May 2000.

[11] Furnell, S.M., Morrissey, J.P., Sanders, P.W., and Stockel, C.T. 1996. "Applications of keystroke analysis for improved login security and continuous user authentication", Katsikas and Gritzalis (eds), Proceedings of 12th International Conference on Information Security (IFIP SEC '96): *283-294*.

[12] NCR. 1999. "NCR announces iris recognition trials with Nationwide Building Society". http://www3.ncr.com/product/financial/press/sensnat.htm

[13] Mukherjee, B., Heberlein, L.T., Levitt, K.N. 1994. "Network Intrusion Detection", IEEE Networks, Vol. 8, No. 3: 26-41.

[14] Dowland, P.S., Furnell, S.M., Illingworth, H.M., and Reynolds, P.L. 1999. "Computer Crime and Abuse: A Survey of Public Attitudes and Awareness", Computers & Security, vol. 18, no. 8: 715-726.

[15] KPMG. 1998. *Information Security Survey 1998*, KPMG Information Risk Management, UK, http://www.kpmg.co.uk.

[16] Deane, F., Barrelle, K., Henderson, R., and Mahar, D. 1995. "Perceived acceptability of biometric security systems", Computers & Security, vol. 14, no. 3: 225-231.

[17] Furnell, S.M. and Dowland, P.S. 2000. "A conceptual architecture for real-time intrusion monitoring", Information Management & Computer Security, vol. 8, no. 2.

A conceptual intrusion monitoring architecture and thoughts on practical implementation

P.S.Dowland and S.M.Furnell

Network Research Group, Department of Communication and Electronic Engineering, University of Plymouth, United Kingdom
e-mail : sfurnell@plymouth.ac.uk

Abstract

The paper presents a conceptual description of the Intrusion Monitoring System (IMS) architecture, which is designed to facilitate detection of system penetration and other anomalous activity in a networked environment. The architecture is based upon eight functional elements, distributed between a monitoring host and a series of monitored client systems. The discussion also considers how the approach could be integrated within the Windows NT environment.

1 Introduction

The concept of real-time intrusion monitoring has been of interest in the IT security domain for a number of years, with the original idea having been proposed by Denning (1987). Such an approach is valuable as a means of combating a number of classes of system abuse, including penetration by unauthorised persons and misuse of privileges by registered users. In addition, abuse may be perpetrated by malicious software, such as viruses and Trojan Horse programs. Although a number of IDS have been developed (Mukherjee et al. 1994), these have generally targeted large systems or specific domains (e.g. military). Commercial implementations are generally restricted in their monitoring functionality. However, the increasing interconnection of corporate systems, coupled with reported increases in computer abuse incidents (CSI 1999), suggests that the use of more advanced intrusion monitoring functionality would be advantageous. This paper presents the conceptual architecture of the Intrusion Monitoring System (IMS), which aims to detect anomalous activity in a networked environment, followed by consideration of how to realise the approach in practice under Windows NT.

2 Intrusion Monitoring System overview

The IMS architecture was originally proposed by Furnell (1995) and is based upon the concept of a centralised *Host* handling the monitoring of one or more *Clients* on local workstations. The Clients collect the required data relating to system activity and respond to any suspected intrusions detected by the Host. Monitoring is based upon the comparison of current activity against two categories of stored information, namely user behaviour profiles and generic intrusion rules. These approaches are common to other intrusion monitoring architectures, such as the Intrusion Detection Expert System described by Lunt (1990). User profiles could conceivably hold a range of identification, authentication and behavioural information relating to registered users. Examples of potential characteristics would include system access times and locations; typical levels of system resource utilisation; application and file usage; methods of user interaction; and biometrics (i.e. physiological and behavioural characteristics). Biometric monitoring is considered to be particularly appropriate to prevent impostor penetration and a number of options exist that could be employed in this context, including keystroke analysis, face recognition and voice recognition (Cope 1990). Other well-known biometrics, such as fingerprint recognition, are less strongly favoured in the IMS context, as less opportunity exists to integrate them in a manner that is transparent (and, hence, non-intrusive) to the legitimate user.

Some classes of intrusion or misuse can be trapped without identifying departures from historical patterns of user behaviour. The occurrence of some events will be suspicious in themselves and, therefore, the system requires a means to monitor for these as well. Examples of generic indicators would include consecutive access violations, out of hours access, account overuse / simultaneous access, use of inactive accounts and extensive use of help systems. While none of these alone would provide sufficient indication to state that an intrusion was in progress, the combination of two or more could be more persuasive. In the IMS context, these attack signatures would be represented via Intrusion Rules that, if satisfied, would increase the alert status of the system.

3 The IMS Architecture

Anomaly Detector. The *Anomaly Detector* analyses activity for suspected intrusions, comparing it against the behaviour profile of the current user's (claimed) identity, as well as against the generic intrusion rules. The detector is comprised of further sub-modules, each handling specific monitoring tasks (e.g. keystroke analysis, tracking of resource usage etc.). The detector maintains an *alert status table*, with entries existing throughout the life of each user-initiated session or process to indicate the level of detected anomalies and thereby the confidence of a potential intrusion. This information would be examined and updated each time activity data relating to the relevant user / process is analysed. The alert status level would increase in response to departures from the user-specific behaviour profile or the satisfaction of generic intrusion indicators. The level would be reduced after successful challenges or after a sufficient period of normal activity to allow the system to discount the previous anomaly. The alert status level can be linked with the types of activity that a subject is permitted to perform. In this way, a phased reduction of permitted behaviour would occur as the level increases. Sensitive activities / information could, therefore, be denied if doubt exists over the legitimacy of the current user, whilst still allowing more mundane activities to continue. The approach would demand that a maximum alert status threshold be associated with each of the activities or objects that IMS is to control.

Profile Refiner. User behaviour may legitimately alter over time. The *Profile Refiner* aims to provide an automatic means to account for such changes, using neural networks to analyse and recognise behavioural characteristics that might not be apparent to a human observer. In this way, the effectiveness of the system has the potential to improve over time. It might also be possible to determine which of the profiled characteristics provide the best discriminators for each user and thereby establish various levels of behavioural indicator (with the primary level representing the most reliable verifiers). This hierarchy could also be extended to allow for the fact that some characteristics may represent negative indicators (i.e. those that, despite refinement, are found to cause a high level of false rejection).

It would be undesirable for the *Profile Refiner* to utilise data that is later found to be anomalous. Refinement should, therefore, only take place after the termination of non-anomalous user sessions. User-specific profile records would also incorporate a series of flags to indicate whether the individual behaviour characteristics are ready to be used in supervision or still being developed. This will allow a gradual training period to be defined for new user profiles without the IMS continually generating intrusion alerts (the flags would also allow a specific 'refinement only' period to be established for existing profiles that have proved to be inadequate for the legitimate user). The purpose of associating flags with each profile characteristic is so that some degree of monitoring could still continue whilst other aspects are being (re)trained.

Recorder. The *Recorder* handles the short-term storage of user-related activity data during a session and focuses specifically upon the collection of data relating to the profiled characteristics of a given user (e.g. collection of keystroke data in relation to the typing profile). Upon termination, the information will be used as input to the *Profile Refiner*, provided that the session was not considered anomalous. In the event of a proven anomaly, the *Recorder* can discard the session data.

Archiver. The *Archiver* collects data relating to *all* system activity and stores it in a long-term archive, providing a more permanent record of activities and suspected anomalies. The storage occurs regardless of whether sessions / processes are regarded as anomalous and details of all security relevant events are archived. Such events include login failures, intrusion alerts, authentication challenges and suspended sessions. However, in order to conserve storage space, it may be desirable to only record details of certain types of event. The *Archiver* is therefore configurable to suit the preferences of the organisation involved (note that the same would *not* be true for the *Recorder* as this would always need to collect information on any activities for which profile refinement may later occur). The long-term retention period of archived details would be organisation-dependent.

Collector. The *Collector* is responsible for obtaining information on all relevant client-side activity. The module must operate in such a way as to encompass, but be independent of, all system applications. It is envisaged that this could be best achieved by implementation at the operating system (OS) level, such that key events also lead to IMS notification. For example, a significant proportion of data collection could be based around the interception and redirection of selected OS service requests (such as file input / output, application execution, keyboard input). In some cases, data could be obtained directly from audit trail records – as in previous systems, such as Wisdom & Sense (Leipins and Vaccaro 1989.). However, with certain aspects (e.g. keystroke analysis) the required information will not be held by audit trails and implementation may, therefore, require a significant number of OS links.

Whilst this would serve to make this aspect of IMS very system specific, it would be more efficient than attempting to modify individual applications to provide relevant information to IMS.

Responder. This module resides in the Client and responds to anomalies detected by the Host. The operation centres around the continuous monitoring of the alert status transmitted by the Host, with increases in the level triggering appropriate actions. The nature of response at different levels would vary and a detailed discussion of the possible options is beyond the scope of this paper. However, appropriate responses might include: issuing of an explicit challenge for further authentication; recording of details in an intrusion log for later investigation; notification of the system manager; phased reduction of permitted behaviour; locking of the intruder's terminal; and termination of the anomalous session / process.

Communicator. The *Communicator* provides the network communications interface between the Host and the local Client(s). The principal functions include transmitting user and process information to the Host and then subsequently keeping the Client(s) informed of the current alert status. If implemented in a heterogeneous environment, the Client side would be responsible for resolving any operating system differences that exist within the monitoring domain, so that information could be presented to the Host in a consistent, standardised format. The actual communication could then be handled via a standard sockets approach, with protection provided by a technology such as Secure Sockets Layer (Frier et al. 1996).

Controller. This module allows the operation of the IMS system to be configured. On the Host side, this applies to the *Anomaly Detector* (e.g. behaviour characteristics to utilise, generic rules in operation), the *Profile Refiner* (e.g. frequency of refinement, acceptable thresholds for challenges) and the *Archiver* (e.g. level of detail required, specific events to record or exclude from logging). On the Client side, configuration relates to the *Collector* (e.g. the level of data collection, which could be automatically linked to the characteristics being monitored by the *Anomaly Detector*) and the *Responder* (e.g. the level of response required at each alert status level). These settings would be controlled and recorded through the Host system. Local Client(s) would then be configured at the time of session initiation. Other features would also be provided under the auspices of the *Controller*, including user profile management and update of the generic rulebase.

4 IMS Implementation

Work is currently being conducted to develop an implementation for Windows NT. This requires replacement of the Graphical Identification aNd Authentication (GINA) Dynamic Link Library (DLL) - the interface through which a user can provide his/her identification, typically in the form of a username and password. However, it can be replaced with any desired authentication method (e.g. commercial products are available using fingerprint and faceprint methods). In addition to the GINA replacement, the IMS would also require software to provide the required continuous monitoring, together with a remote security server. The security server would be used to store, maintain and update the user profiles. This server would process all authentication requests together with local system audits and updates to profiles. This role is slightly different to that of a network server, which usually only authenticates requests for access at the beginning of a session. Instead, the security server would be responsible for ongoing authentication throughout a session. A user login would be performed locally (or remotely via a domain controller) and once the user's credentials are confirmed the monitoring program would be loaded to provide continuous user authentication. To prevent tampering, the IMS system would store user profiles remotely on the security server. These would be encrypted and downloaded at login (although for higher security the profiles could be maintained on the server with authentication requests being handled by the server). To ensure monitoring hardware has not been tampered with, a local machine audit can also be initiated together with checks for dependent entries in configuration files or registry keys.

To reduce network traffic, it is envisaged that the user authentication would be performed on the local computer with only warnings or profile updates being fed back to the security server. Under certain scenarios it may be necessary to lock local computers if contact is lost with the security server to ensure an intruder had not removed a computer. However, it should be noted that this creates a weak point and appropriate measures will be needed to prevent a single server stopping the entire network. This could take the form of a backup server, in a similar fashion to a secondary DNS server. Alternatively, the range of facilities available to the user can be restricted until they can be re-authenticated.

Once a user has been authenticated by the replacement GINA DLL, the IMS client would be activated. The IMS client would then check the IMS security server (host) for the users monitoring characteristics and rules, to allow it to select the most appropriate monitoring

programs. At this point, the selected characteristics would be loaded and initialised. To ensure ease of implementation and future modification, each distinct monitor program would be implemented as a system DLL. Taking the keystroke analysis example, the monitor program would install a system-wide hook to intercept all keystrokes received by the keyboard buffer and pass these to an analysis algorithm within the DLL. To ease the processing burden, the DLL would pass periodic samples of keystroke activity (either time or quantity based) to be analysed. The results of this analysis would then be passed on to the IMS client program to be either compared to the local copy of the user's profile or to be returned to the IMS Host for remote verification and subsequent action. In the event that the IMS detects a potential intruder, a call can be made to the GINA DLL to provide a request for further user authentication. (e.g. question and answer challenge or biometric identification request). As the GINA DLL provides the login interface for NT, it is impossible to perform a local user login without authenticating via the GINA DLL. This can be used to enforce a variety of security rules. For example, the system may refuse login without the presence of the IMS host, the system may only accept a user once a secondary authentication has been made or the system may disallow local logins if monitoring hardware (e.g. a camera) has been removed.

Under Windows NT, the *Anomaly Detector, Profile Refiner, Recorder* and *Archiver* would be implemented on the IMS Security server (Host) as a software suite. To facilitate future upgrades, each component would be contained within a separate DLL, with a front-end provided through a single executable. Depending on the size of the system being monitored, it may be necessary to distribute the tasks over multiple hosts to cope with the level of data analysis and profile updating. It may also be beneficial to implement these programs as services under NT – ensuring they are loaded at host boot-up. The *Collector* would be implemented as a mediator on the Client, collecting information via hooks that intercept system messages (e.g. keystrokes, mouse movements etc.) and forwarding this information on to the *Communicator*. This would again be held in a DLL, which would be called by the replacement GINA. The *Responder* would be implemented within the GINA DLL and provide a replacement login interface for NT. This would also interface to the *Communicator* and *Collector* for data acquisition and client-host communication. The *Communicator* provides the interface between the client and server IMS software. It would be implemented as a shared DLL (used by both client and host). The *Communicator* would use standard TCP/IP communication, with all data being encrypted using a standard algorithm. The *Controller* provides a management interface to the IMS server software and would be implemented on the host as a single executable linking to the *Anomaly Detector*, *Profile Refiner*, *Recorder* and *Archiver* DLL's. Further details of the practical implementation approach can be found in Dowland and Furnell (2000).

5 Conclusions

Intrusion detection systems have the potential to provide an important contribution to system security, protecting against abuse by both external persons and organisational insiders. The IMS architecture represents an example approach and the paper has sought to describe the main functional elements. A system such as IMS is considered to represent a useful addition to Windows NT, which increasingly has a role in enterprise-level IT and, hence, an increasing requirement for strong protection. The paper has provided an indication of how IMS would be realised in the NT environment. The detailed mapping of the IMS approach to the NT architecture is currently in progress.

References

Cope, B.J.B. 1990. "Biometric Systems of Access Control", *Electrotechnology*, April/May: 71-74.

CSI. 1999. "Issues and Trends: 1999 CSI/FBI Computer Crime and Security Survey", USA, March 1999.

Denning, D.E. 1987. "An intrusion-detection model", *IEEE Transactions on Software Engineering*, SE-13(2):222-232.

Dowland, P.S. and Furnell, S.M. 2000. "Enhancing Operating System Authentication Techniques", Proceedings of INC 2000 (3-6 July, Plymouth, UK).

Frier, A.; Karlton, P.; and Kocher, P. 1996. "The SSL 3.0 Protocol", Netscape Communications Corp., Nov 18, 1996.

Furnell, S.M. 1995. *Data Security in European Healthcare Information Systems*. PhD Thesis. University of Plymouth, UK.

Leipins, G.E and Vaccaro, H.S. 1989. "Anomaly Detection: Purpose and Framework", In *Proceedings of the 12th National Computer Security Conference* (USA), 495-504.

Lunt, T.F. 1990. "IDES: An Intelligent System for Detecting Intruders", In *Proceedings of the Symposium: Computer Security, Threat and Countermeasures* (Rome, Italy, Nov. 1990).

Mukherjee, B.; Heberlein, L.T.; Levitt, K.N. 1994. "Network Intrusion Detection", *IEEE Networks 8*, no.3: 26-41.

Enhancing Operating System Authentication Techniques

P.S. Dowland and S.M. Furnell

Network Research Group, University of Plymouth, Plymouth, United Kingdom
e-mail: pdowland@plymouth.ac.uk

Abstract

The need for enhanced user authentication has been evident for some time; but has not been addressed at the operating system level to any degree. Whilst all mainstream operating systems offer some level of user identification and authentication, this is generally based on the username/password combination. Although a number of extensions to operating system security have been proposed (with some reaching implementation) none, as yet, have been integrated into the core operating system kernel. Although there are examples that extend the operating system security model with additional measures (e.g. plug-in fingerprint scanners), these merely extend the operating system security rather than replace it with a more secure version.

This paper will consider the need to improve operating system security focussing upon the enhancement of user identification and authentication. In particular, the security weaknesses of the Microsoft Windows NT environment will be considered, leading to a discussion of supervision techniques that may be integrated within the NT security model. Finally, the conceptual integration of an Intrusion Monitoring System (IMS) architecture is considered.

Keywords

User authentication, user supervision, security, intrusion monitoring, Windows NT.

Introduction

The most commonly used form of operating system user authentication is the username/password pair. In most systems, the allocation of passwords (and sometimes usernames) is entirely at the discretion of the users and, as such, is the cause of many security loopholes. The weaknesses of passwords as the primary form of user authentication have been documented in previous works (Jobusch and Oldehoeft, 1989; Cherry et al, 1992) and will not be covered in detail here. However, typical weaknesses include passwords being easily guessed, shared among users, the use of dictionary words (which are more vulnerable to attack) and being written down near PCs. Even when passwords are more selectively chosen, they are still vulnerable to brute force attack, especially with the fast processors and distributed password cracking software now freely available (Savill, 1999).

It is clear that the 'out of the box' configuration for an operating system is inadequate for most systems. For example, most UNIX installations leave many security 'back-doors' into the system wide open by default (e.g. default password settings that administrators *should* change, but often do not), which provide an easy target for hackers (Stoll, 1989). Similarly, a standard installation of Microsoft Windows NT requires many steps before it can be considered secure (Microsoft, 1999a). Relying on passwords in their common form is inadequate and, therefore, some form of advanced

user identification is desirable. Ideally, this should also be combined with some form of user monitoring; thus ensuring that a user's session cannot be hijacked. Hijacking occurs where a users' active session is taken over by another user (intruder). This can occur on a number of levels; firstly an intruder can simply resume a session by waiting for the user to leave their desk and then taking advantage of an unprotected computer. Alternatively, an intruder may connect a device (computer) to the target computers' network connection and masquerades as the target computer. Whilst hijacked sessions are most likely to occur in a corporate networked environment, there are still risks to SME's and individuals – this is especially true with the trend towards e-commerce and the increased confidence in purchasing on-line (NOP, 1999). An intruder may be able to capture a credit-card purchase and then either modify or replay that same exchange of data to their advantage. Enhancing user authentication is, therefore, of value to both the commercial and private sectors.

Another problem, which is often overlooked during the selection of appropriate security systems, is that of internal misuse of computer systems. Most systems rely on the username/password pair to identify and authenticate a user. Once this authentication has been given, the user is often free to access the system without further checks or monitoring. Whilst most systems offer the ability to selectively exclude users and/or groups from specific shared resources, this is not usually the default setting. For example, under Windows NT, shares are, by default, accessible to all users and an administrator must specifically set access rights to ensure a shared resource is protected from internal misuse. A similar issue relates to private use of computing resources. Although this is not usually considered to be a security risk, it can represent a loss to a business either through physical resource usage or loss of computer processing time. Often the biggest loss to a company is that of lost employee time; not just through the time lost by the employee concerned but also in the time taken to investigate the problem and prevent further misuse (Audit Commission, 1998).

Operating system security weaknesses

With operating systems such as Microsoft's Windows NT4 comprising several million lines of code, it is, perhaps, no surprise that security weaknesses should occur. However, it is often surprising to see the scope and frequency with which such fundamental flaws are found. Using Microsoft Windows NT4 as an example, the Microsoft Product Security Notification Service issues several warnings each week, each identifying a potential security problem with the operating system or its sub-components (Microsoft, 1999b). Of course, Microsoft Windows NT is not the only operating system to suffer with such security problems – the many flavours of Unix also generate hundreds of security patches each year (see http://www.faqs.org/faqs/computer-security/most-common-qs/index.html). However, the wider distribution of Windows means that the consequences of security vulnerabilities are potentially more wide reaching. A further drawback with a "popular" operating system is that as its popularity increases, it becomes a greater target to hackers partly due to the increased usage (and, therefore, potential targets) but also because of the greater availability of information relating to security weaknesses. This has been particularly prevalent with the appearance of "script-kiddies" (young inexperienced hackers), who frequently use the many resources (called "filez") which are available

from hacking sites on the Internet. A noticeable side effect of this is the use of alternative operating systems where security is of prime concern. For example, the US Army has switched to a MacOS-based web server platform, following a hacking incident when the server was running Windows NT (Donoghue, 1999). This is not to say that MacOS is any more secure than Windows NT, just less widely targeted.

Despite the frequency of these vulnerabilities, the only standard form of security provided by these operating systems for authentication purposes is the password.

Enhancing Windows NT security

Windows NT security can be considered on two levels, local machine and domain or remote login (Figure 1).

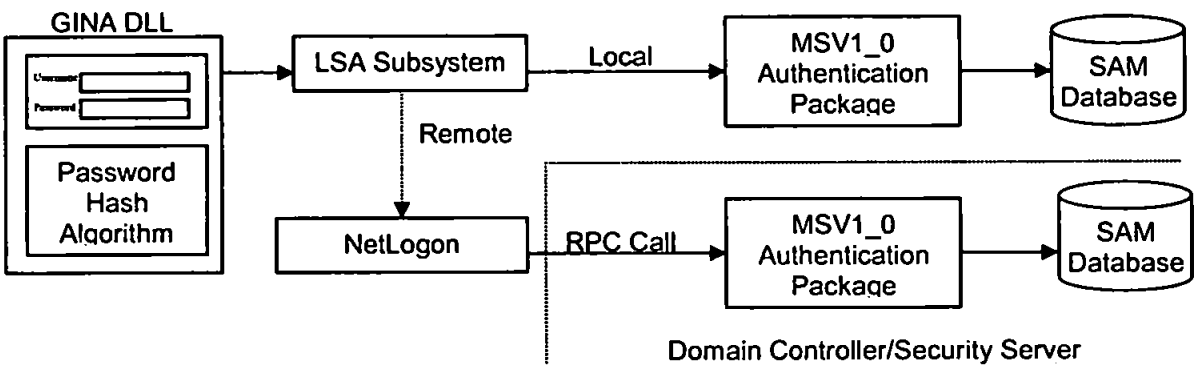


Figure 1 Local/remote user authentication

When a local user presses the "Control-Alt-Del" combination to initiate a login they are prompted to enter their username/password pair. The NT hash algorithm is then applied to the password and is passed on to the Local Security Authority (LSA) which calls the MSV1_0 authentication package. This hash is finally compared with the hash stored in the local Security Account Manager (SAM) database by the authentication package. Once a users' password is authenticated, an access token is issued that is valid for that users' session.

When a user wishes to be authenticated across a network (to log-in to a domain controller or for access to a remote machine), the password hash must be transferred across the network. When the user is prompted for their username/password they are also required to enter a valid domain. When the authentication package identifies that the account is not held locally, a call is made to the NetLogon service which sets-up a secure Remote Procedure Call (RPC) session to the domain controller to authenticate the login. The domain controller then issues a 16-bit challenge (the nonce). This challenge is then encrypted together with the password hash and is returned to the domain controller for authentication. Finally, the domain controller returns an access token which is valid for that users' session.

One of the main problems of the above technique is that once the challenge (nonce) has been intercepted and with knowledge of the encryption algorithm it is possible to

determine the password hash. Given a known hash, it is feasible (with today's technology) to guess (using a dictionary and/or brute-force attack) the original password.

To achieve a more comprehensive approach under Windows NT would require a replacement GINA Graphical Identification aNd Authentication DLL (core user login system library e.g. username/password prompt). The GINA DLL provides an interface through which a user can provide his/her identification. This typically takes the form of the traditional username/password, but can be replaced with any form of identification (e.g. fingerprint scanner, iris scanner etc.).

There are a number of "add-on" software/hardware packages that can be used to enhance Windows NT security. One of the most common packages currently available is the fingerprint scanner. This is a small device that connects to the PC and provides a cost-effective way of authenticating a login attempt. These devices typically provide an additional security module that integrates into the NT security model. Similar devices are also available to capture handprint geometry, facial patterns and there are devices appearing that are capable of iris scanning. Although these packages allow the enhancement of NT security by removing the need for the user to remember a password, they are not completely integrated into the operating system and only provide a replacement for the username/password prompt. There is also a significant cost overhead to be considered (for example, a fingerprint based authentication system would require the purchase of sufficient scanners for all the PC's in an organisation). Many of these solutions also depend on additional hardware that is dedicated to the task of providing enhanced authentication and, therefore, provides no additional benefit to the organisation concerned (i.e. no purpose other than security).

Even if these techniques were integrated into the NT security model, there are still gaps which leave significant security weaknesses. For example, even with a fingerprint scanner, once the user has logged-in using their finger, there is no guarantee that the same user will sit down and continue with the session. Similarly, if a user leaves their workstation, there is no means of checking if the user who continues the session is the same that started it. (Although all versions of Windows allow the configuration of a screensaver with password protection, this is not set by default. It should also be noted that the computer is unprotected from the time the user leaves their desk to the point at which the screensaver is activated, unless they explicitly lock the terminal). Due to these risks, some form of ongoing user supervision is required to ensure that the current user is the same as the user who activated the session. The remainder of this paper considers the adoption of an Intrusion Monitoring System (IMS) and the technical aspects involved in integrating into the Windows NT security model.

Description of an IMS

Following previous research work, a proposed IMS architecture is shown in figure 2. The specific functionality of this architecture has been described in a previous paper and will not be described in detail here (Furnell et al, 1997). At the basic level, the approach involves an IMS host monitoring activity occurring on a series of client systems. The client/server relationship of the IMS architecture shown fits neatly into the Windows

NT security model architecture and the proposed IMS integration is described later in this paper. Further research work is necessary to fully integrate the IMS architecture into the Windows NT security model and will be the subject of a later paper.

The **Anomaly Detector** analyses the data gathered by the IMS client for signs of suspected intrusion. This data can be compared against both the user's behaviour profile and the generic intrusion rules (i.e. attack signatures).

The **Profile Refiner** allows the automatic modification of a user's profile in response to a valid session profile. This recognises the fact that a user's behaviour pattern may change over time (e.g. in a scenario where typing style has been profiled, their typing skill may improve) and allows a user's profile to evolve. Due to the nature of the data and the difficulty in recognising gradual behavioural pattern changes, it is likely that this would be implemented using some form of neural network (Furnell, 1994).

The **Recorder** stores a temporary record of system and user activity during a session (session profile) which can be used by the Profile Refiner to update the user profile, providing the session was not considered anomalous.

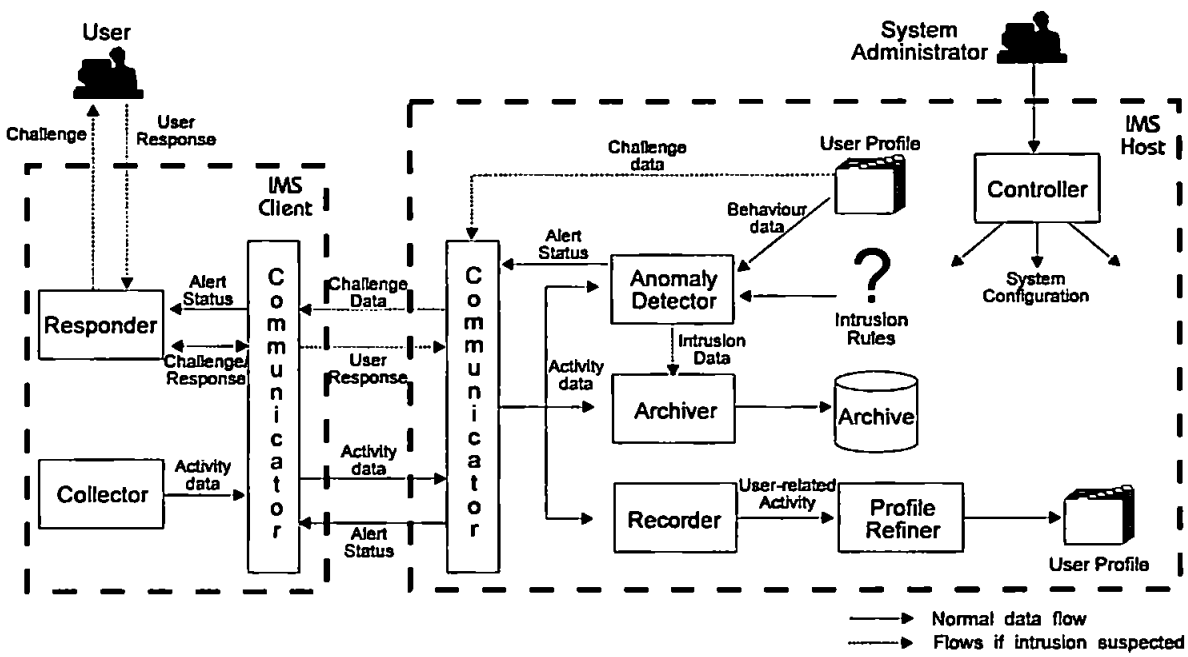


Figure 2 Proposed IMS Architecture

The **Archiver** provides an audit log, storing all security relevant events. This could also be extended to monitor *all* events if an organisation requires a more detailed log of user activity (e.g. to monitor user performance).

The **Collector** provides an interface between the IMS client and the applications running on the client computer. The collector is responsible for gathering information relevant to the user and his/her system activities. Under Windows NT the collector would be implemented as a mediator, collecting information gathered by low-level

system functions that intercept system messages (e.g. keystrokes, mouse movements etc.) and forwarding this information on to the communicator.

The **Responder** provides user interface between the IMS software suite and the end-user. Its main task is that of monitoring the signals send from the server to the client and taking appropriate action where necessary. Possible actions include; issuing a user authentication challenge, suspending a session, limiting a user's actions or cancelling a process.

The **Communicator** provides the interface between the client and server IMS software. The communicator is responsible for ensuring a consistent, reliable and secure exchange of data between the client and server. Where an IMS system is implemented in a heterogeneous environment, the communicator is also responsible for data translation to provide consistent data formatting between different client platforms.

The **Controller** provides a management interface to the IMS server software allowing an administrator to configure the IMS system-operating parameters. The controller also allows an administrator to configure client-monitoring characteristics on a global, group, machine or individual user basis.

An Intrusion Monitoring System incorporates identification and authentication of users, monitoring of users for unusual behaviour or characteristics, together with the ability to modify the profile of a user to reflect changing patterns of use/behaviour. An IMS can rely on many physiological characteristics of the user (e.g fingerprint, voice etc.) and can also monitor behavioural traits such as keystroke patterns, mouse dynamics and application/resource usage. However, it should be noted that the majority of commercially available IMS systems rely on traditional methods of user authentication

A strong potential candidate for a monitoring characteristic is that of keystroke analysis. This is a particularly attractive characteristic, as it requires no additional hardware (cost) or proprietary drivers (development time). By monitoring a user's typing profile it is possible to determine, with some accuracy, the identity of the current user. The use of a users' typing pattern as an authentication characteristic has been described in a number of papers (Furnell et al, 1996; Brown and Rogers, 1993) and has shown to be a strong distinguishing factor in certain contexts with overall False Acceptance Rate (FAR) figures as low as 4.2% being observed.

Although keystroke analysis is a good characteristic upon which to base user authentication, there are limitations. One of the major drawbacks of this characteristic is the very fact that users have a broad range of typing patterns. An inexperienced typist will use a keyboard in a slow deliberate manner, having a slow typing rate and most probably a high error rate. A trained touch-typist will type quickly with a low error rate. However, most inexperienced typists will type equally slowly and most touch-typists will type equally quickly. It is quite possible that the inter-keystroke time will be such that two typists may be indistinguishable in normal working environments.

Keystroke analysis may also be inappropriate depending on the environment in which it is used. For example, if a user is typing in numeric data for a prolonged period, it may

be impossible to achieve a statistically valid sample of keystroke data upon which to base the authentication judgement. Similarly, if a user were drawing with a mouse, there would be no keystrokes to analyse.

From this, we can see that a composite approach is needed, where several appropriate authentication and monitoring techniques are applied. For example, a user may be initially authenticated by their fingerprint, after which their typing profile and application usage can be monitored. Similarly, if that user then starts to draw using the mouse, data can be recorded to determine if the dynamic movement of the mouse is consistent with the users' profile. This technique can also be applied where users *hotdesk*. If a user moves to a desk with an additional security-relevant device (e.g. a camera for faceprint recognition), the additional measures can be detected during an audit and then utilised for that user depending upon the settings in their profile.

Integrating an IMS into the Windows NT security model

If we consider the concept of an IMS, the username/password pair could be used to identify the user with a partial degree of certainty, whilst the continuously evaluated characteristics would allow the user to be monitored throughout the session. Using the previous example of keystroke analysis, a users' typing pattern can be monitored throughout the active session and compared with a historical profile. Deviation from this profile can be flagged and a threshold set beyond which further authentication of the user would be required (Furnell, 1995). This trust level can also determine the frequency of monitoring and, where further authentication is considered necessary, the degree of certainty needed (and, hence, the form of authentication to request).

To achieve an Intrusion Monitoring System (IMS) under Windows NT would require a replacement GINA DLL and an additional piece of software to provide the required continuous monitoring together with a remote security server. A security server (or some form of centralised system) would be used to store, maintain and update the user profiles. This server would (in an ideal system) process all authentication requests together with local system audits and updates to profiles. This role is slightly different to that of a network server, which, usually, only authenticates requests for access at the beginning of a session. Instead, the security server would be responsible for ongoing authentication of a user throughout a session.

A user login would be performed locally (or remotely via a domain controller) and once the user's credentials are confirmed the monitoring program would be loaded to provide continuous user authentication (Figure 3). To prevent tampering, the IMS system would store user profiles remotely on a security server. The profiles would be encrypted and downloaded at login to the local computer (although for higher security the profiles could be maintained on the server, with authentication requests being handled by the server). To also offer security for the hardware (to ensure monitoring hardware had not been removed) a local machine audit can also be initiated, together with checks for dependent entries in configuration files or registry keys. An IMS system would also allow updating of the user profiles, to take into consideration changing user behaviour (e.g. keystroke patterns, application usage etc.) or appearance (e.g. facial recognition).

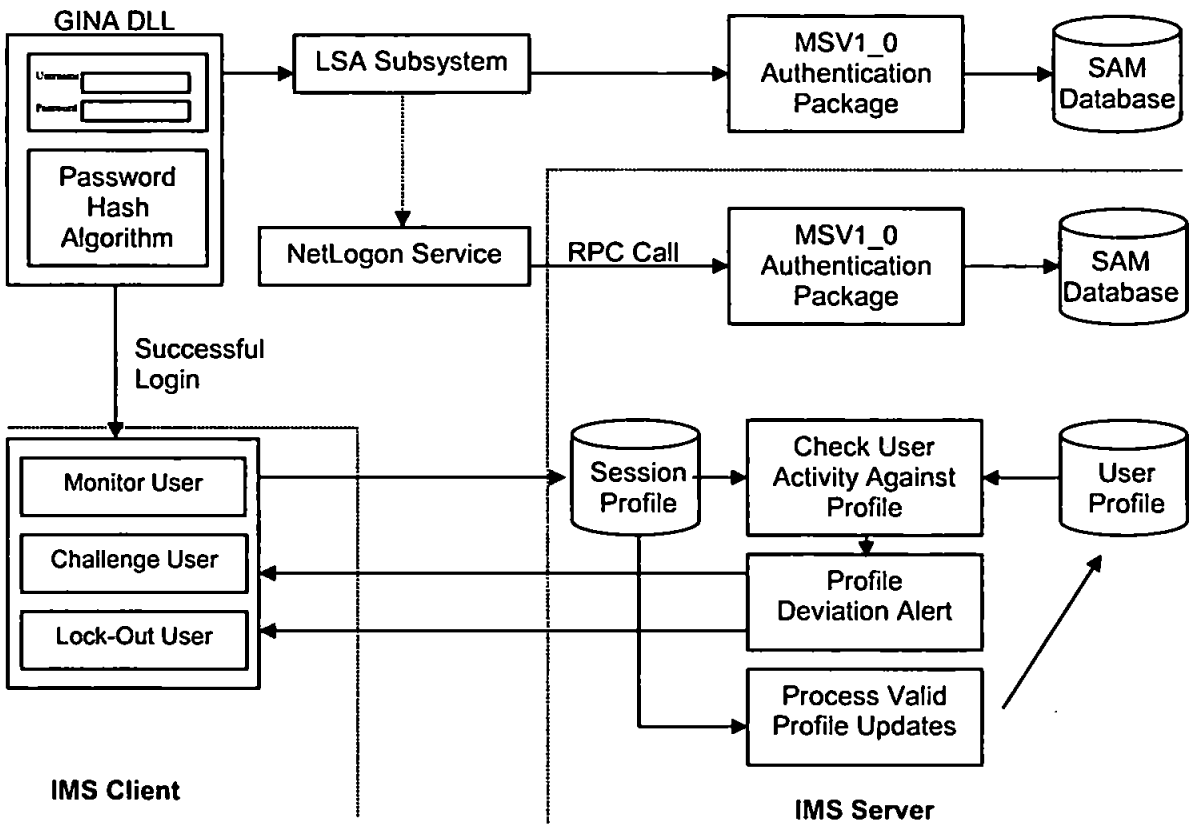


Figure 3 Prototype IMS-NT Integration

To reduce network traffic, it is envisaged that the user authentication would be performed on the local computer with only warnings or profile updates being fed back to the security server. Under certain scenarios it may be necessary to lock local computers if contact is lost with the security server to ensure an intruder had not removed a computer. However, it should be noted that this creates a weak point and appropriate measures will be needed to prevent a single server stopping the entire network, this could take the form of a backup server (in a similar fashion to a secondary DNS server in an Internet context). Alternatively, the range of facilities available to the user can be restricted until the user can be re-authenticated. Another possible weak-point is the profile update process. It is important that the profile update is only performed once a user authentication confidence level is exceeded and it is established that the computer concerned has not been tampered with. In the event that a users' authentication threshold has been uncertain and/or the computer may have been tampered with, any proposed changes to the user profile should be discarded.

One of the most important factors in the implementation of continuous user monitoring is ensuring the transparency of the monitoring process. A system that requires users to continuously re-authenticate themselves will not be successful. Therefore, an IMS should allow background monitoring of an authenticated user, only interrupting the user in the event that further authentication is necessary (e.g. in the form of a challenge-response question).

Clearly an IMS system can provide enhanced user authentication. However, there is no single system configuration that will meet all the needs of all the users. Instead configuration of the security server and client monitoring software is dependent on the level of security required by the organisation and amount of inconvenience that is tolerable to the users (the classic False Acceptance Rate versus False Rejection Rate dichotomy) (Cope, 1990).

Conclusions

As the need for enhanced user authentication grows, operating systems will be extended to provide the necessary services. Windows NT already allows the use of a replacement GINA DLL, which allows OEM security vendors to supplement the Windows NT username/password login with additional/replacement authentication techniques. Alternative login techniques (e.g. fingerprint identification) allow the system confidence in user validity to be increased, but further security is needed to ensure the continued confidence in the user once past the initial login process. A process of continuous user authentication and monitoring, as described in the paper, is therefore desirable.

References

Audit Commission (1998), *Ghost in the Machine – An Analysis of IT Fraud and Abuse*, Audit Commission Publications, UK, ISBN 1-86240-056-3.

Brown, M. and Rogers, S. J. (1993), "User identification via keystroke characteristics of typed names using neural networks", *International Journal of Man-Machine Studies*, p999-1014.

Cherry, A., Henderson, M.W., Nickless, W.K., Olson, R. and Rackow, G. (1992), "Pass or fail: a new test for password legitimacy", Argonne National Laboratory, Mathematics and Computer Science Division, Paper Ref.: MCS-P328-1092,
http://www-proto.mcs.anl.gov/division/publications/abstracts/abstracts92.htm

Cope, J.B. (1990), "Biometric systems of access control", *Electrotechnology*, p71-74, April/May 1990.

Donoghue, A. (1999), "US Army scraps NT for MacOS", *Computing*, p14, 7th October 1999.

Fausett, L. (1994), *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*, Prentice-Hall International, New Jersey, USA, ISBN 0-13-042250-9.

Furnell, S.M. (1995), *Data security in European healthcare information systems*, PhD Thesis, University of Plymouth, UK.

Furnell, S.M., Morrissey, J.P., Sanders, P.W. and Stockel, C.T. (1996), "Applications of keystroke analysis for improved login security and continuous user authentication", *Proceedings of IFIP Sec '96*, Island of Samos, Greece, 21-24 May 1996, pp283-294.

Furnell, S.M., Illingworth, H.M., Katsikas, S.K., Reynolds, P.L. and Sanders, P.W. (1997), "A comprehensive authentication and supervision architecture for networked multimedia systems", *Proceedings of IFIP CMS '97*, Athens, Greece, 22-23 September 1997, pp227-238.

Jobusch, D.L. and Oldehoeft, A.E. (1989), "A survey of password mechanisms: Weaknesses and potential improvements. Part 1", *Computers & Security*, p587-603.

Microsoft Corporation Web Site (1999a),
http://www.microsoft.com/security/issues/deployingc2.asp

Microsoft Corporation Web Site (1999b),
http://www.microsoft.com/security/services/bulletin.asp

NOP Research Group (1999), "E-Commerce in Britain to reach £9.5 billion by 2000",
http://www.nopres.co.uk/survey/internet/internet_item8.htm

Savill, J. (1999), *NT FAQ Web Site*,
http://www.ntfaq.com/ntfaq/security21.html#security21

Stoll, C. (1989), *The Cuckoo's Egg*, Doubleday, New York.

A conceptual architecture for real-time intrusion monitoring

Steven M. Furnell
School of Electronic, Communication and Electrical Engineering, University of Plymouth, Plymouth, UK
Paul S. Dowland
School of Electronic, Communication and Electrical Engineering, University of Plymouth, Plymouth, UK

Keywords
Information technology, Access control, Monitoring, User studies, Computer security

Abstract
The detection and prevention of authorised activities, by both external parties and internal personnel, is an important issue within IT systems. Traditional methods of user authentication and access control do not provide comprehensive protection and offer opportunities for compromise by various classes of abuser. A potential solution is provided in the form of intrusion detection systems, which are able to provide proactive monitoring of system activity and apply automatic responses in the event of suspected problems. This paper presents the principles of intrusion monitoring and then proceeds to describe the conceptual architecture of the Intrusion Monitoring System (IMS), an approach that is the focus of current research and development by the authors. The main functional elements of the IMS architecture are described, followed by thoughts regarding the practical implementation and the associated advantages (and potential disadvantages) that this would deliver. It is concluded that whilst an IMS-type approach would not represent a total replacement for conventional controls, it would represent an effective means to complement the protection already provided.

Information Management & Computer Security
8/2 [2000] 65–74

[ISSN 0968-5227]

Introduction

In typical IT systems, protection against unauthorised user activities is usually provided via login authentication and access controls. The majority of authentication schemes are based upon traditional password methods. The weaknesses of passwords are well-known (Jobusch and Oldehoeft, 1989), but their simplicity (from both user and developer perspectives) serves to ensure their continued use. A significant issue with passwords is that they typically provide a one-off authentication judgement at the beginning of a user session. From that point, protection against unauthorised user activity is reliant upon access controls applied to specific data and resources. Whilst these can be utilised in an effective manner, they are themselves reliant upon appropriate system administration to grant suitable access rights and privileges to users. However, depending upon the level of control imposed, this scenario still offers the potential for unauthorised activity. The normal means of monitoring and identifying this is via audit trails, which maintain a record of nominated security-relevant activities within the system and can be inspected at a later time in order to identify anomalies. The problem with this approach is that any detection of unauthorised activity will be retrospective, when significant damage may already have been done. In addition, previous research findings suggest that while many organisations may maintain audit trail information, only a small percentage (10 per cent) of them actively follow up the information collected (Gliss, 1990). As such, breaches of security may potentially remain unnoticed for some time. What is, therefore, required is an automated, proactive means of detecting and responding to unauthorised access/activity. Such a solution is provided by intrusion detection systems.

The concept of intrusion detection can be traced to original work by Denning (1987), who proposed a model for an intrusion detection system (IDS). This work led to a number of practical IDS implementations from various organisations, in particular the IDES system from SRI (Lunt, 1990). Indeed, intrusion monitoring is still an area of active research[1], indicating that the overall issue has yet to be resolved.

This paper presents an approach to guard against these classes of intrusion and misuse, in the form of the Intrusion Monitoring System (IMS) architecture. The discussion begins by presenting the principles that underlie the intrusion monitoring process, followed by a conceptual description of the IMS architecture. The paper then proceeds to examine a number of issues relating to practical implementation aspects.

Principles of intrusion monitoring

This section presents a number of general principles that underlie the concept of intrusion monitoring and detection, which will enable the architectural approach proposed by IMS to be more fully understood and appreciated.

Categorising system intrusions and misuse

At the highest level, intrusions or misuse will be the result of actions by users or processes, which will operate on one or more targets (which may include data (files), system devices and other users or processes). The purpose of introducing supervision will be two-fold:

1. to ensure that systems are only accessed by authorised users;
2. to ensure that systems are only used for authorised purposes.

User actions can be categorised as being either legitimate or illegitimate. However, it

is useful if a more detailed breakdown than this can be derived for the different potential classes of illegitimate activity. For example, all of the following scenarios represent types of illegitimate activity that should be monitored:

- an illegitimate action that is still within the normal authorisation of a valid user (i.e. abuse of privileges);
- an action by a valid user which is outside the normal limits of authorisation; and
- any action by an unauthorised user.

In addition, it is necessary to recognise differences in the types of potential system abuser. These have already been comprehensively categorised by Anderson (1980), and are described in Table I.

These groupings are considered appropriate for describing the different types of user-related abuse within an intrusion monitoring framework and will, therefore, be adopted for the remainder of the discussion. Whilst it is also possible to develop a deeper profile of potential intruders, by considering factors such as the common motivations behind abuse (e.g. money, ideology, egotism etc.), these are not explored here as knowledge of them would not contribute to the process of detection.

It should be noted that Anderson's categorisations do not take into account any of the categories of abuse that may result from software activity (e.g. viruses, Trojan Horses etc.). This is understandable given that the analysis was made in 1980 before such incidents had become commonplace. However, there has been a significant increase in such attacks over the last decade and evidence suggests that viruses are now the major cause of security breaches in both networked and standalone PC systems (National Computing Centre, 1998). It is now extremely unlikely that the problem will ever disappear and, therefore, countering such activity should also be within the scope of a comprehensive monitor. As a consequence, a further category of intrusion, called malicious process (or malware), can be added to Anderson's list. These may introduce various undesirable consequences, including the alteration or destruction of data, creation of false data, degradation of system performance, crashing of systems or other effects that might render data or systems inaccessible (Brunnstein *et al.* 1990).

Table I
Categories of system abuser

Abuser type	Description
External penetrators	Outsiders attempting to gain unauthorised access to the system
Internal penetrators	Authorised users of the system who access data, resources or programs to which they are not entitled. Sub-categorised into:
	Masqueraders — Users who operate under the identity of another user
	Clandestine users — Users who evade access controls and auditing
Misfeasors	Users who are authorised to use the system and resources accessed, but misuse their privileges

Monitoring and detecting intrusions

The supervision of activities (and resulting anomaly detection) can be based upon user behaviour profiles and generic intrusion indicators. These approaches are common to other intrusion monitoring architectures, such as the IDES system mentioned earlier.

User profiles could conceivably hold a range of identification, authentication and behavioural information relating to registered users. Examples of potential profiled characteristics would include:

- system access times and locations;
- typical levels of system resource utilisation;
- application and file usage;
- methods of user interaction (e.g. GUI versus command line), and
- biometric information (encompassing both physiological and behavioural characteristics).

The use of biometric monitoring is considered to be particularly appropriate to prevent impostor penetration and masquerade attacks. A number of options exist that could be employed in this context, including keystroke analysis (i.e. monitoring of the current user's typing style), face recognition and voice recognition (Miller, 1994).

It is also recognised that some classes of intrusion or misuse can be trapped without identifying departures from historical patterns of user behaviour. As such, generic intrusion rules (also known as attack signatures) may be utilised to identify the occurrence of events that are suspicious in themselves (i.e. irrespective of the user involved). Examples of such generic indicators would include the following:

- consecutive access violations;
- out of hours access;
- account overuse/simultaneous access;
- use of inactive accounts;
- copying of password file;
- extensive use of help systems; and
- modification of an executable file.

While none of these alone would provide sufficient indication to state that an intrusion was in progress, the combination of two or more could be considered more persuasive. In the IMS context, the

occurrence of any such events would increase the alert status of the system (which, as discussed later, could result in a range of potential responses as different threshold values were reached).

A full IMS would operate by comparing current system activity against information held in a knowledge base. The knowledge base would effectively maintain two models of activity for reference by IMS:

1 normal activity (i.e. the user behavioural profiles); and
2 intrusive activity (i.e. the generic rules).

These models will determine what types of activities and events the system will look for and, as such, an event will be judged to be indicative of a suspected intrusion if:

- it is compatible with intrusive activity OR
- it is incompatible with normal activity.

Having considered these principles, the proposed architecture for a practical monitoring system can now be presented.

Intrusion monitoring system architecture

At a high level, the IMS architecture is based upon the concept of a centralised host handling the monitoring and supervision of one or more networked clients running on local workstations. The purpose of the clients is to collect the required data relating to user and process activity and respond to any suspected intrusions detected by the host.

All behaviour profiles, generic rules and such like are maintained securely at the host, which also handles all of the analysis and the main bulk of other processing associated with the supervision. By contrast, the client involves no local data storage and acts almost exclusively as an agent of the host.

At a lower level, the host and client systems will be comprised of a number of modules, each handling a different aspect of the overall intrusion monitoring task, as illustrated in Figure 1. The modules shown are intended to represent the conceptual elements of the system, but could also equate to the coded functional elements in a full implementation. The key aspects of this design are defined in the sections that follow.

Anomaly detector

The *anomaly detector* analyses user and process activity for signs of suspected intrusion, comparing it against the behaviour profiles (class and user-specific) that apply to the current user's (claimed) identity as well as against the generic intrusion rules. In practice, this module will be comprised of a number of further sub-modules, each handling a specific aspect of anomaly detection and behaviour monitoring (e.g. keystroke analysis).

The detector maintains an *alert status table*, with entries existing throughout the life of each user-initiated session or process to indicate the level of detected anomalies and thereby the confidence of a potential intrusion. Each entry contains the basic information shown in Table II, which is examined and updated each time activity data relating to the relevant user/process analysed.

It is envisaged that, at its most basic, the "alert status level" could be a simple aggregate value based on the number of behavioural anomalies detected and intrusion rules satisfied (with the monitored characteristics and rules having been weighted to indicate their significance). The entry relating to "idle time" will be used to allow the phased reduction of the alert status level after certain periods of inactivity. Recording a tally of previous challenges would then be used as a safeguard to determine whether the level of IMS response should be escalated in response to an anomaly even if the alert status is currently low (i.e. as a result of the phased reduction). As Table II illustrates, the alert status table might also be used to store other information, such as the time of session/process initiation or the number of access violations incurred. These would be used for the purposes of on-going comparison against behaviour profiles (for example, session start time could be used to derive the current session length) and would also be required to be maintained throughout the life of the session. It should be noted that some of the table entries are most applicable in the context of monitoring user sessions and will be redundant in the case of process supervision.

The alert status level would increase in response to both departures from a user's historical behaviour profile and the satisfaction of generic intrusion indicators. Under normal circumstances, the detector would commence supervision of a session with an alert status of zero (i.e. no suspicion of an intrusion). However, factors such as failed login attempts, system configuration anomalies and the like could cause it to begin with a non-zero status so that it is essentially more sensitive to further anomalies in the initial instance. The alert status would be reduced after successful challenges or after a sufficient period of normal activity to allow the system to discount the previous anomaly.

Figure 1
IMS architecture

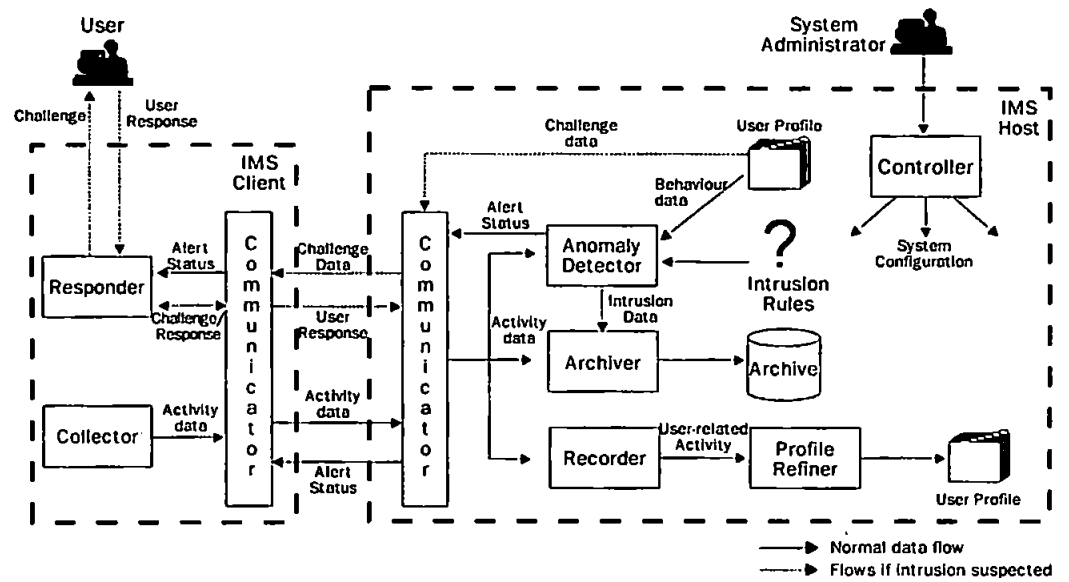


Table II
Structure of alert status table entry

User/process ID	Alert status level	Idle time	# previous challenges	Session start	# access violations

Profile refiner

It is desirable for IMS to utilise user-related activity data in two ways – to analyse for anomaly detection and as the basis for updating behaviour profiles. This second point recognises the possibility that user behaviour may legitimately alter over time (e.g. as a result of access to new applications, improvements in typing ability etc.). The purpose of the *profile refiner* would, therefore, be to provide an automatic means for user-specific profiles to be updated to account for such changes.

It would be most appropriate for the *profile refiner* to be based upon a neural network approach (Bishop, 1995), given that the inherent ability to analyse and recognise patterns could allow behavioural characteristics to be identified that might not be apparent to a human observer. In this way, the effectiveness of the system would have the potential to improve over time, in that it could gradually learn more patterns of legitimate activity for each user (building upon the foundation provided by the generic rules and the initial profiles). It might also be possible to determine which of the profiled characteristics provide the best discriminators for each user and thereby establish (for example) primary, secondary and tertiary level behaviour indicators (with the primary level representing the most reliable identity verifiers). This hierarchy could also be extended to allow for the fact that some characteristics may represent negative indicators (i.e. those that, despite refinement, are found to cause a high level of false alarms).

It would be undesirable for the *profile refiner* to utilise data that is later found to be anomalous. Refinement should, therefore, only take place after the termination of user sessions (provided, of course, that no intrusions were proven during this time). However, it is also considered sensible to allow refinement to proceed if any challenges that were generated were correctly answered by the user (the reason being that the generation of the alert may be indicative that legitimate behaviour has departed from the profile and that refinement is, therefore, necessary). However, in order to help guard against the recognised problem that misfeasors will answer challenges correctly, refinement should be performed on the proviso that the number of alerts raised is small relative to the length of the session (i.e. two alerts in a three hour session would be acceptable, whereas the same number in a ten minute session would be very suspicious). Additionally, any activity occurring during periods where supervision

of the relevant aspect is suspended cannot reliably be used for profile refinement.

User-specific profile records would also incorporate a series of flags to indicate whether the individual behaviour characteristics are ready to be used in supervision or still being developed. This will allow a gradual training period to be defined for new user profiles without the IMS continually generating intrusion alerts (the flags would also allow a specific refinement only period to be established for existing profiles that have proved to be inadequate for the legitimate user). The purpose of associating flags with each profile characteristic is so that some degree of monitoring could still continue whilst other aspects are being (re)trained. The flags could also be used to allow the total disablement of some aspects of monitoring if, for example, some characteristics are found to be inappropriate to certain users.

Data relating to process activity would not be used for refinement as the generic rulebase would remain static (unless specific information on new intrusion methods is introduced by the system administrator).

Recorder

The *recorder* handles the short-term storage of user-related activity data during the period of a user session and focuses specifically upon the collection of data relating to the profiled characteristics of a given user (e.g. collection of keystroke data in relation to the typing profile). Upon termination, the information will be used as input to the *profile refiner*, provided that the session was not considered anomalous. In the event of a proven anomaly, the *recorder* can discard its stored information for the session.

Archiver

The *archiver* collects data relating to all system activity and stores it in a long-term archive (in the same manner as a traditional audit trail), providing a more permanent record of activities and suspected anomalies. The storage will occur regardless of whether sessions/processes are regarded as anomalous and details of all security relevant events will be archived. Such events will include login failures, intrusion alerts, authentication challenges, suspended sessions and the like. The basic format of the archive records would be as shown in Table III.

However, in order to conserve storage space, it may be desirable in some scenarios to only record details of certain types of event. The *archiver* would, therefore, be configurable to suit the preferences of the establishment involved (note that the same would not necessarily be true for the *recorder* as this would always need to collect information on any activities for which profile refinement may later occur). The long-term retention period of archived details would be determined by the security policy of the organisation involved.

Collector

The *collector* represents the interface between the IMS and the existing information system/applications, with the responsibility for obtaining information on all relevant user and system activity. The module would be required to operate in such a way as to encompass, but be independent of, all system applications. It is envisaged that this could be best achieved by implementation at the operating system (OS) level, such that key events also lead to IMS notification. For example, a significant proportion of data collection could be based around the interception and redirection of selected OS interrupts and service requests (such as file input/output, application execution, keyboard input). These would be monitored with two objectives:

1. to collect data on those events which pertain to monitored behaviour characteristics; and
2. to identify those events which may affect the security of the system (for comparison against generic intrusion indicators).

In some cases, the required data could be obtained directly from existing audit trail records on the underlying system. However, with certain aspects (e.g. keystroke analysis) the required information will not be maintained in audit trails and implementation may, therefore, require a significant number of operating system links. Whilst this would serve to make this aspect of IMS very system specific, it would be considerably more efficient than attempting to modify each individual application to specifically provide relevant information to IMS. The system specific coding of the *collector* would only need to be done once, whereas modifications would be required to all current and future applications (which would be likely to be a non-trivial

Table III
IMS archive record structure

Date	Time	User/process ID	Logged event	Privileges	Resources utilised

undertaking and potentially impossible in the case of commercial packages where the source code may be unobtainable).

As with the configuration of the *archiver*, the resolution of data collection would be determined at the host by the system administrator.

Responder

This module resides in the IMS client and handles the task of responding to anomalies detected by the host. The operation of the *responder* would centre around the continuous monitoring of the alert status transmitted by the host, with increases in the level triggering appropriate actions. The nature of the response might include the issue of a user authentication challenge, suspension of a session or cancellation of a process. The issue of appropriate response is discussed in more detail later in the paper.

In some implementation scenarios, the *responder* might also be responsible for handling the initial user identification and authentication process that is required to gain access to the system in the first instance.

Communicator

The *communicator* provides the network communications interface between the host and the client(s) operating on the local systems. As such, the functionality of this module is duplicated on both sides of the link. The principal functions would include transmitting user and process information to the host and then subsequently keeping the client(s) informed of the current alert status. If implemented in a heterogeneous environment, the client side of the module would be responsible for resolving any operating system differences that exist within the monitoring domain so that information could be presented to the host in a consistent, standardised format.

Controller

This module is provided for use by the system administrator to allow the operation of the IMS system to be configured. On the host side, such a configuration would apply to the following modules:

- *anomaly detector*, e.g. behaviour characteristics to consider/prioritise, generic rules in operation;
- *profile refiner*, e.g. frequency of refinement, acceptable thresholds for challenges within a session; and
- *archiver*, e.g. level of detail required, specific events to record or exclude from logging.

For the client side, the operation of the following modules would be controlled:

- *collector*, e.g. the level of data collection (linked to the characteristics being monitored by the *anomaly detector*), and
- *responder*, e.g. the level of response required at each alert status level.

These settings would obviously be controlled and recorded through the host system. The configuration of the local client(s) would then be established at the time of session initiation.

In addition to the above, several other features would also be provided under the auspices of the *controller* module. These would include facilities such as user profile management, update of the generic rulebase and the like.

User profiles

IMS profiles could conceivably hold a range of identification, authentication and behavioural information relating to legitimate users. The profiles would use a number of methods to represent measures of user behaviour:

- frequency tables (e.g. for file access);
- means and standard deviations (e.g. for keystroke/typing profiles);
- ranges (e.g. valid access times);
- lists (e.g. for valid access locations); and
- a combination of methods (e.g. a list of valid access locations which also indicate the relative frequency of use).

The profile data obviously require secure storage to prevent unauthorised browsing or tampering by potential impostors. If users were able to modify profile information it would be possible for them to adjust the records of other users to match their own (and, therefore, allow them to access the account in place of the legitimate owner). Whilst disclosure of the profile statistics may not initially appear to pose such a threat, it could still be a problem in the case of a determined impostor. For example, if the characteristics of the "target" user were known, the impostor would have a concrete statement of what he/she would be required to mimic. An alternative option would, of course, be to subsequently enlist the help of an accomplice with a comparable profile. At the very least, this dictates a requirement for encrypted storage, as used with the password files in the majority of commercial operating systems (Morris and Thompson, 1978).

Issues related to intrusion monitoring

This section presents further discussion of a number of the issues that were mentioned during the description of the IMS modules. The issues in question are the restriction of user activities, suspension of supervision and types of response to suspected intrusion.

Restriction of user activities

It is considered feasible for the alert status level to be inter-linked with the types of activity that a subject is allowed to perform, such that a phased reduction of permitted behaviour would occur as the level increases. In this way, highly sensitive activities and/or information could be denied if there is any doubt over the current user's legitimacy, whilst still allowing more mundane activities to continue. The approach would demand that a maximum alert status threshold be associated with each of the activities or objects that the IMS is to control. If the current status level was then to exceed this, the activity or object would become unavailable. For example, consider the thresholds in Table IV associated with two objects (wordprocessor and database) and the activities create and delete file. If the current alert status level were five then the user would not be permitted to access the database or to perform any file deletion. However, the creation of a file using the wordprocessor application would still be possible.

Such a threshold table would be maintained within IMS, but the values would initially need to be assigned (and, if necessary, subsequently updated) by the system administrator. It must be said that the potential for error would make this approach inappropriate in many scenarios (for example, the denial of data access in sensitive applications could be most unwelcome). In any case, it would be advisable for the system administrator to be notified whenever behaviour restrictions were being imposed so that the situation could be investigated (in case legitimate users were being unintentionally impeded).

Table IV
Alert status threshold table

Activity/object	Alert status threshold
Wordprocessor	8
Database	2
Create file	8
Delete file	3

Suspension of supervision

In some cases it is envisaged that continuous behaviour monitoring at all times throughout a user session may not be strictly necessary or even advantageous. This is especially true in the case of the mechanisms aimed solely at the detection of penetrators (e.g. keystroke analysis). The rationale here is that, after a reasonable amount of uninterrupted behaviour analysis (i.e. with no challenges and no significant periods of user inactivity), the monitoring system should have been able to accurately determine the legitimacy of the current user (e.g. previous research has indicated that, using keystroke analysis, a reliable authentication judgement should be obtainable within 400 keystrokes in a real-time monitoring context (Furnell, 1995)). If an impostor is not suspected at this point then it is extremely unlikely that further monitoring will detect one (indeed, monitoring for longer than is necessary would simply allow more opportunity for false rejections to occur and place an additional load on the system). In view of this, it is considered that monitoring activity during the following periods is likely to be most crucial in terms of impostor detection (with supervision being temporarily suspended at other times):

- during the period immediately after the start of the session (when the authenticity of the user has yet to be conclusively proven);
- during the time after any significant period of inactivity (during which an impostor could potentially have replaced the legitimate user).

Important considerations here would obviously be the period of monitoring necessary before suspension of supervision and also what length of time would constitute the significant period of inactivity necessary for it to be resumed. Suggested periods would depend upon the monitored characteristics (e.g. monitoring of keystroke dynamics could yield an authentication judgement more quickly than monitoring of application usage), but a general rule could be to monitor up to five minutes of non-anomalous activity before suspension in order to allow a sufficient appraisal of the user to be made. Approximately two to three minutes of inactivity would then be seen as a suitable trigger for monitoring to resume, as this length of time could have allowed sufficient opportunity for impostor intervention. In a practical implementation, both of these aspects would be configurable so that the optimum levels could be established.

It should be noted that this approach would not be adequate for the detection of misfeasor activity, as this could very well proceed after authentication has been established. Therefore, if suspension of monitoring was still to be incorporated, it would be sensible to periodically reintroduce supervision at random intervals as an additional safeguard (this would also help to guard against a situation where an impostor/penetrator might be able to replace the authorised user without there being a significant period of inactivity).

This idea is primarily suggested as a means of minimising the likelihood of false rejections in the practical context. However, a further advantage in the context of practical implementation would be that it would reduce the significant processing overhead that would be associated with continuous monitoring in an environment with a large number of client machines.

Response to suspected intrusions

The existence and operation of IMS should ideally remain transparent to the user unless an anomaly is suspected. As previously stated, a suspected intrusion will cause IMS to automatically perform some further action (the nature of which will vary depending upon the type of intrusion involved). Options here include:

- issuing of an explicit request (or challenge) for further authentication;
- recording of details in an intrusion log for later inspection/investigation;
- immediate notification of the system manager (i.e. an intrusion alarm);
- phased reduction of permitted behaviour;
- locking of the intruder's terminal;
- termination (or suspension) of the anomalous session/process.

The degree of automatic response is an important consideration and, as indicated above, must be matched to the severity of the suspected intrusion. For example, if there is high confidence that an activity represents an intrusion or if a particularly serious breach is suspected, then the maximum countermeasure response should result. However, in lesser scenarios more limited responses will be appropriate (e.g. simply writing details to the intrusion log).

There is an obvious danger that any option which allows the user to continue working whilst the anomaly is investigated would also allow more time for an intruder to cause damage. At the other extreme it would be undesirable for the system to terminate a session or process without a very high degree of certainty that an intrusion was in progress. Therefore, the first two options above are considered to be the most appropriate as initial forms of response.

In practice, there are several possibilities for the type of challenge that the system could issue in the event of a suspected intrusion. The original system password would obviously be inadequate, given that it may have already been compromised in order for an intruder to have gained access in the first place. It is desirable that the challenge be such that it allows any legitimate user to resume work quickly with minimal interruption (i.e. it should be easy for them to overcome, whilst still trapping impostors). A suggestion is that a (short) series of question and answer type challenges be posed to the user (Haga and Zviran, 1991), who would then need to answer them correctly in order to proceed further. These could be based upon cognitive and/or associative information, with valid responses having been obtained and stored in conjunction with the original user profiling. If several (e.g. five to ten) such questions were to be obtained from users during profiling then the challenge could be based upon a random selection from the set (further reducing the chance of impostors being able to compromise the system).

There are, however, a number of scenarios in which this approach would be ineffective. Firstly, it must be remembered that any form of authentication-based challenge would be an inadequate countermeasure against misfeasors. They would obviously be able to respond correctly to such challenges (having originally supplied the information themselves) and then continue with unauthorised activity. There is a solution here in the realisation that continuing anomalies would lead to a succession of intrusion alerts; an event that would be suspicious in itself. At this point, the IMS response could then change to a method that would effectively combat misfeasors as well (e.g. a session lock or a trigger for system manager investigation). Nevertheless, this would still enable misfeasors to continue for longer than other classes of intruder (albeit with intermediate challenge(s)) before the system locks them out.

A second problem/exception relates to suspected malicious processes – these cannot be issued with a challenge to which they may respond and verify their legitimacy. This in turn places more importance on the correctness of the resulting IMS response (e.g. the dangers of suspending/deleting a legitimate, and possibly essential, process or failing to take positive action against a genuinely destructive one).

Finally, some classes of anomaly (for example, login failures based on unrecognised user identities) cannot be tied to a specific user and, as such, the issue of a challenge based upon profile information is again inappropriate. However, it is conceivable that some form of generic challenge could be issued (the answer to which would be known by legitimate system users), with invalid responses causing the IMS to proceed to its next level of countermeasure (e.g. system manager notification, terminal lockout).

IMS implementation issues

The IMS concept is considered most appropriate to implementation in a networked environment, for the following reasons:

- Standalone systems will most often be dedicated to a single user. As such, more traditional authentication and access controls (e.g. passwords) will probably be sufficient to ensure security if they are correctly implemented.
- Implementation of a full IMS would be likely to degrade the performance of a standalone system.
- Networked systems provide more potential for collecting monitoring information. Many statistics (e.g. access location, resource usage) would not be appropriate to a standalone environment.

In this scenario, the host would be centralised with multiple IMS clients being used to monitor activity on the individual workstations. The purpose of the clients would be to collect any activity data that is generated locally (e.g. keystroke timings) and to enforce IMS restrictions in suspected intrusion scenarios (e.g. issue a challenge, lock access to the system etc.). In such a scenario it would be necessary to maintain the security of the IMS clients on the individual machines to ensure that their operation cannot be compromised (e.g. by a malicious user trying to avoid detection).

The realisation of the IMS approach is considered to have a number of advantages, as listed below.

- *Improved security.* This is advantageous in any information system, and is achieved here due to the continuous nature of supervision. User authentication is no longer restricted to the discrete judgement(s) possible with passwords' and misuse will be identifiable much earlier than with traditional auditing. In addition, the fact that much of the supervision is based upon behavioural characteristics makes it more difficult for users themselves to undermine security (e.g. by allowing colleagues unauthorised access to their accounts) as they cannot easily transfer these abilities to other users.
- *Cost.* Advantages here result from the fact that it is possible to implement the concept entirely in software at the user end, whereas many frequently suggested authentication enhancement schemes (e.g. smart cards, other biometric methods) are reliant upon specialised equipment at each user workstation. This makes the technique particularly suited to financially constrained environments.
- *Convenience.* This comes from the fact that the supervision can be performed transparently, in a non-intrusive manner. In addition, the fact that the IMS would demand nothing special from the users (e.g. they are not required to remember additional password-type information or possess any physical token) means that its operation should not undermine the existing staff culture, which is recognised as an important issue in the introduction of security (Warren *et al.*, 1995).

There are also a number of inherent disadvantages in the concept of IMS (and, indeed, any other type of comprehensive monitoring and supervision system). The principal concerns are highlighted below.

- The operation of IMS clients and/or data collection will consume system resources and may degrade overall performance. The collection of detailed audit trail data typically degrades machine performance by between 5 and 20 per cent (Wolfe, 1992; Mukherjee *et al.* 1994). An IMS performing full behavioural monitoring and testing of generic intrusion rules would be envisaged to introduce a similar burden.
- Transmission of data from clients to the host will result in a loss of network bandwidth and a loss of timeliness of data.
- Maintenance of the IMS itself would entail a more significant management/ administration burden in the affected systems. For example, correcting problems with behaviour profiles would be a more complex operation than cancelling a forgotten password. At the same time, however, other duties (such as inspection of audit trails) would be reduced, so the new demands would at least be somewhat offset.
- The overall concept of continuous supervision raises a question of user acceptance. It is conceivable that there may be mistrust and resentment of the

system on the grounds of it being seen as a means of monitoring legitimate work and staff performance as opposed to just guarding against intruders. It would, therefore, be important to ensure that the system is perceived as a caring mother rather than a big brother.

In general terms, the likely advantages when compared to other means of protection are considered sufficient to outweigh these points. In view of this, the authors are currently developing an implementation of the IMS approach for the Windows NT environment (Dowland and Furnell, 2000).

Conclusions

The paper has described the concept of intrusion monitoring and a potential approach by which it may be realised in modern networked systems. The IMS concept is not intended as a total replacement for conventional authentication and access control methods (although in some cases it will offer an opportunity for more dated approaches to be replaced). In the majority of systems, supervision could be incorporated alongside other methods to complement the security already provided. In addition, it will have little or no effect upon the need for physical security and personnel-related measures within an organisation. There are also some important aspects of logical security that are not addressed (e.g. protection of data communications) which further highlight the potential need for a wider IT security framework.

Note

See, for example, http://www.securitysearch.net/Research_and_Education/Intrusion_Detection_/

References

Anderson, J.P. (1980), *Computer Security Threat Monitoring and Surveillance*, James P. Anderson Co., April, Fort Washington, PA.

Bishop, C.M. (1995), *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford.

Brunnstein, K., Fischer-Hubner, S and Swimmer, M. (1990), "Classification of computer anomalies", in *Proceedings of 13th National Computer Security Conference*, 1-4 October, Washington, DC, pp. 374-84.

Denning, D.E. (1987), "An intrusion-detection model", *IEEE Transactions on Software Engineering*, SE-13(2), pp. 222-32.

Dowland, P.S. and Furnell, S.M. (2000), "Enhancing operating system authentication techniques", to appear in *Proceedings of the Second International Network Conference (INC 2000)*, 3-6 July, Plymouth.

Furnell, S.M. (1995), "Data security in European healthcare information systems", PhD Thesis, University of Plymouth, UK.

Gliss, H. (1990), "A survey of computer abuse", in *Proceedings of Compsec 90 International*, 10-12 October, London, pp. 495-517,

Haga, W.J. and Zviran, M. (1991), "Question-and-answer passwords: an empirical evalution", *Information Systems*, Vol. 16 No. 3, pp. 335-43.

Jobusch, D.L. and Oldehoeft, A.E. (1989), "A survey of password mechanisms: part 1", *Computers & Security*, Vol. 8 No. 7, pp. 587-604.

Lunt, T.F. (1990), "IDES: an intelligent system for detecting intruders", in *Proceedings of the Symposium: Computer Security, Threat and Countermeasures*, November, Rome, Italy.

Miller, B. (1994), "Vital signs of identity", *IEEE Spectrum*, February.

Morris, R. and Thompson, K. (1978), "Password security: a case history", in *UNIX Time-Sharing System: UNIX Programmer's Manual, seventh edition, Vol. 2*, Bell Laboratories (1983), pp. 595-601.

Mukherjee, B., Heberlein, L.T. and Levitt, K.N. (1994), "Network intrusion detection", *IEEE Networks*, Vol. 8 No. 3, pp. 26-41.

National Computing Centre (1998), *BISS '98 – Information Security. The True Cost to Business*, National Computing Centre Limited, Manchester, UK.

Warren, M.J, Sanders, P.W. and Gaunt, P.N. (1995), "Participational management and the implementation of multimedia systems", in *Proceedings of MEDIACOMM 95 – International Conference on Multimedia Communications*, Southampton, UK, pp. 131-5.

Wolfe, A.D. (1992), "Securing the distributed environment: a question of trust", *Patricia Seybold's Network Monitor*, Vol. 7 No. 1, pp. 3-19.