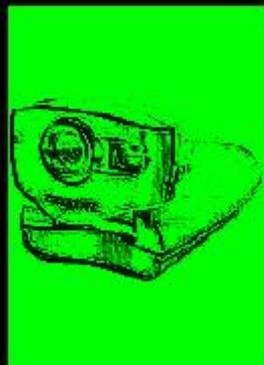




Advances in Networks, Computing and Communications 4

Proceedings of the MSc/MRes Programmes from the
School of Computing, Communications and Electronics

2005 - 2006



Edited by

Paul S Dowland
Steven M Furnell

Advances in Networks, Computing and Communications 4

**Proceedings of the MSc/MRes Programmes from the
School of Computing, Communications and Electronics**

2005 - 2006

Editors

Dr Paul S Dowland

Prof Steven M Furnell

School of Computing, Communications & Electronics
University of Plymouth

ISBN: 978-1-8410-2180-5

© 2007 University of Plymouth
All rights reserved
Printed in the United Kingdom

No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means – electronic, mechanical, photocopy, recording or otherwise, without the prior written permission of the publisher or distributor.

Preface

This book is the fourth in a series presenting research papers arising from MSc/MRes research projects undertaken by students of the School of Computing, Communications and Electronics at the University of Plymouth. These one year masters courses include a significant period of full-time project activity, and students are assessed on the basis of an MSc or MRes thesis, plus an accompanying research paper.

The publications in this volume are based upon research projects that were undertaken during the 2005/06 academic year. A total of 33 papers are presented, covering many aspects of modern networking and communication technology, including security, mobility, coding schemes and quality measurement. The expanded topic coverage compared to earlier volumes in this series reflects the broadening of our range of MSc programmes. Specifically contributing programmes are: Network Systems Engineering, Robotics, Information Systems Security, Web Technologies and Security, Computing and Interactive Intelligent Systems

The authorship of the papers is credited to the MSc/MRes student in each case (appearing as the first named author), with other authors being the academic supervisors that had significant input into the projects. Indeed, the projects were conducted in collaboration with supervisors from the internationally recognised research groups within the School, and the underlying research projects are typically related to wider research initiatives with which these groups are involved. Readers interested in further details of the related research areas are therefore encouraged to make contact with the academic supervisors, using the contact details provided elsewhere in this publication.

Each of the papers presented here is also supported by a full MSc or MRes thesis, which contains more comprehensive details of the work undertaken and the results obtained. Copies of these documents are also in the public domain, and can generally be obtained upon request via inter-library loan.

We believe that these papers have value to the academic community, and we therefore hope that their publication in this volume will be of interest to you.

Prof Steven Furnell and Dr Paul Dowland

**School of Computing, Communications and Electronics
University of Plymouth, May 2007**

About the School of Computing, Communications and Electronics

The School of Computing, Communication and Electronics has interests spanning the interface between computing and art, through software, networks, and communications to electronic engineering. The School contains 61 academic staff and has over 1000 students enrolled on its portfolio of taught courses, over 100 of which are at MSc level. In addition there is a similar number of postgraduate research students enrolled on a variety of research programmes, most of which enjoy sponsorship from external sources.

The bulk of the staff in the School are housed in the Portland Square building, a purpose built state of the art building costing over £25million and situated near the centre of the historic city of Plymouth on the University campus. The laboratories are located in the newly refurbished Smeaton Building, and the Clean room for nanotechnology also recently refurbished courtesy of a Wolfson Foundation grant is situated in the nearby Brunel Building. All buildings are a short walk from each other, enabling a close collaboration within our research community.

This School sits alongside two other Schools in the Faculty of Technology, the School of Engineering (the merged School of Civil and Structural Engineering and Department of Mechanical and Marine Engineering), and the School of Mathematics and Statistics. There are research and teaching links across all three schools as well as with the rest of the University. The closest links are with the Faculty of Science, principally the Centre for Computational and Theoretical Neuroscience which started in Computing, and Psychology through Artificial Intelligence and Human Computer Interaction research.

Prof Phil Dyke
Head of School

Contributing Research Groups

Centre for Interactive Intelligent Systems

Head: Professor E Miranda & Professor A Cangelosi

Email: eduardo.miranda@plymouth.ac.uk

Research interests:

- 1) Natural language interaction and adaptive systems
- 2) Natural object categorisation
- 3) Adaptive behaviour and cognition
- 4) Visualisation
- 5) Semantic web

http://www.tech.plymouth.ac.uk/Research/computer_science_and_informatics/

Centre for Robotics and Intelligent Systems

Head: Dr G Bugmann

Email: guido.bugmann@plymouth.ac.uk

Research interests:

- 1) Cognitive systems
- 2) Social interaction and concept formation through human-robot interaction
- 3) Artificial intelligence techniques and human-robot interfaces
- 4) Cooperative mobile robots
- 5) Visual perception of natural objects
- 6) Humanoid robots

<http://www.tech.plymouth.ac.uk/socce/ris/>

Fixed and Mobile Communications

Head: Professor M Tomlinson BSc, PhD, CEng, MIEE

E-mail: mtomlinson@plymouth.ac.uk

Research interests:

- 1) Satellite communications
- 2) Wireless communications
- 3) Broadcasting
- 4) Watermarking
- 5) Source coding and data compression

<http://www.tech.plymouth.ac.uk/see/research/satcen/sat.htm>

<http://www.tech.plymouth.ac.uk/see/research/cdma/>

Interdisciplinary Centre for Computer Music Research

Head: Professor E Miranda

Email: eduardo.miranda@plymouth.ac.uk

Research interests:

- 1) Computer-aided music composition
- 2) New digital musical instruments
- 3) Sound synthesis and processing
- 4) Music perception and the brain

<http://cmr.soc.plymouth.ac.uk>

Network Research Group

Head: Professor S M Furnell

E-mail info@network-research-group.org

Research interests:

- 1) Information systems security
- 2) Internet and Web technologies and applications
- 3) Mobile applications and services
- 4) Network management

<http://www.network-research-group.org>

Contents

SECTION 1 Network Systems Engineering

Implementing Network Monitoring Tools V.C.Asiwe and P.S.Dowland	3
Recording end-users security events: A step towards increasing usability D.Chatziapostolou and S.M.Furnell	11
An investigation on the relationship of aggregated TCP and UDP traffic M.Davy and B.V.Ghita	19
User Awareness of Biometrics B.J.Edmonds and S.M.Furnell	30
Analysis of End-to-End Techniques for Bottleneck Bandwidth & Path Capacity Estimation T.Edwan, B.Ghita and X.Wang	38
Public awareness of biometrics K.Evangelatos and S.M.Furnell	51
The Awareness and Perception of Spyware amongst Home PC Computer Users M.Jaeger and N.L.Clarke	60
Investigate Placement of Relaying Node in Wireless Mesh Networks D.Jing and X.Wang	72
Evaluation of Pervasive and Ubiquitous Healthcare Systems K.A.Khan and X.Wang	80
Web-Based Risk Analysis and Education for Home Users J.Marston and N.L.Clarke	89
The Art of Network Monitoring A.Mohyuddin and P.S.Dowland	100
Security Considerations for a Wireless Local Area Network O.I.Nwobodo and X.Wang	106
VoIP Security Threats and Vulnerabilities S.M.A.Rizvi and P.S.Dowland	114

Advances in Networks, Computing and Communications 4	
Evaluation of Grid Computing Security E.Vahedi-Sarrigani and X.Wang	123
A Performance of network coding in randomised settings I.C.Tjhai and L.Mued	130
Investigation on Static Network with Network Coding L.Yang and L.Mued	138
A guide for small and medium enterprise of implementing security and firewall system R.Zhang and P.S.Dowland	146

SECTION 2 Information Systems Security & Web Technologies and Security

Information Security Awareness & Training H.Al-Ghatam and P.S.Dowland	157
Security Technologies: Why are they not used correctly? M.Al-Tawqi and S.M.Furnell	164
Web-based Plankton Data Visualisation T.Y.Aung and P.S.Dowland	173
Improving protection and security awareness amongst home users P.Bryant, S.M.Furnell and A.D.Phippen	182
User security awareness of social engineering and phishing A.Karakasiliotis, S.M.Furnell and M.Papadaki	191
Evaluating the Perceptions of People towards Online Security N.K.Jayakumar and A.D.Phippen	199
Intrusion Detection System for mobile devices D.S.Michalopoulos and N.L.Clarke	205
Keystroke analysis as an authentication method for thumb-based keyboards on mobile handsets S.Karatzouni and N.L.Clarke	213
Strengthening the Human Firewall G.C.Tjhai and S.M.Furnell	222

SECTION 3 Computing, Robotics & Interactive Intelligent Systems

Evolution of musical lexicons by singing robots E.Drouet and E.R.Miranda	233
eGovernment take-up in the city of Plymouth, UK G.Ford and A.D.Phippen	247
The ‘Speech Music’ application and language variant analysis C.C.Ford and S.L.Denham	255
Autonomous robot navigation in buildings: a case study using the Evolution ER1 robot D.A.E.Harewood-Gill and T.Belpaeme	264
Information systems integration in virtual learning environments N.Mcilree and A.D.Phippen	273
Real-time granular synthesis with spiking neurons J.Murray and E.Miranda	278
Emotional speech recognition: a neural network pruning approach S.Saqib, G.Bugmann and E.Miranda	286
Author Index	294

Section 1

Network Systems Engineering

Implementing Network Monitoring Tools

V.C.Asiwe and P.S.Dowland

Network Research Group, University of Plymouth, Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

One very vital tool for networks is the network monitoring tool. These tools are part of the network because of their role in allowing administrators to observe and analyse the network at any time. This observation and analysis are fundamental to the smooth running of the network because they aid in managing network resources and ensure that the performance of the network is not hindered. This paper describes the implementation network monitoring tools that have functionality capable of performing simple monitoring tasks as well as packet capturing. The tool was designed and implemented under a Windows environment and it was designed to be user-friendly and simple while at the same time providing user with enough functionality for monitoring their networks. The effectiveness of the tools implemented was evaluated with the aid of a survey and the results analysed shows that it meets it requirement specification.

Keywords

Networks, Network monitoring, Network monitoring tools, implementation, survey

1. Introduction

Networks evolved out of the need for sharing information and for communication between two or more computers. As these needs increase, networks grew in size, operational cost and complexity. Presently the size of a network can span the globe and can be as complex as having many networks connected together as a single network. The operational cost involves service interruption as a result of abnormal conditions and failure of a network device. This growth in turn presented severe problems because many organisations could not justify the use of the network as a result of the poor Return on Investment (ROI) recorded. The first step in the solution of the problems brought by the growth of the network is to monitor the network to identify certain trends and proactively solve these problems (Held, 2000; Gaglio et al. 2006).

Monitoring a network involves the use of specialized tools that can keep track of the status of all the various devices on the network; identify and analyse incoming and outgoing traffic; identify problem areas and check for certain trends by alerting the network administrator of their occurrence. These tools provide means by which the configuration settings of the network can be managed, the performance of a network can be evaluated and any faulty condition diagnosed. This research was conducted to implement a set of simple network monitoring tools under the Windows platform. The rest of this paper will be structured to give a thorough understanding of the concepts involved in the implementation. Section 2 introduces the concepts involved and detailed the approach used in creating the network monitoring tools. Section 3

introduces the research methods used in conducting this research. Section 4 introduces how the implementation of the tools was evaluated. Finally, section 5 presents the conclusions.

2. Network Monitoring Tools

2.1 Network Monitoring

Network monitoring involves observing and analysing the status and behaviour of part or the entire network that create what needs to be monitored and managed (Wisniewski, 2003). Network monitoring takes place at layers 1, 2, 3 and 4 of the Open Systems Interconnection (OSI) reference model. At layer 2, network monitoring uses MAC addresses to capture frames as they enter and exit the network. At layer 3, network monitoring uses source and destination IP addresses for packet capture and at layer 4 monitoring is done using source and destination port numbers and protocols like TCP and UDP (Held, 2000; Miller, 1999). This is depicted in figure 1 below.

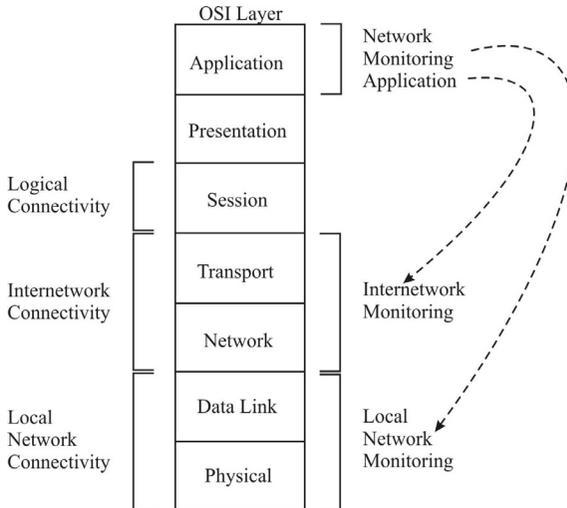


Figure 1: Network monitoring within the OSI framework (Miller, 1999)

Network monitoring is very vital to network management because a network to be managed must first be monitored. Leinward and Fang (1993) described network management as a process of controlling a complex network to maximise its efficiency and throughput. Burke (2004) noted that network management is mostly a combination of local and remote configuration and management with software. In this vein, network monitoring can be said to be a subset of network management. Liska (2003) noted that logging is a very vital aspect of network monitoring and it notifies an administrator via e-mail, telephone or Simple Message Service (SMS) when trappable conditions occur.

2.2 Network Monitoring Information

Network monitoring information can be classified into static, dynamic and statistical information. Static information is characterized by the current information, as such, it will rarely change. Static information is always generated by the various network devices that are monitored. Dynamic information is related to events happening in the network in real-time. Statistical information can be generated by network devices that have access to dynamic information. This information is collected, analysed and summarised statistically producing bar chart, pie chart, histogram and graphs as the need arises (Wisniewski, 2003).

2.3 Network Monitoring Tools

Network monitoring tools gather information about the general condition of the entire network to identify those areas that are failing and need managing. These tools can also check the overall performance of a network against a baseline taken when everything was working perfectly well. With these tools, the network administrator can observe the operation and performance of network infrastructures (Held, 2000). These tools provide network monitoring information with which the network is analysed.

2.4 An Overview of Existing Network monitoring Tools

Ping is a command line utility that tests for connectivity by checking if a destination can be reachable from a source in an IP network. It uses the sends ICMP echo request packets and listens for ICMP echo reply packets to accomplish the connectivity test (Cheswick and Bellovin, 1994; Leinward and Fang, 1993; 2006; Wisniewski, 2003). Traceroute is a command line utility that determines the route a packet takes from its source to reach its destination. Tcpcmdump is the most used tool for network monitoring and data acquisition. It allows capturing and display of TCP/IP packets as they are being sent and received to and from a network adapter. It allows us to precisely see all network traffic. It provides a standard packet capture interface, a common dump format, basic packet decoding features and can filter based on user specification (ComLab, 2006; IEPM Website, 1999). Windump is the equivalent of tcpcmdump, but used for Windows. Windump can be used to watch, diagnose and save the network traffic based on the rules the user specify (Windump, 2006). WildPackets Etherpeek is a portable Ethernet-specific network analyser that allows for visibility into every part of the network. (WildPackets, 2006).

This research implements a suite of network monitoring tools from within a single Graphical User Interface (GUI). The remaining subsections discuss the implementation of the tools.

2.5 The Network monitoring Tools Implemented

The network monitoring tools implemented by this research provide a means of gathering information from frames, packets and protocols. As with all monitoring tools, it has the capability to track outstanding problems by using administrative alerts via Yahoo! Mail when faulty or undesirable network conditions occur. As part

of its functionality, it can translate between IP address and host name and vice versa, display the arp table, test for connectivity between two network devices using ping, trace the route taken by a packet to its destination, capture packets for analysis, monitor network traffic by displaying the packets sent and received on the network interface and determine the throughput of the interface.

2.6 The Network Monitoring Tools Design Approach

The design approach taken makes use of a software solution based on layers 2, 3 and 4 of the OSI model. This approach entails making the implementation user-friendly, simple while still maintaining the key features and easy to expand and customize.

2.7 The Network Monitoring Tools Implementation

The network monitoring tools were implemented using a PC having an AMD Athlon™ XP 1.5GHz processor with a 256MB of physical memory and a network adapter as the hardware platform. The software platform for the implementation was done using Microsoft Visual Basic (VB) and Microsoft Windows XP. Visual basic was used because the systems development methodology used was prototyping and this is a Rapid Application Development (RAD) methodology and Visual Basic supports RAD. The implementation relied extensively on the use of Application programming Interface (API) and Windows socket (Winsock) programming.

2.8 Forms Used for the Implementation

The software was implemented using Microsoft VB. VB uses forms as its visual elements. The software comprises a lot of forms. The notable ones are discussed below:

Once the application has started, access can only be granted while a user is logged on. The form that handles login is shown in figure 2.



Figure 2: The login form

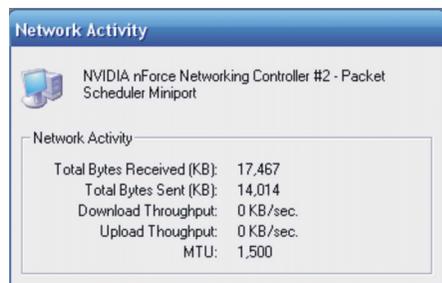


Figure 3: The network activity form

The form shown in figure 3, displays the total bytes sent and received on the network adapter and the download and upload throughput. Figure 4, displays the arp table being used by the LAN. Arp is used to resolve an IP address to a MAC address.

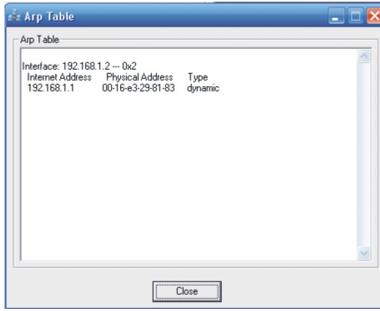


Figure 4: The Arp form



Figure 5: The network address lookup form

The form shown in figure 5 is used to resolve a hostname to an IP address. It takes as input any hostname and produces as output a list of IP addresses corresponding to the hostname.

As shown in figure 6, this form is used to resolve an IP address to a hostname. It takes as input an IP address and gives as output the corresponding hostname.

The network adapter status form shown in figure 7 below retrieves the status information of all the network adapters found in the PC.



Figure 6: The domain name lookup form

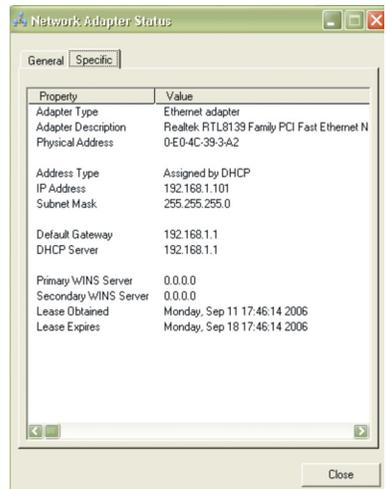


Figure 7: The network adapter status form

The ping form tests for connectivity between a source host and a destination host. Figure 8 is the form used to ping a local or remote host. The form shown in figure 9 is used for tracing the path a packet takes from source host to destination host.

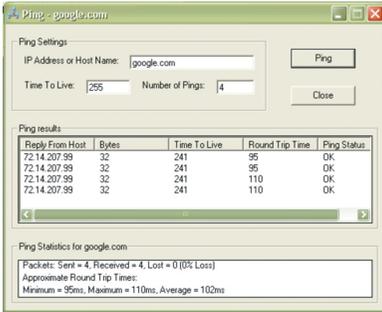


Figure 8: The ping form



Figure 9: The trace route form

The form shown in figure 10 is used for monitoring the network in general, based on source and destination IP address, protocols and based on port numbers. The original source of the form was from the Planet Source Code website (<http://www.Planet-Source-Code.com/vb/>). The general monitoring is based on the network adapter's available IP addresses generated by the code for the form. These addresses can be selected by using the combo box below the form's title bar. The monitoring based on source and destination IP addresses, protocols and port numbers can be done by using filters. The filters can be accessed by using the toolbar below the title bar. The outputs from the form are the inflow of traffic in and out the network adapter and the contents of each packet.

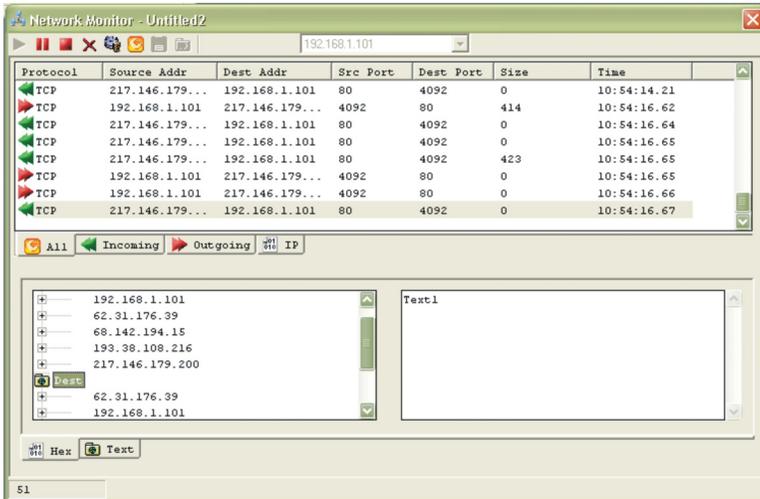


Figure 10: The network monitor form

3. Research Methods

The aim of this research is to implement a range of network monitoring tools within a single GUI under Microsoft Windows. The functionality of the implementation was specified to include live monitoring of network traffic; monitoring based on IP addresses, MAC addresses, protocols and port numbers; network address look up; domain registry look up; traceroutes and wireless network monitoring. The research

methods used in conducting this research include interviews with systems and network administrators to determine the extra features required for inclusion as a tool; review of existing tools and software to identify what is to be improved upon; experimental design for the actual implementation of the network monitoring tools and finally, a survey by using questionnaire to evaluate the tools implemented.

4. User Evaluation

As part of the research method, a survey was carried out using a questionnaire to evaluate the performance of the network monitoring tools implemented. This was done after the testing of the tools. The tools were sent to twenty five targeted users; only 17 participated in the survey by testing the tools and filling in the questionnaire. The analysis carried out on the results from the questionnaire showed that:

Of the 17 respondents:

59% strongly agreed and 41% agreed that the software was simple to use. 71% strongly agreed and 29% agreed that the software was user-friendly. 12% strongly agreed and 88% agreed that the software made good use of a graphic interface. 53% strongly agreed, 35% agreed, 6% do not know and 6% disagreed that the software had no adverse effect on the system while it was running. 65% had no error, 29% had between 1 to 5 errors and 6% had between 6-10 errors. 18% strongly agreed, 76% agreed and 6% do not know that the software recovered from the error(s). 18% strongly agreed and 82% agreed that the software satisfied its aim. 12% strongly agreed and 88% agreed that the software satisfied the objective of having the capability to monitor live network traffic. 35% agreed, 29.5% do not know, 29.5% disagreed and 6% strongly disagreed that the software satisfied the objective of having the capability to monitor a network based on Internet Protocol (IP)/Media Access Control (MAC) addresses, protocols and port numbers. 65% strongly agreed and 35% agreed that the software satisfied the objective of having the capability to perform network address and domain name lookups. 88% strongly agreed, 6% agreed and 6% do not know that the software satisfied the objective of having the capability to determine the connectivity of two network devices. 65% strongly agreed and 35% agreed that the software satisfied the objective of having the capability of finding the routes taken by a packet from source to destination devices. 35% agreed and 65% do not know that the software satisfied the objective of having the capability to monitor a wireless network.

5. Conclusion

Network monitoring tools play a vital role in every network and it is a must have if an organisation is to achieve its ROI. The benefits inherent in using network monitoring tools cannot be over emphasised. It provides that valuable network monitoring information needed for the management of any network. It enhances network stability, reliability, performance and allows for the controlling of the complexities in modern day networks.

Analysis of the survey on the evaluation of the network monitoring tools implemented showed that the research achieved its objectives to an appreciable level.

It is believe that this research will go a long way in creating that awareness on the need for constant monitoring of the network.

6. References

Burke, J. R. (2004) *Network Management: concepts and practice, a hands-on approach*, Prentice Hall, New Jersey, ISBN: 0-13-032950-9

Cheswick, W. R. and Bellovin, S. M. (1994) *Firewalls and Internet Security: Repelling the Wily Hacker*, Addison-Wesley, Reading, ISBN: 0-201-63357-4

ComLab Website (2006) 'Tools for modeling the user-traffic' [Online], Available: <http://www.comlab.uni-rostock.de/research/tools.html> [Accessed August 2006]

Gaglio, S., Gatani, L., Lo Re, G. and Urso, A. (2006) 'A Logical Architecture for Active Network management' *Journal of Network and Systems Management*, vol. 14, No. 1, pp127-146

Held, G. (2000) *Managing TCP/IP Networks: Techniques, tools and security considerations*, Wiley, Chichester, ISBN: 0-471-80003-1

IEPM Website (1999) 'Monitoring with tcpdump' [Online], Available: <http://www-iepm.slac.stanford.edu/monitoring/passive/tcpdump.html> [Accessed August 2006]

Leinward, A. and Fang, K. (1993) *Network Management: a practical perspective*, Addison-Wesley, Reading, Mass, ISBN: 0-201-52771-5

Liska, A. (2003) *The Practice of Network Security: Deployment Strategies for Production environments*, Prentice Hall, New Jersey, ISBN: 0-13-046223-3

Miller, M. A. (1999) *Managing Internetworks with SNMP*, M & T Books, Foster City, ISBN: 0-7645-7518-X

Subramanian, M. (2000) *Network Management: principles and practice*, Addison-Wesley, Reading, Mass, ISBN: 0-201-35742-9

WildPackets Web Site (2006) 'WildPackets- Etherpeek' [Online], Available: <http://www.wildpackets.com/products/etherpeek/overview> [Accessed August 2006]

Windump Website (2006) 'Windump: tcpdump for Windows' [Online], Available: <http://www.winpcap.org/windump/> [Accessed August 2006]

Wisniewski, S. (2003) *Advanced Network Administration*, Prentice Hall, New Jersey, ISBN: 0-13-097048-4

Recording end-users security events: A step towards increasing usability

D.Chatziapostolou and S.M.Furnell

Network Research Group, University of Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

End-user security is nowadays an integral part of our everyday life since modern computer applications frequently dedicate parts of their functionalities to security. As a consequence computer end-users potentially come across with security related events, which may be either system- or user-initiated. However, computer security is often viewed as a difficult and complicated task, which eventually prevents end-users from achieving the protection that they desire and anticipate. This paper presents the results of an initial study from 26 participants, the purpose of which was to investigate the usability of security events that were encountered over a two week period. The results reveal difficulties in dealing with the security events, with more intense problems encountered when end-users attempt to make use of security intentionally.

Keywords

Security, Usability, Security event, End-users

1. Introduction

Nowadays, it is a common observation that end-users are much closer to security than in the past. With the increasing volume of IT threats, end-users more often come into contact with security-related events. Indeed, security functionality is now frequently integrated within software such as operating systems, and tools and applications. For instance, in Windows XP, the integration of security has significantly improved since the introduction of Service Pack 2 in 2004 (Microsoft Corporation, 2006b). However, as a consequence, end-users potentially come across security terminologies such as pop-up blockers, software update messages, and security alerting messages for possibly unsafe attachments. Additionally, use of security-oriented software, such as firewalls, antivirus and antispymware products has significantly increased as the associated threats become more widespread and recognised. Moreover, general applications now often incorporate security functionalities. For example, applications within the Microsoft Office suite employ encryption functionality in order to protect misuse of data (Microsoft Corporation, 2006a).

Unfortunately, the reality of this situation is serious. Users can be deterred if they are not able to understand the security presented to them. In fact, this is often the case as security is frequently not optimally designed for end-users. Software designers often give less attention to usability when designing security within products. Usability of security applications has critical importance because an unusable product might

prevent end-users from enabling security features in their systems. This means that end-users, either at work or home, might be left unprotected. Therefore, usability considerations include ensuring that end-users are able to find the security available for them and determine the protection they require at any time.

This paper presents an investigation into the usability of security and the challenges that end-users face in using the related software features. The first part of the investigation examines prior works on usability and security, with references and examples. The rest of the paper presents the results from a related study, the aim of which was to examine end-users' understanding of security events encountered while making ordinary use of their computers.

2. Examples of unusable security

Examples of security usability problems have been witnessed numerous times by security researchers. In the area of security-oriented tools, a prominent example is Whitten and Tygar's (1999) evaluation of the usability of PGP version 5.0. Their work is one of the first standard examinations of the usability of security applications. Specifically, they had carried out a cognitive walk-through analysis along with a heuristic evaluation, which was completed with user testing. Their findings showed that PGP 5.0 user interface had severe problems which made public key cryptography a difficult task for an average user to accomplish.

Unfortunately, this is not the end of the list. General applications are also found to lack usable security. Internet Explorer (IE), the standard Web browser of Microsoft Windows, has been used to illustrate improper implementation of usability and security. Furnell (2005) indicated that "Users may struggle to make appropriate use of IE's security features". Although IE is a general application rather than a security-specific tool, it includes security functionalities within its options. According to Furnell, the related security options of IE violate key principles of Human Computer Interaction (HCI). These key principles apply for friendly visual state and informative feedback to the end-users. In reality IE seems to have been designed with security features to be primarily meaningful for advanced users, who have prior security knowledge. This lack of usability could possibly reduce end-users' protection rather than encourage its use.

3. The study

With the above points in mind, a study has been conducted in order to investigate end-users' encounters with actual security events. The aim was to record the participants' experiences over a two-week period. A recording sheet was created and distributed to participants for use during this time. In addition to one-off collection of background details about the participants, the sheets sought to record two specific categories of ongoing information, relating to system- and user-initiated events. These two broad categories encompass the types of security event that end-users might experience, as described in the sections that follow.

3.1 System-initiated events

These types of events occur with intention to inform the end-user about security. Thus, that type of events initiates from the computer system and targets the end-user. This could be done in different ways, such as security messages and warning screens, pop ups etc. depending on the computer system, operating system, and the applications installed. We define system-initiated events as: *'Events initiated by a computer system with the intention to advise and inform end-users' operations'*.

For example, many users may be familiar with seeing pop-up dialogs in their web browser asking them whether or not they wish to allow an event, such as that in Figure 1. In such cases the participants make an entry on the recording sheet providing details of the application that initiated the security event. Additionally there is a series of key questions which contribute the investigation of usability. In brief, the included questions concerned whether participants understood the event, if they had to take a decision, if there was a help feature and whether it was used, and if that event prevented participants from completing the task they were trying to perform. From the participants' comments the usability level of an application could be assessed.

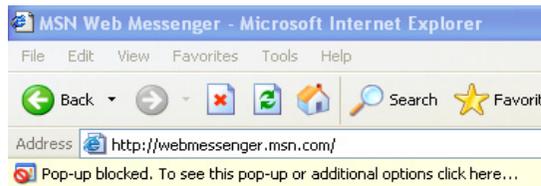


Figure 1: A system-initiated event in the form of pop up message

3.2 User-initiated events

User-initiated events differ from system-initiated events as at this time an end-user initiates an event with intention to deal with security. Specifically, this applies when end-user intends to take control of a computer system by configuring security-related features within applications and tools. We define user-initiated events as: *'Events initiated by an end-user of a computer system who intentionally wishes to utilize security toward distinct goals'*.

As the definition states, these types of events are requests from an end-user who has settled a goal relating to security and wants to accomplish it. An example of a user-initiated event is shown in Figure 2. In this example the application used is Internet Explorer (IE), in which security functions are available under the options tab. As Figure 2 shows, an end-user might intentionally attempt to configure security options, such as whether ActiveX controls, plug-ins, scripts and other security-related operations should be enabled or not. Imagining this was a real case, a participant experiencing this user-initiated event could make an entry in the recording sheet providing information of the application used, the actual intention and whether or not they were able to accomplish the task.

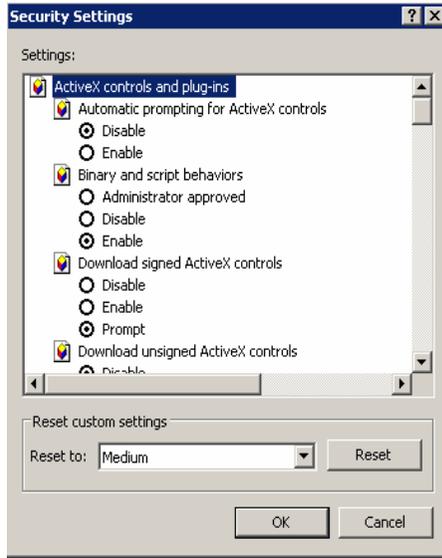


Figure 2: A user-initiated event in Internet Explorer

4. Study results

The total number of participants was 26 people with an equal split between genders. Most of the participants (68%) were in 21-29 group of age, with the rest evenly split between participants under 20, and those aged 30-39 and 40-49. The focus in the age category of 21-29 expected the participants to have a good appreciation in information technology as part of their everyday lives. This is confirmed as 92% of the participants used a computer on a daily basis and 88% rated themselves as ‘intermediate’ or ‘advanced’ users. Moreover, the participants’ level of education is considered high, as 88% claimed to hold a university level qualification.

The results showed a total of 87 recorded system-initiated events. The majority of them were recorded from security-specific applications, which translate to 76% from the total system-initiated events. A tabulation of the applications and tools that initiated the events can be seen in Table 1.

The type of system-initiated events experienced by the participants were primarily in the form of ‘warning messages’ (41%), followed by ‘security alerts’ (38%), ‘update messages’ (15%) and less commonly ‘password requests’ (6%). From the total of 87 events, 82% were fully understandable by the participants, while in the remaining 18% of cases respondents claimed that they were not able to fully understand. This translates to 16 system-initiated events out of 87 that were not fully understood by the participants.

The results also revealed that 66% of the system-initiated events required the participants to take a decision. The participants were asked to specify whether they were clear on what to do as a result, and the replies were as shown in Table 2. The

majority were clearly comfortable, but this still left more than a third of cases in which participants claimed to be confused.

Application / Tool	No. of recorded events	Amount in %
Windows Security Centre	25	30%
Zone Alarm	15	17%
McAfee	14	16%
Norton	11	13%
Internet Explorer	9	10%
Firefox	7	8%
MSN Messenger	2	2%
Safari	2	2%
Word	1	1%
Outlook	1	1%
Total	87	100%

Table 1: Ranked listing of the system-initiated events recorded during the study

Was it clear what to do next?	No. of Times	Amount in %
Totally clear	25	29%
Mostly clear	28	32%
Mostly unclear	20	23%
Not at all clear	14	16%
Total	87	100%

Table 2: Participants' understanding of what to do next at the occurrence of system-initiated events

The results relating to the help and assistance that participants used in the incidents of system-initiated events show that only 11% of the recorded events involved use of a 'help' feature, as shown in Table 3. Meanwhile, in 48% of the total system-initiated events a 'help' feature was not used, whereas in the remaining 41% the participants recorded that there was no help available. In terms of other guidance, the answer was 'no' in 92% of cases, while in 6% participants referred to the Internet, and to other people for the remaining 2%.

In response to the final question, participants were asked if the system-initiated event prevented them from completing a task they were trying to perform at the time. Again, while the majority (78%) were not prevented, it is notable that in 22% of cases the user was effectively defeated.

Did you use a help feature?	No. of Times	Amount in %
Yes	10	11%
No	41	48%
N/A	36	41%
Total	87	100%

Table 3: Usage of a 'help' feature from the total system-initiated events

The results concerning the user-initiated events recorded a total of 29 events, which is a significant drop when compared with the system-initiated category. As with the system-initiated events, the applications and tools that were primarily recorded were

security-oriented, accounting for 66% of the total user-initiated events. Table 4 represents in detail the applications/tools and their related occurrence in user-initiated events.

Application / Tool	No. of recorded events	Amount in %
McAfee	7	25%
Norton	5	17%
Zone Alarm	4	14%
Windows Security Centre	3	10%
Router security configuration	3	10%
Back up	2	7%
Firefox	2	7%
MS Word	2	7%
Internet Explorer	1	3%
Total	29	100%

Table 4: Ranked listing of the user-initiated events recorded during the study

In most of the cases (59%) the participants were asked to take a decision at the time they initiated an event. This situation demands good understanding of the event by the participant in order to correctly take decisions. In fact the study results revealed that in more than half (15) of the total recorded user-initiated events the participants did not have a clear view when asked what to do next in the event as shown in Table 5.

At this point, considering the fact that more than the half of the user-initiated events claimed 'not clear of what to do next', the presence of a 'help' feature is considered imperative. In reality the study results revealed that for 58% of the user-initiated events recorded by the participants that there was no help available, as shown in Table 6.

How clear was it to do what you had to do?	No. of Times	Amount in %
Totally clear	10	34%
Mostly clear	4	14%
Mostly unclear	6	21%
Not clear at all	9	31%
Total	29	100%

Table 5: Participants' understanding of how to perform user-initiated events

Did you use any help feature?	No. of Times	Amount in %
Yes	4	14%
No	8	28%
N/A	17	58%
Total	29	100%

Table 6: Usage of a 'help' feature from the total user-initiated events

The absence of a help feature reduces the usability which is one of the main considerations in HCI. In the remaining 14% of user-initiated events (i.e. four instances) the participants actually made use of an available 'help' feature. This result indicated that end-users did not often use a 'help' feature, considering that the

option was present twelve times, and eight times the participants did not use it. Additionally, in the majority of the cases (23 times) no other guidance was drawn upon. Only six times did the participants look after for some additional help, with four cases on the Internet, and in two cases they turned to other people.

The last question asked if participants were able to complete their intended action. Even though in the majority of the cases (66%) they managed to accomplish their tasks, there were ten user-initiated events (34% of the total cases), in which participants did not manage to complete their task. This certainly suggests problems in terms of the clarity and usability of the provided security, and represents an area for further attention.

5. Discussion

The participants' feedback was analysed in order to investigate the usability of security in applications that end-users normally use. The main objective was to indicate if they are capable to deal with them. The study results relating to the system-initiated show that 18% of the total events were not fully understandable. This is much more intense when considering the user-initiated events, since the survey results revealed that ten out of the 29 events were not able to be completed. If these findings are representative of wider user experiences, then they certainly highlight a significant problem. Moreover the infrequency of user-initiated events in the study suggests that many end-users do not actually use security intentionally, and instead rely upon the default features of their applications. Furthermore, some recorded incidents indicated that when participants attempted to accomplish an advanced task, such as setting firewall rules, they failed and ended up frustrated. Some participants underlined the fact that there was no appropriate help, which made their tasks even more difficult. Additionally, plenty of times there was no help available, which made participants simply give up. Participants eventually spend time and effort without any outcome. This has as a consequence that end-users probably avoid security related tasks in the future.

6. Conclusion

This paper highlighted some real incidents that end-users are facing when they come across security events. The results from the survey indicate the importance of the situation. End-users barely make intentional use of security. This abstention has as a consequence that end users are not able to have full usage of the available security. Ideally, they should derive confidence to use security by having complete control over security events in order to fulfil their tasks. Security functionality within applications has been seen to demand experience and knowledge from end-users. This leads to an immediate discrimination between users: on one side some users are able to protect themselves, whereas on the other side are users that simply cannot do so. It is very difficult for an end-user with limited computer literacy and experience to be able to use the available security features.

In terms of future research, it is recognised that the user population involved in this initial study was relatively small. As such, it would be desirable to undertake a

wider exercise involving more participants, with a wider range of backgrounds. In addition, it would be beneficial for such a future study to further simplify the task of recording events, so as to prevent participants from neglecting to do so.

7. References

Furnell, S. M. (2005), 'Computers and Security: Why Users Cannot Use Security', *Computers and Security* 24(4), 275–279

Microsoft Corporation Web Site (2006a), "Security", Available at: <http://office.microsoft.com/en-us/assistance/ha011403111033.aspx>, (Accessed on 20 August 2006)

Microsoft Corporation Web Site (2006b), "*Windows XP Service Pack 2 Overview White Paper*", Available at: <http://msdn.microsoft.com/security/productinfo/xpsp2/default.aspx>, (Accessed on 21 August 2006)

Whitten, A. and Tygar, J. D. (1999), 'Why Johnny Can not Encrypt: A Usability Evaluation of PGP 5.0.', in *Proceedings of the 8th USENIX Security Symposium* pp. 23–26. 56

An investigation on the relationship of aggregated TCP and UDP traffic

M.Davy and B.V.Ghita

Network Research Group, University of Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

The internet quality of service and reliability is currently managed by the TCP protocol. TCP is responsible for setting up the streams speed in a managed way, and if only a limited bandwidth is available, for slowing down the transmission rate of the streams. However, the internet is changing, no longer the dominant applications are HTTP, FTP or POP, but multimedia streaming. As the far most of them are based on the UDP protocol, which doesn't the control features of TCP, the streams oppose a threat to TCP transmissions which are no longer able to manage the quality of service. This paper investigates the relationship between TCP and UDP traffic, by using a network simulation approach and the analysis of real network trace data to set the fundamentals for a more extensive exploration of the aggregated traffic issues and some solutions that could be developed. This research has been sponsored by France Telecom R&D UK.

Keywords

Aggregated traffic, TCP, UDP, network simulation, trace analysis.

1. Introduction

Since the early development of networks and Internet, most applications have been running on top of TCP due to the quality of service (QoS) it provides. The recent evolution of networking applications towards the delivery of more multimedia content and their ability to use high data rates to deliver content with an in-time delivery preference rather than in-order delivery pushes upwards the usage of UDP as a transport protocol instead of TCP. But lacking the feedback and adaptation TCP implements, UDP tends to use all the bandwidth available on a transmission path, eventually congesting the network and preventing TCP transfers to live normally.

This paper presents the work accomplished in order to investigate the fact itself that a relationship between UDP and TCP transfers exists, and if the answer is affirmative, under which conditions can this effect be demonstrated, and is organised as follows. Section 2 will present an overview of the related research in this domain, trying to give some solutions to the problem of aggregated traffic using adequate queue management, developing new application layer protocols or, more drastically, developing new transport solutions that would be deployed alongside TCP and UDP. Section 3 will describe the network simulation developed to investigate in a controlled environment the apparition of the TCP-UDP aggregate traffic problem and Section 4 will present the methodology and results from the analysis of real traffic trace data to detect if such a problem can be observed on different live networks.

2. Related work

Due to the rising data rates employed by multimedia applications, network nodes responsible for the routing of the traffic like routers have to be able to forward packets faster and better, possibly adapting their queues management has to adapt also to this evolution, the standard management scheme operated on a Round Robin (RR) fashion becoming more difficult to meet, resulting in overgrowing queues. Evolutions of the RR queue management like the Weighted Round Robin algorithm (WRR), attributing a weight to each queue, has been tested in (F de Castro et al 2003) and present better solution, but are not adaptive schemes as they do not depend on the traffic passing through the router interfaces but on the interfaces themselves. The evolutions of active queue management with Random Early Detection (RED) (Floyd and Jacobson 1993), Flow based RED (FRED) (Lin and Morris 1997) allow a better packet dropping scheme more fair towards TCP transfers., as they are based on an algorithm trying to predict that a queue would overgrow and start to drop packets from it before the congestions starts. While RED monitors the queues lengths and discards randomly packets from a queue, it is not able to restrict unresponsive flows such as UDP. FRED however implements preferences in the way it discards packets, trying to drop more packets from unresponsive flows. The implementation of RED-DT is a trial to improve this behaviour by adapting the dropping probability of each queue upon the arrival of each packet (Vukadinović and Trajković 2004).

The easiest way to implement fairness in aggregate transport would probably be to evaluate the impairments TCP transmissions are confronted to while transmitting UDP streams. Several researches in such a direction have been started and a few protocols have been developed for specific purposes. SABUL and later redefined as UDT (Gu *et al* 2003) has been developed on a rate based congestion control basis for high data rates traffic, using UDP to transfer data and TCP feedback messages to provide the application with information to change the UDP sending rate. A similar idea has been developed with TCP Rate Probing Adaptation (TPBA) (Tobe 1999), switching from UDP to TCP in the same transmission to regularly probing the network status while transmitting to adapt the UDP sending rate accordingly.

An alternative, though much heavier would be to develop a new transport protocol. Instead of relying on the existing transport architecture, some researches have been focusing on creating side protocols to TCP and UDP. Among them are the Reliable UDP (RUDP) based on RDP (Partridge and Hinden 1990) and the Datagram Congestion Control Protocol (DCCP) (Kohler *et al* 2006). DCCP implements an unreliable flow of datagrams with acknowledgements, a connection handshake and teardown. The most important feature of DCCP is its evolution capacities, as its congestion control and special features can be changed with the usage of different Congestion Control IDs (CCID). It currently implements two CCIDs, which respectively implements a TCP-like Additive Increase Multiplicative Decrease (AIMD) congestion control (Jacobson 1988), and a TCP friendly rate control (Floyd and Kohler 2006; Floyd *et al* 2006).

3. Network Simulation

In this section we present the simulation results obtained with OPNET MODELER. The simulation has been carried in 4 rounds, implementing two different network topologies under diverse TCP and UDP loads. The simulations started with simple topologies constituted of two single hosts linked to a switch and a server by 10 Base-T or 100 Base-T links, followed by a more complex topology linking sub-networks to a switch and the server by 10 Base-T links. Figures 1 and 2 present the general topologies used.

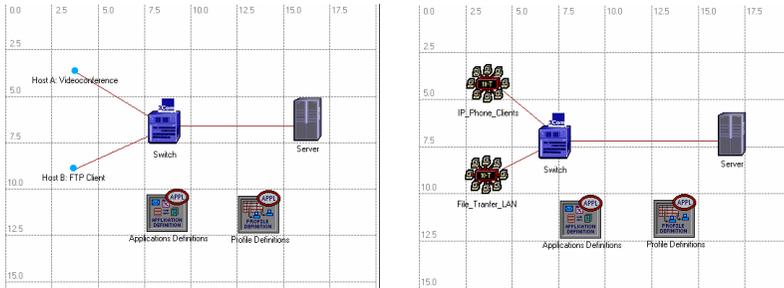


Figure 1: OPNET Topology #1 (left) and topology #2 (right)

The load imposed on the network comprises for all the simulated scenarios a constant FTP transfer as a source of TCP traffic and the UDP traffic was either generated with videoconference or Voice over IP (VoIP) clients following the definitions presented in Table III-1. The projects #1 and #2 are simulated using the topology #1 and the projects #3 and #4 use the topology #2. Project #2 is divided into two sub-projects as at this time in the research, it was easier to change the maximum speed of the links than to modify the default parameters of the applications defined by OPNET to obtain a different scenario exhibiting a load inferior to the bottleneck bandwidth of the network. Project #2-1 uses 10 Base-T links while project #2-2 uses 100 Base-T links.

OPNET Project	Bottleneck bandwidth	TCP Traffic					UDP Traffic				
		Hosts	Applic ^o	Traffic KB/s	Start	Durat ^o	Hosts	Applic ^o	Traffic	Start	Durat ^o
# 1	10 MB/s	1	FTP	250	0s	300s	1	Video-conference	7.6 MB/s	Every 20s	20s
# 2 - 1	10 MB/s	1	FTP	250	0s	900s	1	Video-conference	7.6 MB/s	20s	60s
# 2 - 2	100 MB/s	1	FTP	250	0s	900s	1	Video-conference	7.6 MB/s	20s	60s
# 3	10 MB/s	10	FTP	575	0s	900s	10	VoIP	16 KB/s	100s	300s
# 4	10 MB/s	10	FTP	575	0s	900s	75	VoIP	990 KB/s	100s	300s

The results exhibited by the simulations vary from one project to the other, but the projects #1, #2 – 1 and #4 exhibited some interaction from UDP transfers on the FTP traffic, whereas the projects #2 – 2 and #3 didn't. With the simple topology used for the project #1, it was already possible to observe the problem of the aggregated UDP – TCP transport as exposed by the traffic patterns in figure 3.

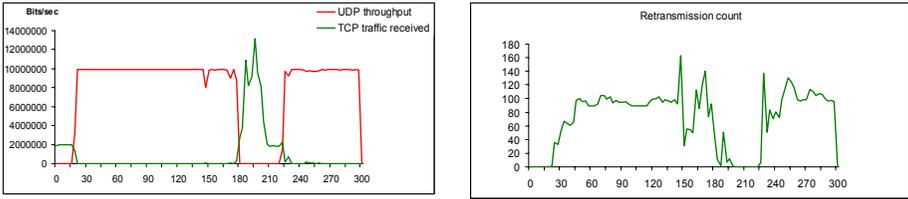


Figure 2: Project #1 traffic patterns (left) and retransmissions (right)

In this simulation round, the TCP traffic is established at 2Mb/s (or 250 KB/s) until when the UDP traffic generated by the videoconference client is triggered at $t=20s$. From this moment until the videoconference stops at $t=180s$, as well as during the second UDP burst after $t=220s$, TCP transfers stop completely and are the subject of a high level of retransmissions as figure 4 shows.

It is important to notice that during this project #1, the bottleneck link is filled up to its maximum bandwidth of 10 Mb/s by the UDP traffic. Similar results have been observed in projects # 2 – 1 and #4, this later one giving a better idea of what can happen on a real network. In fact, the evolution of the project scenarios was elaborated in order to evaluate the possible impact of UDP on TCP starting from small topologies easily encountered on home networks, up to larger topologies closer to a small company network comprising more hosts. Figure 5 present the comparative TCP / UDP traffic of project #4 where it is possible to follow the evolution of the TCP traffic regarding the VoIP calls emitted between $t=100s$ and $t=400s$.

Using the same topology, but with 10 VoIP hosts instead of 75, hence generating UDP traffic figures 7.5 times smaller than the 990KB/s generated during the Project #4, the FTP traffic going through the aggregation link of project #3 fails to present any influence of UDP on TCP, as shown by figure 6.

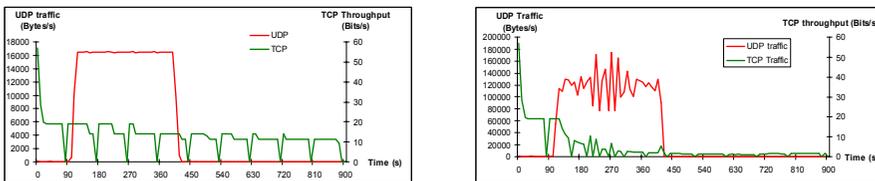


Figure 3: Project #3 (left) and project #4 (right) TCP / UDP traffic patterns

For both these simulation projects, the point-to-point throughput from the server to the aggregating node over the bottleneck link, were used close, or up to the bottleneck bandwidth during UDP transfers. We believe the interaction phenomenon observed between UDP and TCP traffics is due to the fact that for a given network, if the traffic generated by the hosts is not close or greater than the bottleneck bandwidth, the UDP traffic has no influence on the TCP transfers. Additionally, if this UDP traffic tends to be greater than the bottleneck bandwidth, not only the TCP traffic will be affected and will totally back off, but also the UDP transmission themselves will suffer. It can be observed in figure 7 that for the simulated scenarios presenting an influence of UDP over TCP, the overall delay for the real-

time application responsible for generating the UDP traffic was raising dramatically during the duration of the videoconferences or the VoIP calls sessions. For the remaining simulation projects, when no interaction between UDP and TCP was observed, the packet to packet delay during the UDP transmissions didn't vary in the same proportions, if any.

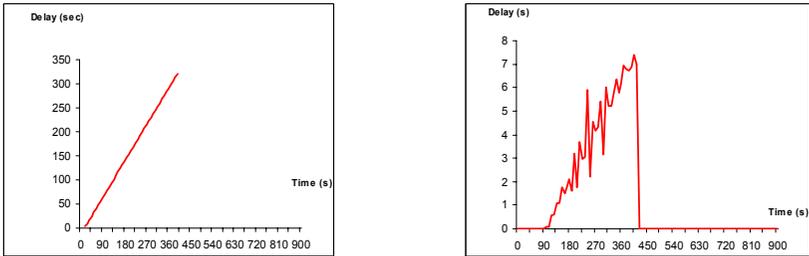


Figure 4: Project #2 Videoconference and Project #4 VoIP packet-to-packet delay

With the simulation of network scenarios implementing different application behaviours, it has been demonstrated in this section that for both small networks and larger topologies likely to be found in real life, the UDP traffic generated by multimedia applications, such as videoconference or VoIP, oppose a threat to the TCP traffic in the case the aggregated bandwidth required to carry both of them is close or greater than the bottleneck bandwidth of the network. Section IV presents the analysis of real traffic trace data aiming to match such behaviour with real traffic.

4. Real Traffic Trace Analysis

In this section, we setup an analysis methodology to monitor the evolution of TCP transmissions in regard to the UDP concurrent traffic present on the same links. While the simulations described in section III propose that the apparition of TCP back off due to unresponsive UDP transfers is triggered by a bandwidth usage close to the bottleneck bandwidth of the network, we investigate here the TCP behaviour in two real traffic data sets.

4.1 Methodology

The analysis of real traffic trace data conducted during this research used off line analysis of trace data captured at aggregation nodes such as a router or a firewall. The methodology required the deployment of a complete process to extract and compute the important parameters from the trace files, including the usage of special tools to protect the identity and the security of the network on which the trace data was captured.

De facto, after collecting trace data from a network node with *Tcpdump*, these privacy requirements imply the usage of a program such as *Tcpurify* to scramble the IP addresses found in the collected data and it has been proposed that due to the complexity of some networks and their range of IP addresses, it was better to scramble not only some of the subnets presented by the trace file, but all the IP

addresses. A special encoding scheme has been added to *Tcpurify* in order to renumber all the trace file IP addresses to 0.0.0.0. Doing so might, at first sight, remove the possibility to detect individual flows in the data collected but as the new encoding scheme doesn't scramble the port numbers this identification could still be carried on. In fact, as presented by (Gleason 2001), the random port numbers used by TCP connections to initiate their requests are unlikely to collide during the average duration of TCP connections, leaving the possibility to identify uniquely different TCP connections and their parameters.

The sanitised trace data obtained from the modified version of *Tcpurify* has been further processed by a series of scripts responsible for the generation of the trace statistics, comprising the evolution of the TCP throughput, retransmission count and reported losses. However, due to some software limitations, namely *Tcptrace*, responsible for generating TCP and UDP connections reports, it has not been possible to generate statistics reports with an update period less than a second. In fact, if the generation of the TCP connections timestamps is accurate up to the millisecond, as *Tcptrace* first focus is not UDP connections, the generation of the UDP parameters reports is subject to much less attentions, leaving the resolution of the timestamp to the interpretation of textual dates instead of the precise UNIX epoch format. This limitation induced a larger amount of work in order to be able to synchronise the TCP and UDP flows datasets extracted from the trace files.

The synchronisation of the two protocols flows information is the critical step in the trace analysis. In fact, as the TCP flows and UDP flows are used by different applications, may have different sources and destinations and last for different amount of time, it is not possible to directly compare both protocols by associating a UDP flow to a TCP connection. Instead, an averaging process has been set up at regular intervals in order to compare the evolution of the previously mentioned TCP parameters regarding the overall UDP throughput encountered in the path. As defined by the lack of capability of *Tcptrace* to present UDP timestamps in a precise fashion, this interval has been reduced to the minimum available after the *Tcptrace* reports generation, e.g. 1 s. The process of synchronisation of the TCP and UDP flows information is considered critical because if not done properly, some important comparison points might be lost by comparing parameters from different times.

4.2 Trace analysis

The data used during this research came from A. a small network constituted of a limited number of hosts at the France Telecom (FT) laboratory, with a bottleneck bandwidth in direction of Internet of. 2Mb/s and B. a large network comprising hundreds of hosts at the University of Plymouth (UoP) using four 150Mb/s connections to connect its backbone its ISP.

To give credit to the methodology presented in section IV. A., it is important to know that this analysis is relevant only if the levels of TCP and UDP usage on the network are sufficient to assume that, would an evolution of TCP parameters be detected, it could be attributed to UDP transmissions and only them. In fact, if other protocols share the same amount of bandwidth than TCP or UDP transfers, it would not be relevant to compare the evolution of those ones only, but would require the

integration of the other protocols too. Fortunately, the distribution of the protocols usage shown in figure 8 and figure 9 is much in favour of TCP and UDP, whose aggregate usage represents at least 95.85% of the traffic encountered at the FT lab and 96.45% for the UoP dataset, allowing us to make the assumption that would an interaction be discovered between UDP and TCP, it would not be dependant on other protocols behaviours.

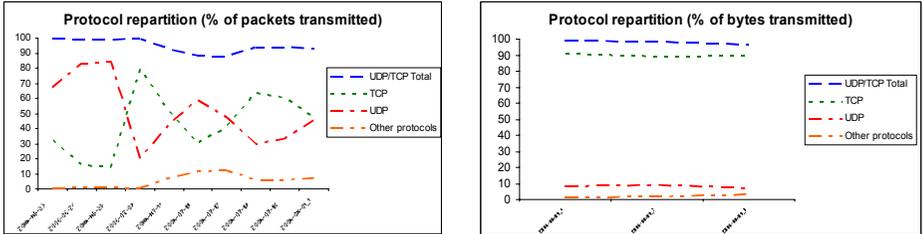


Figure 5: FT (left) and UoP (right) datasets protocol usage distribution

Due to the self adjusting behaviour of TCP, the evolution of TCP throughput in response to a high UDP level on the network would be expected to decrease as the UDP throughput raises. Unfortunately, the figures presented by both datasets for directions incoming to and outgoing from the capture node do not present such a pattern. Figures 10 and 11 present the repartition of the TCP throughput in regard to the UDP throughput, both averaged every second of the capture file as described in section IV. A. It might be noticed that for the FT dataset, the incoming traffic seems to present a reduced TCP throughput for higher UDP throughput values, but we believe this can be attributed simply to a lack of high throughput UDP transmissions rather than a direct influence of UDP on TCP connections.

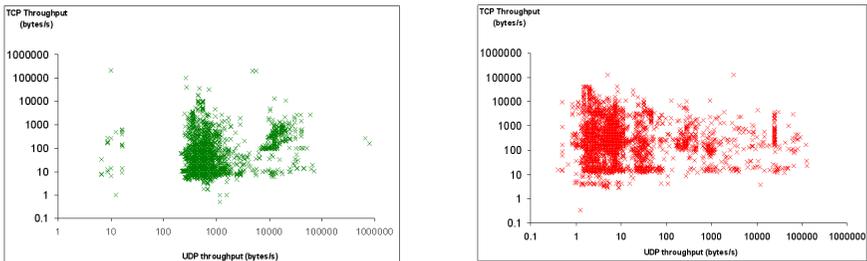


Figure 6: FT traces – Throughput distribution, outgoing (left) and incoming (right) traffic.

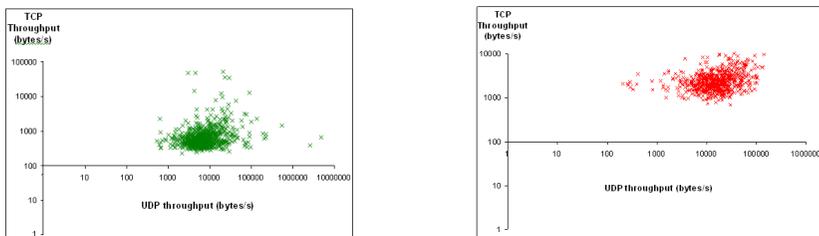


Figure 7: UoP traces – Throughput distribution, outgoing (left) and incoming (right) traffic.

For the UoP dataset, the values obtained appear to be more stable. This is probably due to the higher number of hosts generating some traffic compared to the FT lab, where a single host can have greater influence on the overall throughput figures. The UoP traces even present in the case of the incoming traffic a rising trend towards the higher UDP throughput values, which would be the opposite effect than the expected diminution. However, as indicated previously, the bottleneck bandwidth of the University of Plymouth backbone is much greater than the network load shown in the trace data captured. The rising pattern presented in figure 11 may only be the normal evolution of the traffic for such slow bandwidths. This might come as an element to confirm that for lower levels of usage of the bottleneck link capacity, no interaction of UDP on TCP transfers is observable.

Retransmissions are triggered when packets are not acknowledged, when the TCP retransmission timer expires, or when some packets are missing in the received sequence. The results obtained with both data sets present different patterns. For the FT traces, as for the throughput, it appears that many TCP connections present a higher amount of retransmissions when the UDP traffic is minimal. And despite the fact that 96% of the time, the loss rate for the outgoing traffic remains under 1% and under 0.5% for 90% of the time for the incoming transmissions, no tendency to increase towards the higher UDP throughput has been detected.

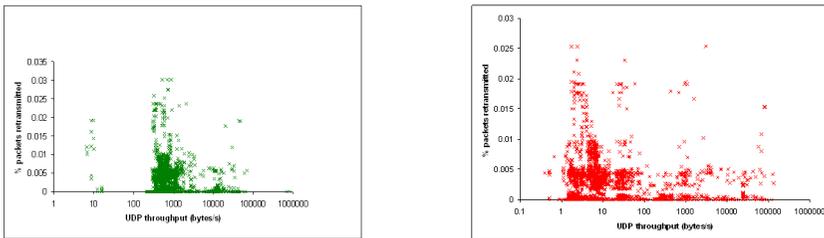


Figure 8: FT traces – TCP Retransmission rate vs. UDP throughput outgoing (left) and incoming (right) traffic.

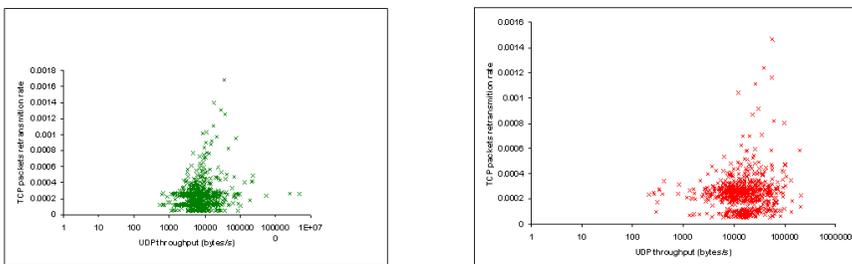


Figure 9: UoP traces – TCP Retransmission rate vs. UDP throughput outgoing (left) and incoming (right) traffic.

However, for the UoP traces, it seems that, for some TCP flows transmitting when UDP throughput is higher than during the rest of the trace data, more retransmissions can be observed in both directions. This may constitute an example in favour of the expected behaviour for the aggregate traffic, but these values represent only 5% of the dataset, 82% of the entries presenting a retransmission rate under 0.03%, even for

the higher UDP bandwidth observed in the trace data, and its impossible to conclude from them.

As for the retransmissions, the expected behaviour of packet loss was an increase as UDP throughput raises. Losses are here reported from Tcptrace statistics which indicate them as post loss acknowledgements (ACKs), e.g. the total number of ACK packets received after losses were detected and a recovered from. If the losses distribution for the FT dataset presents a greater amount of losses when the UDP throughput is lower, the UoP traces show an opposite distribution. The general figures for the outstanding data are again quite low as only 0.4% of the outgoing traffic in the FT dataset are greater than 0.1% and 4% of the incoming traffic present more than 0.5% of losses. For the UoP dataset, 7.6% of the outgoing traffic presents a loss rate inferior to 0.3% while 2% only of the incoming traffic loss is greater than 0.04%.

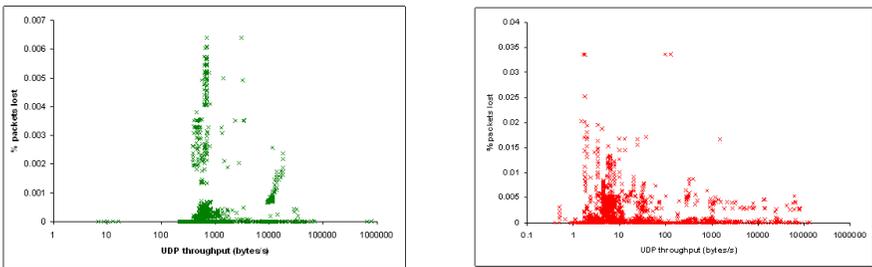


Figure 10: FT traces – TCP Loss rate vs. UDP throughput outgoing (left) and incoming (right) traffic.

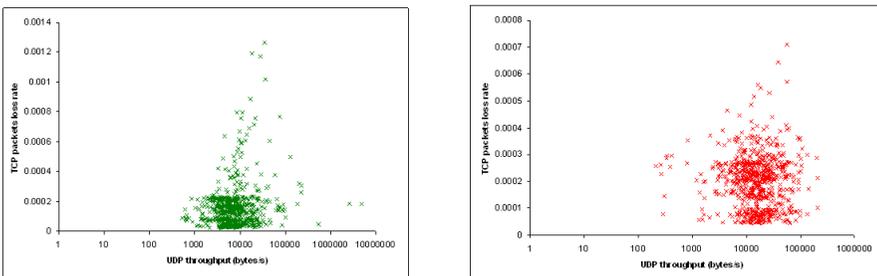


Figure 11: UoP traces – TCP packet loss rate vs. UDP throughput outgoing (left) and incoming (right) traffic.

The really low loss figures presented the UoP trace data, associated to the averaging methodology used to obtain those figures, do not allow us to emit any conclusion in favour of the UDP influence on TCP transmissions, and further investigation, especially for higher values of UDP throughput, need to be conducted.

5. Summary, Limitations and Future Work

In this paper, we have presented the research conducted both with simulation scenarios and the analysis of real trace data in order to detect in which conditions an influence of UDP unresponsive transfers might impact on the TCP traffic. The

simulation scenarios presented that, for a given network, if the aggregated traffic throughput is not close enough or does not exceed the bottleneck capacity of the network, no interference can be detected and both types of traffics can flow without suffering any impairments. However, it has been proposed that in opposite scenarios, TCP connections would start to back off, and eventually starve. In extreme cases, not only TCP transfers would stop, but a dramatically increase of packet to packet delay would be observed in UDP transmissions. However, if this phenomenon has been observed during the simulation rounds, the analysis of two real traffic trace datasets didn't provide any explicit demonstration of this phenomenon. These results have to be tempered by the methodology used for the analysis: in order to obtain data suitable for a cross analysis between the two protocols parameters, an averaging process has been developed, which might hide important information about the connections. As the figures obtained from the trace files used did not present an extensive usage of the network, to validate the hypothesis that effects of UDP can be observed on TCP only if the link usage gets close to the bottleneck bandwidth, the analysis methodology would gain to be revised, possibly implementing live analysis instead of off-line analysis, allowing longer runs and more accurate information to be extracted.

6. References

F. de Castro M., Gaíti D., M'hamed A., Oliveira M., "Comparing Application Performance on Distinct IP Packet Scheduling Configurations", http://www.cefetce.br/Ensino/Professores/mauro/Public/sBC-SBRC_Comparing%20Application_Performance_on_Distinct_IP_Packet_Scheduling_Configurations.pdf, 2003

Floyd S., Jacobson V., "Random Early Detection Gateways for Congestion Avoidance", IEEE/ACM Transactions on Networking, 1993

Floyd S., Kohler E., "Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 2: TCP-like Congestion Control", RFC 4341, March 2006

Floyd S., Kohler E., Padhye J., "Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 3: TCP-Friendly Rate Control (TFRC)", RFC 4342, March 2006

Gleason M., "The ephemeral port range", NCFTP website, http://ncftp.com/ncftpd/doc/misc/ephemeral_ports.html, 2001

Gu Y., Hong X., Mazzucco M., Grossman R., "SABUL: A High Performance Data Transfer Protocol", submitted for publication, <http://www.dataspaceweb.net/papers/sabul-hpdt-03.pdf>, 2003

Jacobson V., "Congestion avoidance and control", Proceedings of SIGCOMM '88, ACM, Stanford, CA, August 1988

Kohler E., Handley M., Floyd S., Datagram Congestion Control Protocol (DCCP), RFC 4340, March 2006

Ling D., Morris R., "Dynamics of Random Early Detection", Proceedings of SIGCOMM '97, 1997

Partridge C., Hinden R., "Version 2 of the Reliable Data Protocol (RDP)", RFC 1151, April 1990

Tobe, Y., "A Host Architecture of QoS Control for Continuous Media Stream Communications", PhD thesis, <http://www.unl.im.dendai.ac.jp/~yoshito/yoshito-thesis.pdf>, 1999

Vukadinović V., Trajković L., "RED with dynamic thresholds for improved fairness", Proceedings of the 2004 ACM symposium on Applied computing, March 2004

User Awareness of Biometrics

B.J.Edmonds and S.M.Furnell

Network Research Group, University of Plymouth, Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

Biometric technologies are slowly becoming more commonplace although their growth has not been as fast as some have predicted. There is a stigma attached to biometrics in that people have concerns over their usage. It may be that they fear what people may do with their personal biometric data or it may be that they do not like the intrusive nature of the devices.

A user-trial was proposed to investigate user awareness of biometrics, in order to determine what the general public know about biometrics and their use. Five biometric devices were chosen with regards to public availability and cost: Fingerprint recognition, Iris recognition, Facial recognition, Signature recognition and Voice/Speech recognition. Thirty participants were asked to answer questions before and after using the devices in order to gain an opinion of the technology. Fingerprint recognition was found to be the most favoured of the technologies, whereas Voice/Speech and Signature were the least liked.

Keywords

Biometrics, User awareness, Iris, Fingerprint, Facial, Signature, Voice Recognition

1. Introduction

The sales of biometric technologies are steadily increasing, as a result of people becoming more aware of security concerns (Lake, 2001). People are beginning to see the limitations of more conventional methods of security such as Personal Identification Numbers and Passwords, which rely upon people being able to remember them. This of course can be quite restricting when on average people have to remember two PIN numbers and up to another eight passwords and user names to go with this. It was also found that one in four of the people asked in a recently conducted survey have twice as many as that again (Scotsman News, 2005). This of course results in people using the same password for many systems, which weakens the value of the protection.

This research has been conducted in order to gauge public perception with regards to biometrics. The study involved exposing participants to the use of biometric devices and asking for their opinions based on a variety of questions before and after this exposure.

An underlying intention of the research was to look for the influence that media presentation and other peoples' views may have upon general public opinion regarding biometrics, and to see if actual use of the devices changed this opinion.

2. What are Biometrics?

Biometrics use who you are to identify you. This obviously saves having to carry a key or remember a password. Biometrics can utilise either physiological or behavioural characteristics. Suitable physiological characteristics can include Fingerprint, Hand Geometry, Iris Scanning, Retinal Scanning, Face, and Facial Thermogram. Alternatively, behavioural biometrics can be: Voiceprint Recognition, Signature Recognition, Keystroke Analysis, and Mouse Dynamics.

Biometric devices are becoming more common as people are beginning to realise that they are one of the most secure and user-friendly methods of securing devices. Biometrics can of course be implemented in most of the places that traditional password systems have been implemented. For example on laptops, where a password system has traditionally been implemented, a fingerprint scanner could be used instead. A door that would usually use a swipe card entry system could use an Iris scanning system.

There are of course many advantages to using biometrics: The main benefit is the potential for added security. It has been estimated that the chances of two people having the same iris pattern is 1 in 10 to the 78th power (CNN News, 2004). This is of course very good odds to support the use of biometrics, given that the population of the world is only 65 to the 10th power (World Population, 2003). Thus an identical match is in reality never going to happen. Biometrics are also very hard to forge, they are not like swipe card or password that can be stolen from you. If implemented appropriately, the user actually has to be there in person for the biometric device to work.

3. Questionnaire

In order to gauge the public's views it was decided that a user-trial be conducted, involving the use of a questionnaire. This was decided the best method to gauge the public's perceptions as it is an accurate measure of what the participant is feeling.

The questionnaire consisted of 28 questions for 30 participants. Following a number of questions about the participant's background, there were a series of questions about biometrics. The participants were then invited to use the biometric devices, answering set questions after each. This allowed for a good gauge of what the participants thought of the devices before and after using them, and also what they believed were good about them.

For the hands-on part of the study it was decided that users would simply enrol with the device then attempt to log in to the system. This was considered to allow the participants to get a realistic view of how these devices could be used in a security application without requiring their prolonged participation. The duration of the enrolment and log in process varied for the different applications, but the participants were given a few minutes with each application to familiarise themselves with the device.

A total of five biometric techniques were involved in the trial, three of which were physiological (face, fingerprint, and iris) and two behavioural (signature and voice). These were selected on the basis of being techniques that were all commercially available for use on end-user systems at a reasonable unit cost.

3.1 Iris Recognition

An Iris in a person is completely unique to themselves, even in identical twins. The iris is one of the best biometric solutions. There are many advantages of Iris recognition over other biometric technologies ranging from: Speed, Stability and accuracy.

The hardware was chosen with regards to availability and also how feasible it was to actually use the hardware for the survey. For the Iris recognition the hardware used was the Panasonic authenticam this hardware was used in conjunction with SecureSuite made by I/O software Inc.

3.2 Facial Recognition

Facial recognition is one of the newer biometric technologies due to its complex nature. The face is an important part of who people are and we as humans use it to identify people from one another. It thus seems to make sense that Facial recognition be used as a biometric technology. The human face has around 80 nodal points that are used by biometric software to authenticate people. Only around 14-22 of these are used for facial recognition. The Nodal points that are recorded are made into a string of numbers represent the face, which are then stored in a database.

For the study, a program called FaceIt by Visionics be used. This program can be used with any camera to authenticate a face, so the authenticam was used again for this application.

3.3 Fingerprint recognition

Fingerprints are unique to each person this is due to them being influenced by the environment around them. The ridges on the fingerprint are formed during the foetal stage of life when the general shape is defined. These ridges remain the same throughout life, enlarging as the person reaches adult size. Fingerprints can reconstruct so long as the injury to them is not too severe.

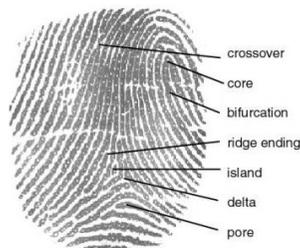


Figure 1: Minutiae of the fingerprint (Biometric Information, 2006)

As Figure 1 illustrates, there are a number of areas involved in a fingerprint although the complete range of characteristics is not always used in recognition, depending on the fingerprint device in use. Collectively the areas of a fingerprint are called Typica (Forensics Information, 2006).

The fingerprint recognition device used was made by Targus and was of the more expensive capacitive type. Again the software used was supplied with the device, and in this case was a security suite called OmniPass.

3.4 Signature

Signature recognition started life in a similar way to that of fingerprint recognition in that we have been signing to verify things for a long time in other contexts. As such, it only seems natural that it be used for biometric authentication.

Biometric signature recognition software does not treat the signature as just an image, but can compare a range of factors. Various signature dynamics (such as speed, relative speed, stroke order, stroke count and the pressure applied) are analysed. So the signature is not only being compared on how it looks but also by how it was generated (PDA Lok Company, 2006).

For the signature recognition a PDA was required due to touch screen. A readily available program called PDA Lok was downloaded as a trial version. Initially the program was tested on an MDA compact device, which is a smaller PDA. Writing a signature on such a small screen proved to be a problem so a Dell Axim was used instead.

3.5 Voice Verification/recognition

Voice verification works by digitising a profile of a person's speech to produce a voice print stored model, or template.

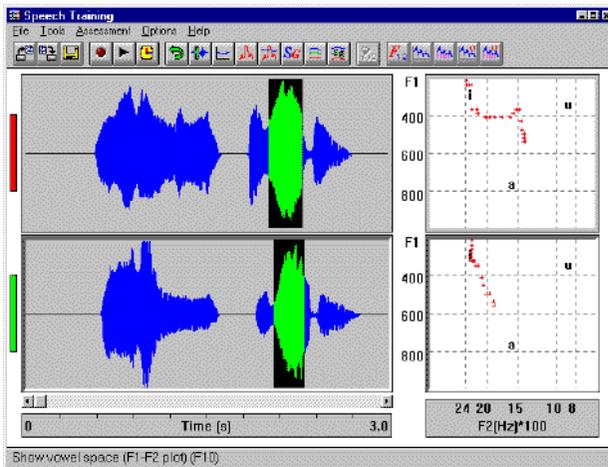


Figure 2: Speech Pattern (Speech Technologies, 2006)

Figure 2 shows two speech patterns from the same person saying the same phrase. It can be seen that there are many similarities between the two recordings. Speech recognition compares these similarities and simply verifies that the user is who they claim to be. Voice verification is one of the least intrusive biometric technologies, as it simply requires the person to say something as simple as their name. Another advantage of the voice recognition idea is that users do not normally need to purchase new hardware in order to implement the solution.

For the speech recognition software a trial version of a program called Anovea made by Anovea Inc was used. This piece of software required a microphone and also some speakers. The program was run on a laptop computer so the internal speakers were simply used for this. A Logitech desktop microphone was used for this application. This, alongside devices from the iris, fingerprint and signature experiments, is shown in Figure 3.



Figure 3: Biometric devices used

4. Results

An important factor that was looked at within the user-trial was how much that the general public knew about biometrics. A question was asked before the participants used each of the biometric devices simply asking whether or not they had ever used the particular biometric technology before.

The results show that the Biometric technology that the participants had used the most was Signature recognition (43% had used before). Although again this most likely due to people believing that signing for goods with a credit card is the same thing. Due to the fact that Signature recognition requires expensive equipment (PDA, or touch sensitive device) and is also one of the newest biometric technologies available, it is highly unlikely that this percentage of people had actually used it many times. Fingerprint recognition is one of the most widely used biometric technologies available and this is reflected in the fact that over one third of the people who conducted the user trial have used it before.

The two most rarely used biometric technologies were Iris recognition and Facial recognition. This is due to the fact that both these technologies have a reputation for being expensive and are more specialised to an application. Thus they are not used so much in applications that the public would have access to. Most of the reason behind Iris recognition being expensive when compared to other biometric methods was due to a twenty-year patent covering the technology, thus not allowing other researchers to develop it and create better-priced versions (ZD Net, 2006).

In conclusion to the question ‘How aware are the public of biometric technologies’ it can be seen that roughly 80% of people have heard of biometric technologies. Whereas from working out the mean 71.8% of people have never used one of the common biometric technologies used in this user trial.

An important factor of the user-trial was to see what type of biometric technology on test users favoured the most? This question was answered by asking the participant about how they perceived the security and usability of each device.

	Very Secure		Very Insecure		
Signature	4 13%	8 26%	6 20%	3 10%	5 16%
Voice	3 10%	8 26%	9 30%	4 13%	2 6%
Finger	14 46%	12 40%	3 10%	2 6%	0 0%
Face	7 23%	14 46%	5 16%	4 13%	0 0%
Iris	21 70%	7 23%	2 6%	0 0%	0 0%

Table 1: How would you now rate the security of each approach?

From the results in Table 1 it can immediately be seen that the majority of people believe Iris recognition to be the most secure of the biometric technologies on test. This is of course wholly accurate; as already indicated, Iris recognition is the most accurate of all the biometric technologies. Fingerprint recognition also gained a good rating from participants this technology has been proven to be in theory very secure. It can be seen that signature and voice recognition scored the lowest in the question this is again hard to decide as their have been very little surveys on both these devices with regards to security. The only information available on security is from the manufacturers of the devices so cannot be taken to be completely accurate.

	Very Easy		Very Hard		
Signature	16 53%	10 33%	3 10%	0 0%	0 0%
Voice	7 23%	12 40%	6 20%	3 10%	2 6%
Finger	20 66%	7 23%	3 10%	0 0%	0 0%
Face	7 23%	11 36%	5 16%	5 16%	2 6%
Iris	10 33%	12 40%	6 20%	2 6%	0 0%

Table 2: How would you now rate the usability of each approach?

Again, Fingerprint recognition (Table 2) can be seen to have the highest percentage of people rating it highly, this was expected from looking at the other results. It is quite probable that people like the idea of Fingerprint recognition because it is not as intrusive (in the sense of the time and effort required) as some of the other technologies, such as Facial and Iris recognition. It was also found from the author's point of view to be the easiest to use of the biometric devices on test.

Signature recognition was also rated highly with regards to usability, again this is most likely due to the fact that it is non-intrusive it is also something that the majority of people are used to doing. Voice and Facial recognition achieved amongst the lowest scores in terms of usability. This can be related to the fact that both of these devices took the longest to enrol with, also it can be seen that both of these devices are very intrusive of the user. For instance a number of people do not like having their photo taken, which is the way in which Facial recognition works.

As Fingerprint recognition has been rated so highly it is interesting to look into what application the general public have decided it to be most useful for. Within the user-trial a question was asked 'How appropriate do you consider Fingerprint recognition in the following scenarios' from this the participants were given a number choices to select: Passports and airport check-ins, Proposed national ID cards, Cash card ATM machines, logging onto computers, Entry into buildings, Verification of mobile phone users and keeping track of employee work hours.

For Fingerprint recognition the vast majority of participants rated it highly for use with Passports and Logging onto a computer. Although interestingly with regards to the proposed national ID cards people did not rate it so highly as for the use in passports with 50% and 66% respectively.

5. Conclusions

This paper has presented a wide-angled view of the way in which the public views biometrics. In that it has shown how aware they are of biometrics and also their reactions to different types of biometric devices.

From looking at the usability and security factors, fingerprint recognition looks to achieve the most well-rounded score. There are of course many reasons as to why the public may believe fingerprint recognition to be the best in these factors. Fingerprint recognition was one of the first biometrics to be developed, so it may simply be that the public are just more accustomed to the idea of having their fingerprint read. Another factor is that fingerprint recognition is one of the least intrusive biometric technologies. But again Iris recognition also scored very well in each of the three sections, which could be considered an intrusive technology as it requires a picture of the participant's eye and also involves a lot of aligning.

6. References

David Lake (2001), "Aye for an Eye – Biometric security systems –Industry Trend or Event", 23/07/01, *The Industry Standard*

Biometrics Information Web site (2006), “Fingerprint Types”
<http://perso.orange.fr/fingerchip/biometrics/types/fingerprint.htm> (Cited 09/06)

CNN News Web site “Iris Accuracy Estimations” www.cnn.com (Cited 11/05)

Forensics Information Web site (2006), “Fingerprint Identification techniques”
http://forensicsHQ.com/fingerprint_identification.php (Cited 08/06)

PDA Lok Company Web site (2006), “Digital Signature recognition”
www.pdalok.com/about_biometrics/digital_signature_recognition.htm (Cited 07/06)

Scotsman News Web site (2005), “Passwords add up to information overload for brain”,
04/10/05, <http://news.scotsman.com/index.cfm?id=2034192005> (Cited 11/05)

Speech Technology Web site (2006), “Feature specification”
www.speechpro.com/production/?fid=45 (Cited 06/06)

World population Web site “Current world population” www.ibiblio.org/lunarbin/worldpop/
(Cited 11/05)

ZD Net Web site (2006), “Foolproof Iris recognition technology?” 26/11/05,
<http://blogs.zdnet.com/emergingtech/?p=88> (Cited 08/06)

Analysis of End-to-End Techniques for Bottleneck Bandwidth & Path Capacity Estimation

T.Edwan, B.Ghita and X.Wang

Network Research Group, University of Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

This paper presents a new passive technique for estimating the bottleneck bandwidth based on transferring the Gaussian kernel density estimation of the packets inter arrival times to the frequency domain, the resultant spectrum contains information about the transmission time of the bottleneck link and can reveal information about multiple bottlenecks if they exist along the end-to-end path. The advantage of the technique is that it provides a model that can be manipulated by the digital signal processing methods which is different than the previous work done in this field that relies too much on statistical methods. The new technique was validated using the ns2 network simulator, several topologies and traffic sources were tried. The post-simulation analysis was done by the programs and Unix/Linux Bourne (and Bash) Shell scripts developed for this purpose. In addition to that, some experiments were conducted to test the strength of the patterns between flows that share a bottleneck by applying K-means algorithm to cluster the average packets inter arrival times of these flows. The document also presents the results of real traffic experiments conducted in an attempt to infer both the bottleneck bandwidth and the capacity of the path using a passive approach.

1. Introduction

With the evolution of the Internet, and the maturity of its infrastructure, along with the increase of network traffic, efforts now are redirected to services associated with it and how it can be improved. One major parameter that has a direct effect on these services is the bandwidth. A lot of effort has been done in the past few years by several researchers trying to find an adequate way to estimate the bandwidth and define several terminologies that will give better understanding of what exactly to measure and at the same time serve as a base for future work.

The necessity of an adequate technique for measuring the bandwidth can be seen from two different perspectives. Firstly, the applications and protocols, whether they are involved in file transfer and content delivery or they are involved in real-time streaming media, they require accurate measurement of the bandwidth, as shown by Kiwior et al (2004). This is crucial especially in future when we will be dealing with multimedia applications, for example an application unaware about the available (or bottleneck) bandwidth can stream a 5GB video over a 19.2Kbps cellular data link or send a text-only version of a web site over a 100Mbps link. Knowledge of the bandwidth along a path allows an application to avoid such mistakes by adapting the size and quality of its content, (Fox 1996), or by choosing a web server or proxy with higher bandwidth than its replicas, (Stemm 1999). Secondly, network operators are also concerned with traffic engineering, routing, network capacity, and network troubleshooting issues, in addition to other issues regarding the verification of service level agreements and Quality of service (QoS) (Harfoush et al 2003).

There are two well known approaches for measuring the bandwidth, one is to measure it hop-by-hop which is considered to be inefficient, and the other is end-to-end which requires only two nodes (the sender and the receiver). End-to-end approach in turn can be used in two different ways : active (intrusive) or passive (non-intrusive), the former is seen to have some problems, particularly it is slow and because of the packet injection to the network (that will compete with the original traffic and might also inject more load in to the network) is considered inaccurate, moreover the Internet traffic is not static, so the measurement using this way (active) will give an indication of the bandwidth just over a certain time interval which is in this case the measurement time interval. On the other hand non-intrusive (passive) methods have less measurement overhead. They basically depend on capturing existing traffic in the path of interest and try to estimate the bandwidth by inferring traffic patterns. Passive techniques can be used to analyse huge amount of traffic captured for several years, detect the trends and the evolution of bottlenecks. It is also believed, Katti et al. (2004), that passive techniques fit large scale network traffic measurements better than active methods that suffer from probing overhead, taking into account that we are always interested in real large scale network traffic as this is the case for the Internet.

First in section 0, the paper gives an idea about what is the bottleneck and what should be measured, then in section 0 it focuses on the passive approach and the new technique used in the estimation of bottleneck bandwidth. In section 0, a discussion of the experiments that were conducted to validate the new technique is provided. An example of inferring the bottleneck bandwidth and the capacity of the path using a passive approach in a real traffic environment is presented in section 46. Finally, the paper concludes with the limitations of the new technique (section 0), the alternatives for improvement (section 0) and an overall conclusion (section 0).

2. Bottleneck link and bandwidth

Some researchers define the bandwidth of a link as the maximum transmission rate that could be achieved between two hosts at the endpoints of a given path in the absence of any competing traffic (Harfoush et al. 2003). Or it can be defined as the ideal bandwidth of the lowest bandwidth link on the route between two hosts, as shown by Thepvilojanapong et al (2002). However this seems to be ideal because we always have traffic along our path. A more practical definition of this link is the link that experiences a significant queuing, so it is the congested link and it is not necessary to be the link with the minimum capacity in an end-to-end network path (Katabi and Blake 2001). In fact a bottleneck occurs when a congestion occurs and this takes place when data arrives on a large capacity link (like a fast LAN) and gets send out to a smaller link (like a slower WAN) or when multiple inputs streams arrive at router whose output capacity is less than the sum of the inputs (Stevens 1994). At both cases the queue will build up. Figure 1 illustrates a typical scenario.

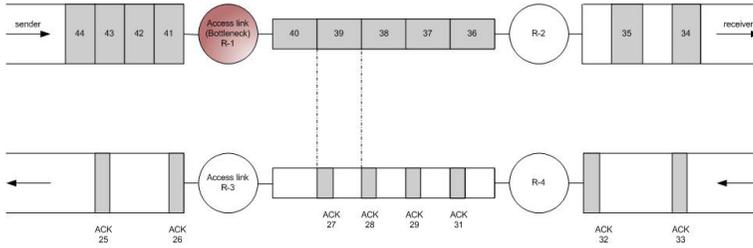


Figure 1: Congestion case

If the path is symmetric, R_1 and R_3 will be the same router as are R_2 and R_4 . If we assume that the packets are arriving to R_2 from a WAN and the router will then pass them to a LAN, the packets will maintain the same spacing as they did on the WAN on the left of R_2 . In the same manner the spacing of the ACKs on their way back is the same as the spacing of the slowest link in the path. This seems to be logical but it is very important when trying to infer the graphs of the packet inter arrival times in the passive approaches.

3. Non-intrusive bottleneck bandwidth estimation

This section discusses the proposed approach, which relates to the automation of equally spaced gaps detection mentioned by Katti et al (2004). These equally spaced gaps are located between the probability density function of the packets inter arrival times, and they are usually multiples of the transmission time (the time to transmit one packet) of the bottleneck link. Rather than using the pdf of the equally spaced gaps (as in *multiQ*, a passive tool developed at the M.I.T. (Katti et al 2004)), the method proposed by this study is to perform Gaussian kernel estimation on the packets inter arrival times of the flows of interest then transform the result to the frequency domain in order to detect the repetition of the mode spikes which will map to the transmission time of the packets on the bottleneck link. Note that theoretically if there is more than one pattern (frequency) it can also be detected (usually the most congested bottleneck dominates the pattern). Before applying this approach and in an attempt to study how the packet inter arrival times for certain flows behave, in the existence of cross traffic, K-means clustering was applied to cluster the average inter arrival time of each pair in a group of TCP flows available at the receiver.

3.1 Classification of flows using K-means

In Pattern Recognition and unsupervised learning of neural networks-means, algorithms like K-means are typically used for clustering. Clustering here means to group the objects based on a certain feature into a number of groups and this is done by minimising the sum of squares of distances between data and the corresponding cluster centroid. As a result the data is classified into K-clusters. All the data in one cluster is very similar (relatively close in values). The procedure is to first take all non repetitive combinations of a group of flows available at the receiver, the flow here is defined as the IP address and port number. The number of non repetitive combinations can be calculated from the following equation:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (1)$$

where n is the number of flows and $k = 2$. After that, the average interarrival time for each pair of flows was calculated, then k-means algorithm was applied to cluster the samples so that the sum of squares (or the Euclidean distance) within a cluster is minimised, according to the following equation (Webb 2002):

$$S_W = \frac{1}{n} \sum_{j=1}^g \sum_{i=1}^n z_{ji} (x_i - m_j)(x_i - m_j)^T \quad (2)$$

Several clustering methods were applied to evaluate how strong is the relation between the flows that do share a bottleneck. The aim of this was to see if it is possible to detect a shared bottleneck directly from packet inter arrival times.

3.2 Novel Technique for estimating bottleneck bandwidth

The new technique for estimating the bottleneck bandwidth proposed in this paper is based on transferring the Gaussian Kernel Density Estimation of the packets inter arrival times to the frequency Domain. Kernel density estimation is a standard technique for constructing an estimate of a probability density function from measurements of the random variable. In fact kernel estimators are extension to histograms whose disadvantages provide the motivation for kernel estimators. In a histogram, it is important to consider the width of the bins (equal sub-intervals in which the whole data interval is divided) and the end points of the bins (where each of the bins start). So the problems with histograms are that they are not smooth, and they depend on the width of the bins and the end points of the bins. These problems can be alleviated by using kernel density estimators which will provide the required smoothing. KDE is expressed mathematically as:

$$f_x = \frac{1}{n} \sum_{i=1}^n k\left(\frac{x - x_i}{h}\right) \quad (3)$$

where K is the kernel function. There are various choices among the kernel function, however several of them have $\int K(t) dt = 1$ and have peaks at the centre (at each point). Because the points here represent the modes it was decided to adjust the kernel function so that the overall graph will look very similar to a sinusoidal plot. By doing this it is possible to transfer the new graph to the frequency domain by performing Fourier Transforms and still obtain all the frequencies in our "signal" in a manner reassembling the ordinary analysis of electrical signals. Any repeated pattern in the original graph will give a frequency bump in the new graph from which the transmission time of the bottleneck link can be determined. The Gaussian kernel function was used (Cosine function may produce better results) and the bandwidth h should be adjusted for not to overestimate the density, we can look at this as a resolution problem: adjust the variable h until you get a bump in the frequency domain (if there is a bottleneck).

The final result after applying the KDE on inter arrival times would be an equally separated modes that usually decrease in amplitude as we move far from the global

mode, this would be similar to a damped sinusoidal plot. The assumption was that the damping frequency (the bump in the frequency domain) will directly map to the transmission time of the packets, in other words it will give an indication if there is a bottleneck and an estimate of its bandwidth. The damped sinusoidal is expressed as:

$$f(x) = \exp(-ax) [\sin(\omega_b x)] u(x) \Leftrightarrow F(\omega) = \frac{\omega_b}{(a + j\omega)^2 + \omega_b^2}, a > 0$$

The motivation for this is to propose a model for the passive approach that can be ported to any digital signal processing tool for better analysis (such as filtering techniques) and from an implementation point of view it can be easily implemented as a hardware device (bandwidth analyser).

4. Validation

4.1 K-means

A simple bottleneck simulation scenario was used: nine sources connected to one destination using transmission control protocol TCP through three bottlenecks *B1*, *B2*, *B3* with 1.5Mbps, 2Mbps, and 2.5Mbps respectively. Each of the nine sources is connected to its gateway by a 10Mbps link and each three share a bottleneck. Following the simulation, the flows were separated by first calculating all possible non-repetitive combinations according to equation (1), followed by calculation of the K-means clusters.

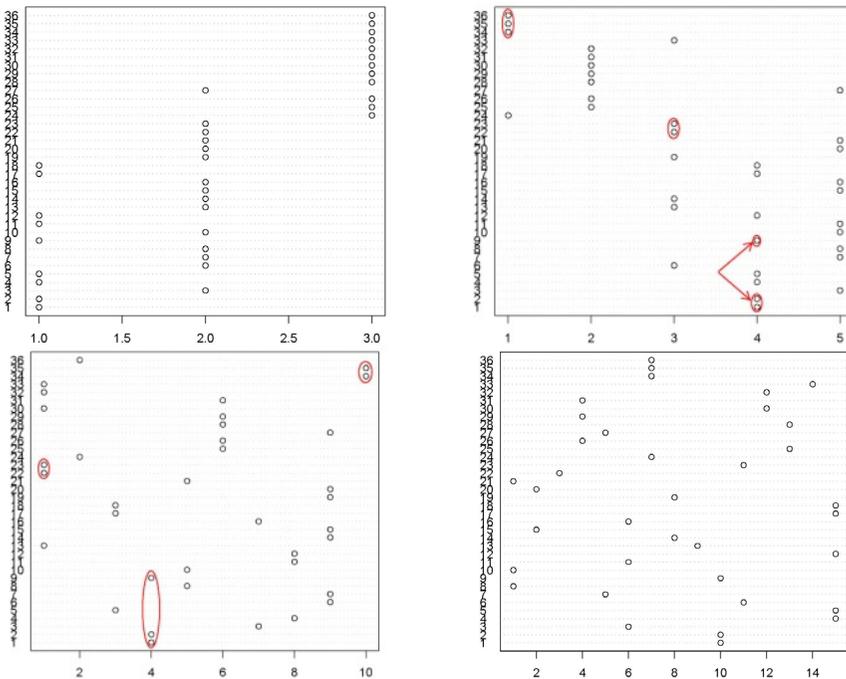


Figure 2 - Resulting clusters when using different K-means clustering sizes

As it can be seen from the above graphs, some patterns persist but a lot are lost especially when increasing the number of clusters. For example when the number of clusters is 5, groups 34-[7,8], 35-[7,9], 36-[8,9] did cluster correctly but with additional group 24-[4,7] that should not be in this cluster. Notice here that flow 4 belongs to the 2Mbps bottleneck and flow 7 belongs to the 2.5Mbps bottleneck and the first bottleneck (1.5Mbps) did not contribute to the error. Also Notice that groups 1-[1,2], 2-[1,3],9-[2,3] all share a bottleneck and they have a strong pattern that persist even when the number of clusters is increased to 15 clusters and 20 (not shown). Also notice that groups 1,2,9 all share the bottleneck with the least bandwidth (1.5 Mbps) compared with other bottlenecks (2 and 2.5 Mbps). On the other hand groups 22,23 and 27 (that belong to the middle bottleneck) were severely affected by the flows from the other two bottleneck as they are both close to its data rate. From this discussion it can be observed that the closer the bottlenecks bandwidths are the more difficult to separate their flows and thus the more difficult to detect them. The second observation is at higher bandwidths the clustering error increase (failed to cluster correctly) while at lower bandwidths strong patterns persist, this seems to be logical since at higher bandwidths the transmission times are small and close to each other in contrast to lower bandwidth bottlenecks.

4.2 Estimating bottleneck bandwidth in frequency domain

To validate the new technique of converting our kernel density signal (the term signal is used as an analogy with the electrical signal) to the frequency domain, the simulation scenario in Figure 3 was constructed : 9 FTP sources, each three belongs to a path and they are connected to their gateways via 10Mbps links, each path contains 5 hops. The middle path contains 4 links from the gateway to the sink (which is the observer's point, in this case the receiver), 1.5Mbps, 2Mbps, 3Mbps, and 2Mbps. At hop 3 and 4 there are 9 cross traffic sources for each, they are using the same path as our packets (path persistent) and are connected to a different sink by 20Mbps link. First constant bit rate cross traffic sources were used then Pareto pdf sources were tried. In some cases the sources were modified for not to generate traffic all at the same time, but rather they will fire their packets three by three.(first three start, then after a while the other three and so on). The simulation time was 22 seconds. At the end of the simulation, our program extracts packets from a group of two flows (Flow-1 in Figure 3) and computes the packets inter arrival times then the kernel density estimation and transfer the results to the frequency domain by applying Fast Fourier Transforms on the resultant ``signal". Figure 4 (a-f) depicts some of the results when the kernel bandwidth was changed.

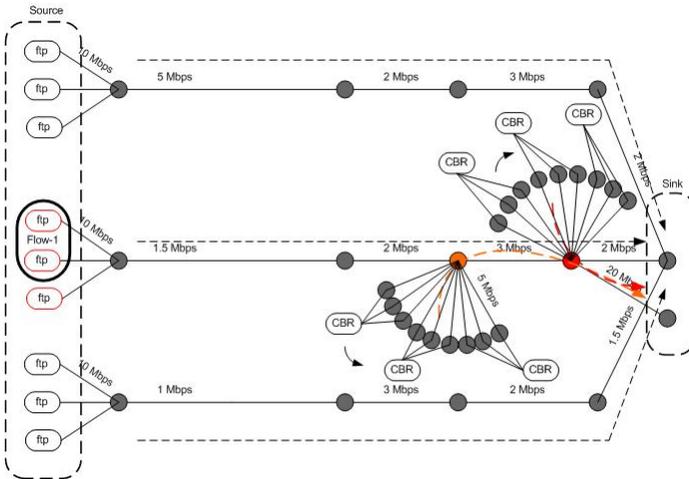
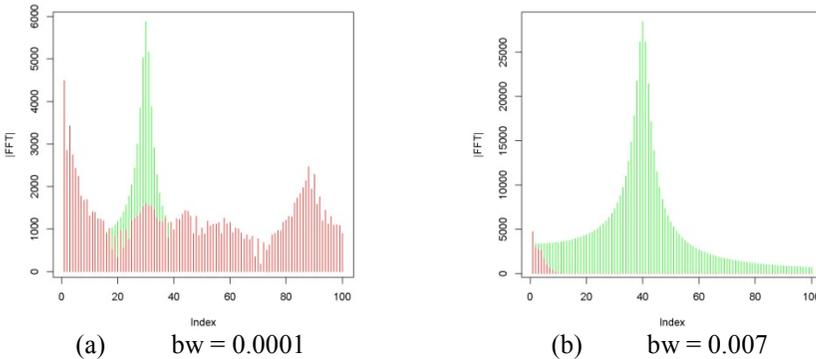
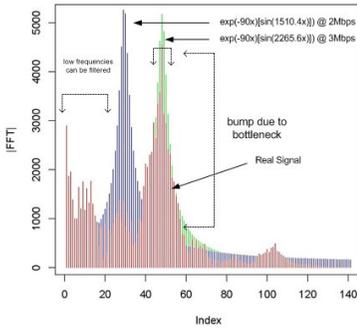


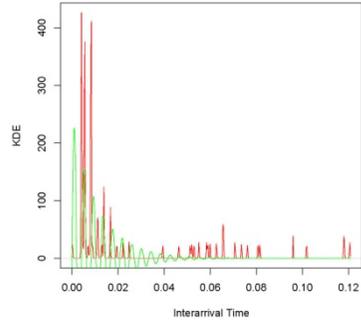
Figure 3 - Simulation topology

For the sake of comparison a damped sinusoidal function was also plotted in the same KDE plot, because as mentioned in section 0 the assumption was that there might be a correlation between this function and the KDE. The period of this function was chosen as the transmission time of the bottleneck (2Mbps in this case), this is the time to transmit a packet of 1040 byte. The packet size was 1000 byte and 40 bytes for the TCP-IP headers}, giving $T = 2.773 \text{ ms} \rightarrow f_p \approx 360.58\text{Hz} \rightarrow \omega_p = 2\pi(360.58) = 2265.6$. In the first case in Figure 4d a small kernel bandwidth was chosen which results in an under smoothed high noise signal which can be seen as a wide spectrum in the figure. While in the second case when the kernel bandwidth was increased by nearly the factor of 10, the signal was lost except for the low frequency envelop. Finally when a kernel bandwidth was chosen between the two extremes a normally smoothed signal was produced which is highly correlated to the reference signal. Notice in Figure 4 the first bump and even the tail of the rest of the plot nearly matches our reference signal. Also notice the persistent low frequency pattern in all graphs that corresponds to the envelop of our signal, this can be removed by applying filtering techniques as it does not contain any significant information in this case.

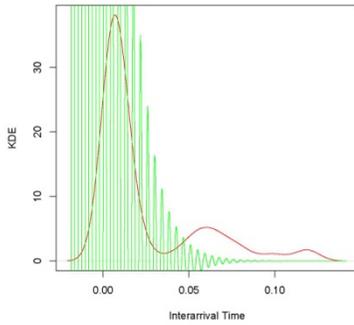




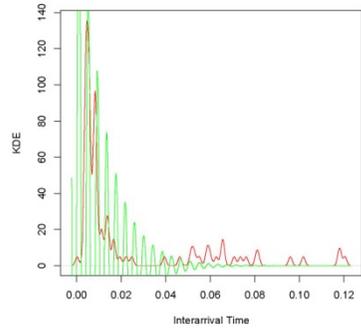
(c) $bw = 0.0008$



(d) $bw = 0.0001$

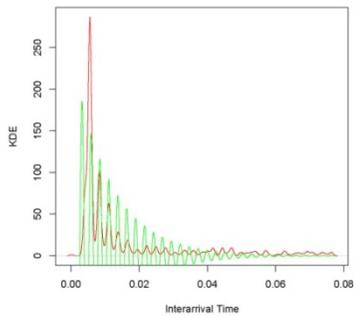


(e) $bw = 0.007$

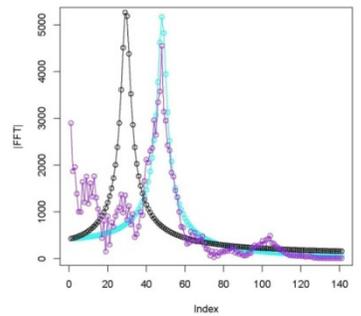


(f) $bw = 0.0008$

Figure 4 - Effect of changing the kernel bandwidth: Under smoothed (wide frequency spectrum), Over smoothed (loss of signal except the low frequency envelop) and normally smoothed (the signal has a bump at the bottleneck frequency reciprocal of transmission time)



(a) KDE and damped sinusoidal correlation

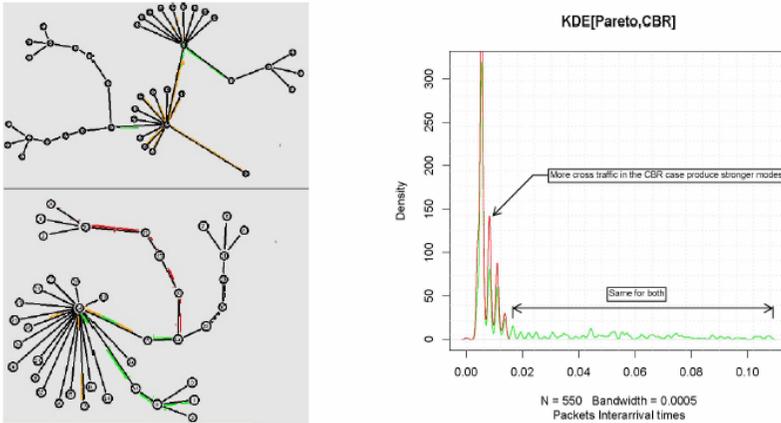


(b) KDE and damped sinusoidal – frequency response

Figure 5 - New technique for estimating the bottleneck bandwidth in the frequency domain

Several alternatives were used to benchmark the proposed method, for example when applying Pareto cross traffic instead of constant bit rate it was found that this has no effect on the overall pattern, the only difference was that the cross traffic is less intense than in the constant bit rate case and thus the modes still exist in their same

places with their same separation but with less amplitude. This can be clearly seen Figure Figure 6b. The Pareto pdf sources were applied to model (interactive) Internet traffic flows. And because the traffic is bursty, transmission only takes place during on periods. This heavy tailed pdf can be expressed as: $f(x) = ax^{1-a}, a > 0, 1 \leq x \leq \infty$, where a is the shape parameter, when it is near to one it gives rise to self-similar traffic, whereas a near to two, has similar fractal properties to exponential traffic. The default 1.5 was used. In general terms, a heavy tailed distribution can give rise to very large file sizes.

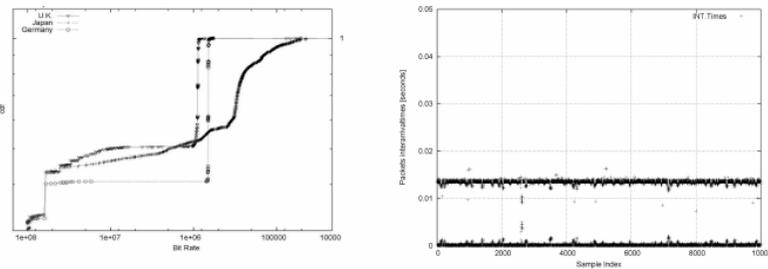


(a) ns2 simulation topology (b) KDE interarrival times for Pareto (green) and CBR (red) sources

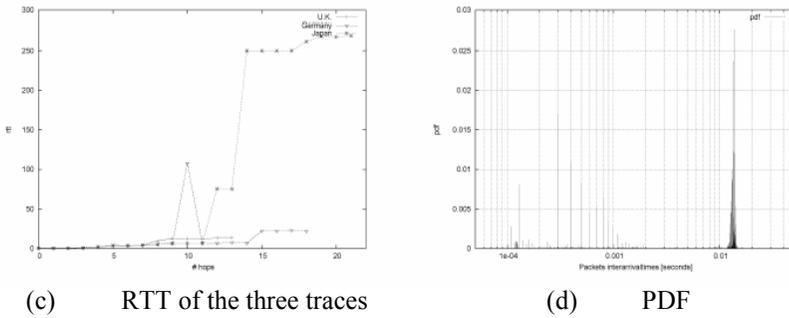
Figure 6 - Effect of applying Pareto vs. CBR

5. PDFs and Real traffic

In order to see how much information the pdf of the packet interarrival times contains on the path capacity and the bottleneck bandwidth, several experiments were conducted in real environment. The UoP network was used to download a file (116.5 MB) from different repositories: [Keihanna, Japan Asia], [Duesseldorf, Germany Europe], [Kent, UK Europe]. 10,000 packets were dumped to a trace file using *tcpdump* from each destination, the packets then were analysed, pdfs and cdfs were calculated and plotted at the same graph.



(a) CDF of the three traces (b) Packet interarrival times



(c) RTT of the three traces

(d) PDF

Figure 7 - Real traffic analysis

By examining Figure 7a it can be seen that in the case of [Duesseldorf, Germany] most of the packets arrived spaced by a time nearly equal to the $1.25\mu\text{s}$ which gives a rate of 800Kbps. This should give an indication about the path capacity (the same applies for the other plots), although inferring the path capacity is not a simple task, some look at it as the minimum inter arrival times for back-to-back packets others see it as the global mode in the distribution of packet inter arrival times, while some researchers see that both can give wrong results and suggest an examination of the pdf and the location of the bumps as an alternative approach (Katabi and Blake 2001), however the global mode here is very clear (no noisy samples) and most of the packets arrived at this rate.

Now to infer the bottleneck the pdf plot must be considered, An important point to note here is that the network interface card for the machine was 100Mbps Fast Ethernet, while the total link capacity was 620Mbps with a physical implementation of 4 STM-1 lines of 155Mbps each. And per-flow basis traffic splitting was assumed so the bottleneck speed would be 155Mbps rather than 620Mbps if there is a significant queuing at that link. It was found that bottleneck usually occur at the access links because of the significant queuing. Consider the [Kent, UK] case were the pdf are equally separated by 0.1 ms - assuming a packet size of 1500 bytes, this gives 120 Mbps as a bottleneck bandwidth estimate. Figure 7d, shows that the spacing between the modes was considered and not the packets inter arrival times which in our case cannot exceed the 100 Mbps network card rate.

Comparing the results with other tools, and by looking at the first hop estimation, both *Clink* and *pchar* estimate the total lines bandwidth 620 Mbps, in addition to that the hop queuing was at its maximum at the first hops and decreased to zero at the third hop which means that access links are more likely to have queuing. *pathneck* was used to determine the choke points along the path. When considering the path to [Keihanna, Japan Asia] it was found that the packets inter arrival times give a high fluctuating pattern this can be seen from the cdf and RTT graphs. In the cdf graph it is still possible to estimate the path capacity as the global mode.

6. Limitations

Applying the K-means approach to group flows that share a bottleneck was just to sense how strong is the relation between these flows without using the *entropy* as a

decision rule as mentioned in (Katabi and Blake 2001), however trying to cluster the average inter arrival times for each pair of flows, did reveal the strong relationship for limited cases with the bottleneck bandwidths used, in fact some flows can still be grouped in one cluster even when the number of clusters was increased to ten, but the rest did not show this relation this is because the averages of other flows are still close to the flows of our interest (that share a bottleneck) and fail to cluster correctly. On the other hand applying our new technique (the KDE in the frequency domain) in a controlled simulation environment did produce the expected results but still needs more testing in large scale environments. In addition to that the idea of using a certain kernel function like the Gaussian or the cosine with varying bandwidth is still foggy. Particularly making the bandwidth of the kernel function variable may change the original signal (the term signal is used as an analogy with the electrical signal) dramatically and this may have a great impact on the frequency response, although it will still indicate if there is a bottleneck or not based on the existence of the bump this may change the estimated rate. One more thing to mention about this new technique is that if for any reason an additional perturbing modes exist between the equally spaced modes the pattern will be disturbed and this will contribute to the error in the frequency measurement. Particularly this will overestimate the frequency. Finally, other parameters like the clock resolution (*tcpdump* timestamps errors), delay can still have their effects on the measurement.

7. Future work

This project lends itself to extension in more than one direction; one of these directions is to consider more complex scenarios and testing the new technique in real traffic environments for longer periods. Another issue is to consider the effects of packet loss and delay which may have impacts on the estimation results. Studying the resultant traffic "spectrum" for longer periods that can be obtained by the new technique is an important issue that must be considered because this will reveal a lot of information about the bottleneck evolution and the network growth. Implementing the method in real-time is still a challenging issue especially with the necessary adjustments needed for the bandwidth of the kernel density which should be done by the user during the measurement. In fact, we must keep in mind that the output of this method is a visual information that need to be inferred logically by the user which means that the output cannot be provided directly to a non-human (non-intelligent) system (an application for example), thus involving artificial intelligence techniques (like Artificial Neural Networks) might be considered in the future.

8. Conclusion

A substantial amount of prior research focused on modelling the Internet traffic and understanding the parameters that affect its performance; amongst these parameters, bottleneck bandwidth is one of the critical ones. Two end-to-end approaches were used to estimate bottleneck bandwidth, one is active and the other is passive. This paper promotes the passive techniques by proposing a modification to a passive algorithm. The focus was on passive techniques mainly because of their less overhead of measurement and because they do not have an intrusive nature that may affect the measurement, in addition to their ability to analyse huge traffic captured

for years which facilitate the understanding of the bottleneck evolution and the Internet growth. One important outcome of this project is the new model that transfers the analysis environment to the frequency domain where digital signal processing techniques can be widely applied and hardware implementation can be considered. In addition to that the new model provides a new perspective in the field of estimating the bottleneck bandwidth.

9. References

Chiu D.-M. and Jain R., 1989. Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks, *Computer Networks and ISDN Systems*, 17(1):1-14, 1989.

Claffy, K., Dovrolis, C., 2004. Bandwidth Estimation: Measurement Methodologies and Applications, [Online] http://www.scidac.org/March2004/ascr_net_2.html [accessed 01 September 2006]

CAIDA, 2006. Cooperative Association for Internet Data Analysis [Online] <http://www.caida.org> [accessed 01 September 2006]

Harfoush, K., Bestavros, A., Byers, J., 2003. Measuring Bottleneck Bandwidth of Targeted Path Segments Proceedings of the IEEE INFOCOM 2003, San Francisco, CA, USA, March 30 - April 3, 2003.

Hu, N., Steenkiste, P., Li, E.L., Mao, Z.M. and Wang, J., 2004. Locating Internet Bottlenecks: Algorithms, Measurements, and Implications, *Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*, Oregon, USA

Hu, N., Steenkiste, P., Li, E.L., Mao, Z.M. and Wang, J., Pathneck, 2006. [Online] <http://www.cs.cmu.edu/~hnn/pathneck/> [accessed 01 September 2006]

Katabi, D., Blake, C., 2001. Inferring Congestion Sharing and Path Characteristics from Packet Interarrival Times <http://www.lcs.mit.edu/publications/pubs/PDF/MIT-LCS-TR-828.PDF> [accessed 01 September 2006]

Katti, S., Katabi, D., Blake, C., Kohler, E. and Strauss, J.(2004) M&M: A Passive Toolkit for Measuring Tracking, and Correlating Path Characteristics, [Online] <http://nms.lcs.mit.edu/~dina/MNM/mmdocs/paper.PDF>, [accessed 01 September 2006]

Key, P., Massoulie, L., 2003. Fair Internet traffic integration: network flow models and analysis, Statistical Laboratory, University of Cambridge, Cambridge CB3 0WB, UK, <http://www.statslab.cam.ac.uk/~frank/PAPERS/kmbk1.PDF>, [accessed 01 September 2006]

Kiwior, D., Kingston, J., Spratt, A., 2004. PATHMON, A Methodology for Determining Available Bandwidth over an Unknown Network, [Online] http://www.mitre.org/work/tech_papers/tech_papers_04/kiwior_pathmon/kiwior_pathmon.PDF [accessed 01 September 2006]

Stevens, W.R., 1994. *TCP/IP Illustrated, Volume 1*, Addison Wesley, ISBN 0201633469.

Teknomo, K., n.d., K-means Clustering Tutorials, [Online] <http://people.revoledu.com/kardi/tutorial/kMean/index.html>, [accessed 01 September 2006]

tcpdump, [Online] <http://www.tcpdump.org> , [accessed 01 September 2006]

Thepilojanapong, N., Tobe, Y. and Sezaki, K., 2002. One-way Delay Measurement and Bottleneck Bandwidth Estimation, *Joho Shori Gakkai Shinpojiumu Ronbunshu Journal*, vol. 2002, no. 15;pp. 39-44

Webb, A., 2002. *Statistical Pattern Recognition*, 2nd Edition, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, England

Public awareness of biometrics

K.Evangelatos and S.M.Furnell

Network Research Group, University of Plymouth, Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

A fair degree of media attention has been devoted to biometrics, with the impression often being created that either they offer a panacea to our authentication needs, or that they are a way for governments and private organizations to monitor our activities. From this basis, a survey has been created to benchmark the public awareness of biometrics and their resulting attitudes, as these have been formed by the media coverage. The results revealed that people are much more accepting of those biometric systems that they are aware of (e.g. fingerprint) and which are more convenient to operate (e.g. signature) rather than the systems that they believe to be more secure (e.g. iris). Thus, it has been concluded that convenience, awareness and practical experience are essential requirements for the public acceptance of biometrics.

Keywords

Authentication, Biometrics, Survey, User Awareness

1. Introduction

Nowadays with the wide adoption of Information Technology (IT) and the Internet, identity thieves continuously find more sophisticated ways to perform frauds. To protect people from this threat, user authentication has been accepted as being the first line of defense against identity theft (Furnell *et al.*, 2000). Until today the public was relying on the “classic” authentication mechanisms of passwords and tokens, considering biometrics as something that belongs in science fiction. But nowadays the increased usage of electronic transactions and the increased level of frauds and terrorist attacks made passwords and tokens inadequate to prevent identity frauds; the ‘crime of the century’ as it has been characterized by the United States Department of Justice (2006). For this reason biometrics are constantly proposed in more and more applications, especially after the terrorist attacks of 9/11. This has unavoidably drawn a fair degree of media attention and as a result it is expected that a far greater number of people have now heard of biometrics. But how are biometric technologies presented by the media and how does this coverage affect the end users? This question is of particular interest since the public perception can impact upon user acceptance of biometric technologies as authentication mechanisms, which is one of the most important barriers to their wide deployment. From this basis, a survey has been conducted to benchmark the public awareness of biometrics and their resulting attitudes, based upon what they have seen or heard from the media. This paper begins by presenting the way that biometrics are presented by the media. Then details of the survey and an analysis of the obtained results are presented, leading to the conclusions that have been drawn.

2. Biometrics and the media

There are numerous movies, such as *Minority Report* and *I Robot*, where the heroes have to use their biometric characteristics in order to gain access to restricted areas, computers and files with sensitive data or even to start their cars. Although such movies present biometric systems as a ‘James Bond’ technology capable of providing absolute security (Black, 2002), they have made known to the general public the existence of the various biometric technologies and their potential applications. In addition there are articles, such as Chowdhary (2006), which suggest that with biometrics people will not have to remember numerous passwords and PINs or to worry if they forget their keys, their passports and of course their passwords. All these have created great expectations and formed the opinion that biometrics can be a panacea to our security problems. As a result the biometric industry has to overcome this ‘Hollywood cure’ (Wait, 2003) if users are to realise the true benefits of biometrics.

Additionally, the larger part of the media expresses concerns that biometrics are a threat to our privacy and civil liberties, especially in the case of passports and ID cards. Numerous articles mention that government agencies will want to obtain records with our biometric data, so that they can link them together with other information (e.g. criminal and tax records) and “have a complete picture of our private existence” (Mogg, 2006), creating the fear of Big Brother. Other articles are going further, considering the security of the centralised databases that will hold our personal data. Characteristic is the statement of Wood (2006): “creating one huge database could be the perfect gift for sophisticated computer hackers”. This is truly a major risk since if the databases are compromised and the citizen’s characteristics are stored as raw data (and not as one-way encrypted templates) they could be easily reconstructed and used by identity thefts. Such coverage most probably will have negatively affected the perception of the public, relating all the applications of biometrics with the initiatives of the various governments.

Although a significant number of commercially available biometric devices and many pilot tests have been run the last few years, only few people have a practical experience with such devices. In this context it is of interest to investigate the perception of the public about biometrics as it has been formed by the controversial media coverage described above.

3. Surveying public awareness of biometrics

In order to determine the user acceptance of biometrics as an authentication mechanism, a survey was conducted to assess public awareness about biometrics technologies and the attitudes of the potential users, as these might have been formed by the media.

The survey consisted of 34 questions, the majority being multiple choice, and was divided into three sections. The first section, questions 1 to 5, aimed to provide some demographic characteristics of the respondents. The second section, questions 6 to 10, included some general questions about the respondents’ attitudes to IT and

security. These helped to identify the extent of IT and security awareness of the respondents. The third section, questions 11 to 34, intended to determine the user awareness and attitudes in relation to biometrics. The questionnaire was made available in two forms, a printed copy and an online version, and distributed through e-mails to a stratified random sample, chosen mainly from the mailing lists of the University of Plymouth, UK. These lists include staff and students from the various schools and departments of the University. Moreover printed versions of the questionnaire had been distributed randomly, to people in various locations.

The study was conducted over a four-month period, commencing in March 2006, during which approximately 350 e-mail invitations were sent to a wide range of individuals (based on their job/topic of study), their age and education level, with 154 completed responses being received. Additionally 80 printed surveys were distributed, yielding 55 responses, representing a total response rate of approximately 49%. In the following sections the results of the survey are analyzed trying to understand the public awareness and attitudes towards biometrics, as these have been formed by the media.

3.1 Demographic characteristics

The vast majority of respondents (81%) were aged below 30 indicating that they would have grown up with IT and therefore be more familiar with using it. This is justified by the fact that 59% of the survey respondents considered themselves as intermediate ability users, 35% as advanced and only 5% as novice. In terms of gender the sample was quite evenly distributed, with 57% of the respondents being female and 43% being male.

The majority of responses (79%) were from people with a higher education, reflecting the fact that the survey was mainly distributed through academic channels. Since this study was concentrated in UK, the majority of respondents (75%) were British citizens while the remaining 25% were foreigners working or studying in UK. This indicates that their perception about biometrics will have been mainly influenced by the way that the British media presents the topic.

In terms of the employment or study background a wide diversity was achieved, with 145 out of 202 respondents (72%) being from non-technological fields. This is of particular importance since the majority of the responses were from people with no or minimum IT education, indicating that their perception will have been most probably formed only from what they have experienced, heard or read about biometrics outside of academia.

Although at first glance the above demographic results do not suggest a truly representative sample of the general public, the achieved diversity can be considered as a fair reflection of the people that will form the base of the potential biometric users.

3.2 Public perception of IT security

The fact that 70% of the respondents consider IT security very important while no-one believes that it is not at all important indicates that the respondents are to some extent security aware. This conclusion is further enhanced by the fact that 75% of the respondents find identity theft a very serious threat, while only 1% of them believe that it is not a threat. Moreover the results clearly indicate that the majority of respondents (45%) have concerns about the ability of the currently employed authentication techniques to prevent theft in large scale systems (such as online banking), realizing the shortcomings of passwords and tokens.

Asking the respondents to rank their preferences from the security mechanisms used for authentication, surprisingly 125 out of 209 (60%) of them chosen as their first preference biometrics, although only 10% has used them before. The detailed results can be seen in Figure 1. This finding is in agreement with the results of a survey carried out by Unisys (2006), which found that 66% of the 1661 consumers worldwide favored biometrics. This fact further suggests the respondents' dissatisfaction with passwords and tokens, presuming that biometrics will be better. This is most probably caused by the media, which by presenting biometrics as a panacea to our security needs has increased the expectations of the users.

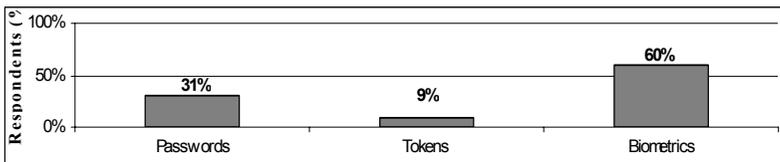


Figure 1: Respondents' preferred authentication mechanisms

3.3 User awareness

By asking the respondents if they have heard about biometrics before the survey it has been revealed that 135 out of the 209 (65%) are aware of them. However, when a list with various biometric characteristics was provided, only 10% of the respondents stated that biometrics does not suggest anything to them. This finding is in contrast with the results of the survey conducted on behalf of Citizenship and Immigration Canada (2003), which found that 44% of the 1,200 respondents "registered very low awareness", indicating that nowadays the public is much more aware of biometrics, which is a result of their extended coverage.

Following this question the 135 respondents were asked to indicate all the biometric technologies that they have either heard or used. The results showed that all of the mainstream technologies (see Figure 2) are well known, while a small percentage is even aware of newer techniques, such as the odor and gait recognition. Investigating further the user awareness, the 135 respondents were asked to indicate the proposed applications of biometrics that they are familiar with. The results showed that only 9 of the 135 (7%) respondents have not heard of any of the proposed applications (see Table 1). On the other hand, and as expected, the most known applications are citizen identification for border crossing (58%), ATM machines (58%), physical

access (53%) and PC/network access (51%), which are among the most discussed applications.

3.4 Media influence

Before investigating the influence of the media, the 135 respondents that are aware of biometrics were asked to indicate all the sources from where they have learned about them. The results showed that TV news (67%), newspapers and magazines (55%) are the primary sources of information, which is in accordance with the findings of Citizenship and Immigration Canada (2003) study. However, the rest of the findings show a major increase in the people that have heard about biometrics from various sources, reflecting the increased media coverage.

Investigating the extent of the media influence, the 135 respondents were asked to indicate any specific information that they could recall hearing or reading about biometrics. Quite few (43%) recalled various movies and articles about the proposed application of biometrics to passports and IDs, indicating the role of the media on informing the public. When these respondents were asked to state their opinion on how the media is treating biometrics, half of them described it as fair. This finding is quite important, showing that probably they have been influenced by both the part of the media that supports biometrics and by the part that criticizes them, forming a more rounded opinion.

To determine further the influence of the media, the 135 respondents were asked to state their opinion as to whether biometric systems can be easily cheated. The results showed that 72% do not believe that it is easy, indicating that the respondents have not been influenced by the media concerns that biometric systems can be cheated using fake characteristics, while 21% stated that they do not know, potentially suggesting that this coverage has confused them.

Continuing to investigate the role of the media, the 135 respondents were asked to indicate the biometric systems which in their opinion can be a health risk. Half of them stated that biometrics are not a health risk, while 37% believe that iris and 36% that retina recognition can be. This shows that quite few respondents have been affected by part of the media which consider iris and retina recognition as a health risk, and that they will most likely be reluctant to use their eyes for authentication purposes.

By asking the respondents to indicate on a 5-point scale the extent to which they believe that the mainstream biometric systems can work reliably, a rank of the of the expected reliability, Figure 2, has been obtained by adding the total number of positive responses (*'extremely reliable'* and *'very reliable'*) and then subtracting the total number of negative responses (*'not at all reliable'* and *'slightly reliable'*).

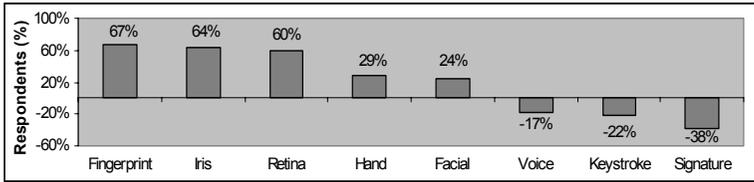


Figure 2: Ranking of the expected reliability of biometric systems

The results revealed a clear level of doubt in behavioural methods. For the case of keystroke an explanation for the negative result most probably is the minimum coverage of this method, while for the case of voice and signature the negative result reflects the fact that these methods are presented as the most easy to cheat. This view is actually reasonable, considering that the behavioural methods typically have worse False Acceptance Rates (FAR) and False Rejection Rates (FRR) than the physiological. It should be noted that no difference was observed between the respondents that have used biometrics and those that have not, indicating further the extent of the media influence.

3.5 User attitudes

One of the main objectives of the survey was to evaluate the users’ attitudes towards biometrics based upon what they have heard or read from the media. This has been achieved by asking the respondents to indicate how comfortable they would be to use the mainstream biometric technologies for proving their identity. A rank of user preferences can be seen in Figure 3.

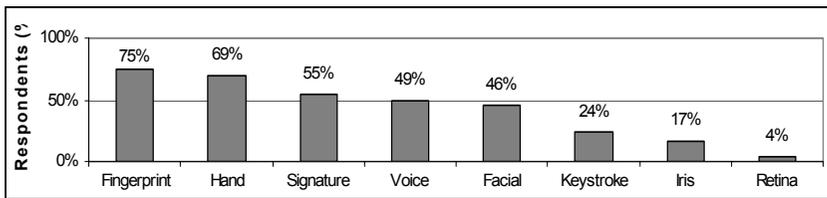


Figure 3: Ranked preference of biometric technologies

As expected, the most popular technology is the fingerprint recognition, despite the fact that it is known for its use by the police for identifying criminals. This indicates that the familiarity of the users with this method makes them extremely comfortable with the idea of using it. The above results are in contrast with the findings of the study carried by TNS / TRUSTe (2005), which found that 81% of the 1,003 American Internet users that participated consider as acceptable the fingerprint recognition, 58% the iris scan, 50% the hand geometry, 48% the voice recognition, and 44% the facial scan. This fact indicates that there are some demographical differences, which most probably are caused by cultural differences and from the different way that the media worldwide are covering the topic.

Maybe the most significant question posed in the survey asked the respondents to indicate how useful they find the use of biometrics in their various proposed applications. The ranked preferences can be seen in Table 1.

Proposed Applications of Biometrics	Percentage
1. Verify identity for passports and airport check-ins	75%
2. Check entry to government buildings	66%
3. Verify the identity of credit card holders	57%
4. Verify identity at ATMs for withdraw	56%
5. Looking for wanted criminals/terrorists at public events	53%
5. Verify identity of citizens (national ID cards)	53%
7. Check entry into schools and child services	48%
8. Check people for welfare fraud	38%
9. Verify identity for online transactions	30%
10. Verify voters during elections	28%
11. Verify identity for login to a PC/laptop/network	23%
12. Verify identity for telephone transactions	8%
13. Keep track of employee work hours	1%
14. Verify identity for using a cell phone	-7%

Table 1: Ranked preference of biometric applications

Observing the first ten preferences of the respondents reveals a pattern, which indicates that they find more useful the use of biometrics in applications relating to prevent terrorism and frauds. On the other hand the lower preferences of the users enhance the conclusion that the public prefer convenience rather than security.

Respondents were asked to indicate the time that they are willing to spend enrolling their characteristics, which then will be used by a biometric system to authenticate them. The majority (70%) of respondents are willing to spend up to half an hour, which nowadays can be sufficient in most cases for creating a profile. Once a user has registered his characteristic there is still a high possibility that the system will falsely reject him. To this end the respondents were asked to indicate the frequency with which they would be willing to tolerate such errors. The results showed that half of them are not willing to tolerate any errors, while 31% stated that they do not know. These results clearly indicate that biometric systems must have a low error rate, which is in accordance with the findings of Furnell *et al.* (2000).

Asking the respondents to indicate where their digital biometric characteristics (templates) should be stored, 40% of them stated that they prefer a central database and 25% that they do not mind. These results are quite surprisingly considering that the majority of the media express concerns for the risks associated by keeping the templates in a central database, once more indicating the users' preference to convenience.

When the respondents were asked to indicate whether they have concerns that their biometrics characteristics will be stolen with the purpose to cheat a biometric system, 56% of them (who in their majority do not believe that biometric systems can be easily cheated) stated that they are extremely or very concerned. This contradictory finding reveals the respondents' belief that criminals will find a way around this technology even though this will be very difficult.

Moreover, by asking the respondents to indicate how confident they are that their biometric characteristics will only be used for authentication purposes, a lack of confidence to both government agencies (57%) and organizations (50%) has been

revealed. This finding is in accordance with the results of TNS / TRUSTe (2005) study, where 64% of the respondents believed that the “potential for governments to misuse the information is too high”, reflecting the fear of Big Brother that the public has. However, 37% of the respondents are willing to lose some of their civil liberties for great security, as revealed by a later question. But more interesting is the fact that 22% of the respondents stated that they are not sure, revealing the confusion of the public which most probably is caused by their desire for greater security and the negative impact of the media.

Trying to further investigate the users' attitudes and the role of the media the respondents were asked to indicate their opinion as to the likely impact of biometrics in security. Surprisingly the results showed that the vast majority (76%) has realized that biometrics will only add another layer of security while only 5% of the respondents believe that they will stop terrorism and frauds.

Looking the future of biometrics the results revealed that two in three respondents believe that biometrics will be widely deployed by the end of the decade, while only one in fifty does not believe in their widespread use. This clearly indicates that the extended coverage of biometrics has prepared the public for their deployment. This conclusion is further enhanced by the fact that 70% of the respondents are willing to use biometrics despite their concerns that their characteristics can be either stolen or misused. However, most of the respondents are either not willing to pay (34%) or willing to pay less than £50 (37%) for devices with biometric capabilities. This finding indicates that cost is and will be an important barrier to the widespread use of biometrics. Moreover, it has been identified that 26% of respondents are willing to pay more than £50 if it is to protect an expensive device (e.g. a laptop above £850), indicating that they are more concerned to protect the device and not their personal data.

4. Conclusions

The survey has shown that today the public appreciates much more the importance of IT security, recognizing the need for stronger authentication mechanisms. But this is contradictory when considering that they do not even use (or at least use correctly) the current authentication methods, suggesting that user convenience is by far more important than the desire for security. This conclusion is further enhanced by the fact that the respondents are more accepting of those biometric methods that are easy to use (e.g. hand), even though they consider them to be less reliable than others (e.g. iris).

Moreover it has been identified that important factors for the acceptance of biometrics are that of awareness and practical experience. A strong relation has been revealed between those methods which the users are aware but most importantly have used and those which they feel more comfortable with the idea of using. Thus, it can be concluded that convenience and awareness (including practical experience) are essential, and that biometrics that the users will actually be willing to tolerate and use must be chosen for each application scenario.

5. References

- Black, J. (2002), “A Growing Body of Biometric Tech”, *BusinessWeek* [Online], http://www.businessweek.com/technology/content/jul2002/tc2002072_9892.htm, [Accessed 18/01/06]
- Citizenship and Immigration Canada (CIC) (2003), “Tracking public perceptions of biometrics” [Online], www.cic.gc.ca/english/press/03/poll-biometrics-e.pdf, [Accessed 18/01/03]
- Furnell, S.M., Dowland, P.S., Illingworth, H.M. and Reynolds P.L. (2000), “Authentication and Supervision: a Survey of Users Attitudes”, *Computers & Security*, Vol. 19, No. 6, pp 529-539
- Chowdhary, S. (2006), “Tech on the runway” [Online], *The Financial Express*, 16 January 2006, http://www.financialexpress.com/fe_full_story.php?content_id=114581, [Accessed 17/01/2006]
- Mogg, W.R. (2006), “Someone to watch over you”, *The Times* (London), 16 January 2006, pp 20
- TNS / TRUSTe (2005), “Consumer attitudes about biometrics in ID documents” [Online], http://www.truste.org/pdf/Biometrics_Study.pdf, [Accessed 14/01/06]
- Unisys (2006), “Consumers Worldwide Overwhelmingly Support Biometrics for Identity Verification, Says Unisys Study” [Online], http://www.unisys.com/about_unisys/news_a_events/04268651.htm [Accessed 21/07/06]
- United States Department of Justice (2006), “Identity Theft and Fraud” [Online], www.usdoj.gov, [Date Accessed 05/01/2006]
- Wait, P. (2003), Great expectations: Biometrics, *Washington Technology*, Vol. 18, No. 13
- Wood, L. (2006), “Feature - Good for nothing, Leanne Wood explains why she will refuse an ID card”, *Morning Star*, 9 January 2006

The Awareness and Perception of Spyware amongst Home PC Computer Users

M.Jaeger and N.L.Clarke

Network Research Group, University of Plymouth, Plymouth, UK
e-mail: info@network-research-group.org

Abstract

Spyware is seen currently as a major threat to personal computer based data confidentiality, whilst seen by criminal elements as a positive moneymaking device through personal data theft from security unconscious home internet users. This paper examines the level of understanding of anonymous home computer users to Spyware, their subjective understanding of it, the level of use of Anti-Spyware software, what level of threat they feel from Spyware, as well as if Spyware has changed their browsing habits. The methodology employed was an anonymous survey via email invitation and of those that accepted the invitation, 205 completed surveys were obtained. From the resulting survey analysis, most respondents do understand what Spyware is, yet there was found to be a lack of understanding of computer security in defending against Spyware, this being shown by 20% of survey respondents who did not use any Anti-Spyware. What was also found was a small proportion of participants who did not correctly understand what Spyware is. Additionally the subjective nature of survey respondent's ideas of Spyware infected websites was established, and compared to past web-crawl research from 2005, here a high proportion of survey respondents opinions was found to be objectively wrong. It was also found respondents see Spyware as a 'High/Some Threat', and due to past infections and news/media articles 72% have changed their browsing habits.

Keywords

Spyware, Anti-Spyware Software, Spyware Web-Crawl, Information Security

1. Introduction

In recent years there has been an emergence of software known as Spyware, programs created both for the covert and overt acquisition of personal or non-personal information from a personal computer (PC) connected to the internet, as stated by (Sariou *et al.*, 2004). This is now a persistent security problem to home PC users; the old security culprits of Malware (the viruses), are being superseded by Spyware, and to a lesser extent other newer forms of Malware e.g. key loggers and email phishing. Spyware programs are to an extent intelligent, able to hide via changing their signature, execute code without user intervention and move through networks invisibly whilst serving their own purpose.

Most research undertaken on Spyware is based on levels of Spyware infection on scanned desktop PC's, for example a survey (AOL/NCSA, 2005) by National Cyber Security Alliance (NCSA) in 2005 found from 354 respondents 38% had no Spyware protection, yet 83% felt safe from online threats; however 61% of the respondents had a form of Spyware or Adware on their PC. Equally an analysis report on

Spyware from (Webroot Software Inc, 2005) stated that during the course of the third quarter of 2005, an important and alarming Spyware trend emerged. Where, many home computer users are admittedly afraid of becoming a victim of identity theft from using the Internet. So some home PC users are worried and others are not.

This research aims to use past research into Spyware and infected website categories to ask home PC users which websites they think are the most likely to have Spyware. Original questions were asked of survey respondents, for example do respondents use Anti-Spyware and if so what vendor, do they understand what Spyware is and what level of threat do they feel from Spyware; respondents were also asked if they had changed their internet browsing habits due to past infections and or world news media articles. This papers format is to outline past Spyware research, followed by the methodology to investigate the attitudes of home PC users to Spyware. Specific aspects of research will be looked at; in terms of understanding what Spyware is, the use of Anti-Spyware software and the subjective belief of Spyware infestation in specified website categories. These results are then discussed and conclusions presented.

2. Background

Most research on Spyware has been completed on identification, categorisation and removal of Spyware; this new research looks at Spyware from the subjective viewpoint of the home PC user, as well as security measures employed. Past research has made little effort into understanding user awareness of Spyware and its activities; however some good research has been done.

This is shown by work produced by (Freeman and Urbaczewski, 2005) which states reasons why people hate Spyware. Whilst a paper by (Zhang, 2005) shows consumer understanding of Spyware in terms of their knowledge level. This is again revealed by a similar study by (Qing and Tamara, 2005), who believe in the concept of educating PC users to remove complacency that they have over Spyware, this research established user awareness factors were most accurate in showing which users took active measures against Spyware. Whilst a research paper created by (Awad and Fitzgerald, 2005) took a different viewpoint looking at what consumers find most deceptive about Spyware. Even though home PC users are worried about internet identity theft they do seem to have a love-hate aspect to Spyware, consumers will allow its usage on their own home PCs if there is a pay off i.e. in terms of peer-2-peer (P2P) software where the ability to download anything they want is assuaged by Spyware monitoring what they are doing, producing advertisement activation in the P2P software itself, these are seen as “Overt Providers” by (Warkentin *et al.*, 2005); this research also points to creating new legislation to segment Spyware software into positive and negative forms, with legal protection for some and prosecution for others.

As Spyware becomes a more prevalent infection vector, more research needs to be done into what home PC users actually think about Spyware in terms of their level of understanding, security implemented and awareness as to where they think Spyware is on the internet. Current Spyware articles point to a major problem in consumer

understanding of Spyware due to complacency, lack of knowledge or an inappropriate safe feeling whilst using the internet, consumers seem to have an ‘it will not happen to me’ attitude to Spyware infections.

3. Methodology

The method selected for this avenue of research was an online anonymous survey. The analysis of the survey data was via Qualitative analysis; this analysis form was chosen as it shows a complete, detailed depiction of collected data; whilst interpreting data distinctions by using logical human deductions. Creation of the survey was done over a number of versions to produce the correct wording, syntax and structural flow. Once the final version was decided upon, the collection process was set out over a series of months; this period lasted from 03/03/06 to 26/07/06. Data collection was via anonymous survey response, allowing for more candid question answering. Survey invitation was via the use of email (containing the survey web link and basic information about the research); this was then sent to as many people as possible.

The chosen design approach was not to use normal demographics e.g. Age, Gender, Education. The reason for this is that computer skills in terms of ability are not dependent on any of the above demographic traits. So the base gauge decided upon was the respondent’s subjective understanding of their own computer skill level. The only problem with this is the subjective nature of under or overestimating ones own ability. As such four catch all levels were created: Novice home PC user, Intermediate home PC user, Advanced home PC user and IT Professional. A skill set example was provided for each level, each level encompassed a series of computer skills which increased in knowledge and complexity dependent on the respondents understanding of computing. The survey was partitioned into 6 sections; grouping together similar avenues of questioning. The sections are as follows: PC Usage, PC Setup/Configuration, Experiences of and with Spyware, Understanding of Spyware, Past Spyware Infections and Other Known Internet Security Threats.

4. Results

An opening question was asked of the 205 survey respondents, ‘what is your home computer used for’ e.g. Email and Online Shopping; results are shown in Figure 1.

An interesting piece of data from Figure 1 points to 50% of respondents using P2P software, P2P software is a known harbour for both Spyware and other forms of Malware. As such these respondents must be tacitly accepting possible infections as a consequence of the possible positive P2P software usage.

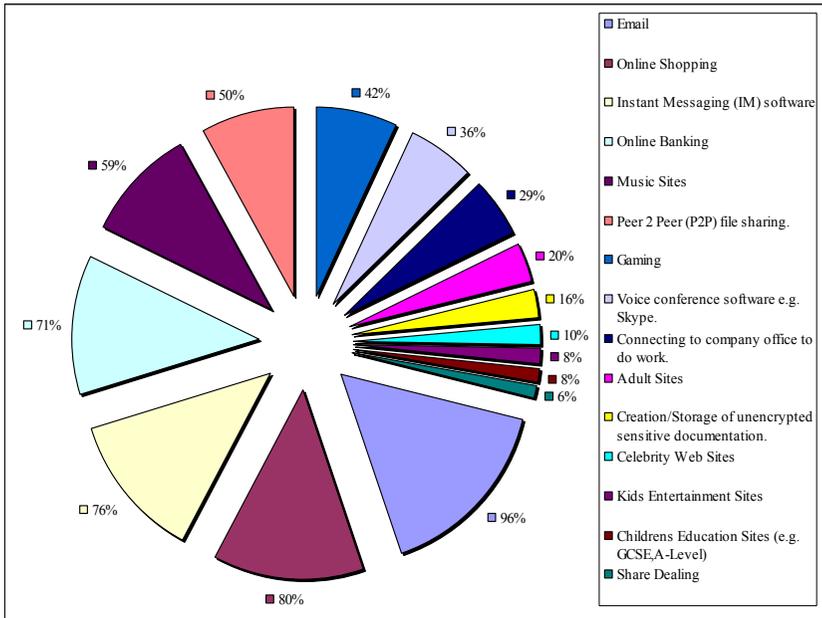


Figure 8: What Is Your Home Computer Used For?

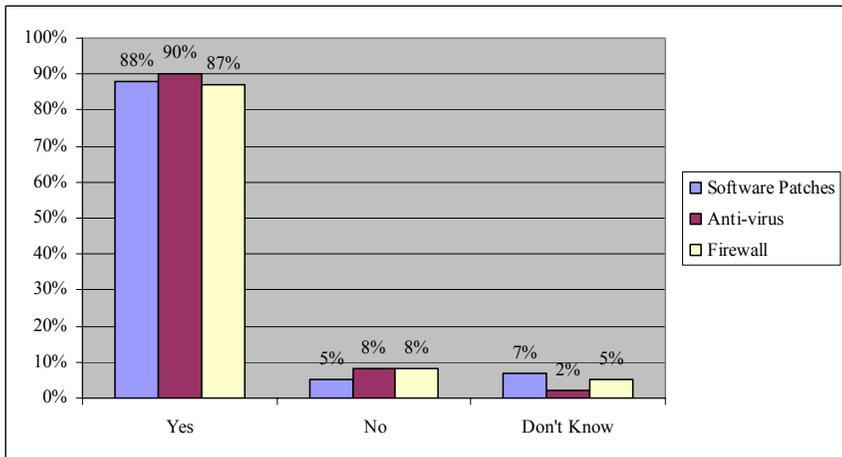


Figure 9: % of Home Computer Users Using Anti-Virus, Firewalls and Software Patching.

This was followed by a question about whether they software patched their home computer, used a Firewall and used Anti-Virus software; these results are shown in Figure 2. What can be seen is that most respondents do use these security controls, however a small proportion do not, whilst some ‘Do not Know’ if they do or do not, this data can also be looked at alongside how many respondents use Anti-Spyware in Figure 3.

	Total	Have you heard of (come across the phrase) “Spyware” before?	
		Yes	No
Total	205	196	9
		96%	4%

Table 1: What Do You Feel Is Your Skill Level With Your Home PC?

Of the 196 respondents who had heard of Spyware, as shown in Table 1, only 82% of these respondents picked the most widely decided definition of what Spyware is (as shown in Table 2), that being ‘Spyware gathers information about what I am looking at on the internet and sends it back to a central computer.’

	Total	Novice Home PC User	Intermediate Home PC User	Advanced Home PC User	IT Pro
Total	196	2	68	75	51
Spyware is a nasty computer virus.	9	1	7	1	0
		50%	10%	1%	0%
Spyware sends me to annoying web sites I do not want to go to.	7	0	4	1	2
		0%	6%	1%	4%
Spyware gathers information about what I am looking at on the internet and sends it back to a central computer.	161	1	44	70	46
		50%	65%	94%	90%
Spyware is a form of annoying popup advertising that appears when I go to certain web sites.	19	0	13	3	3
		0%	19%	4%	6%

Table 2: Which Phrase Best Describes What “Spyware” Could Be?

Conversely for the respondents that had heard of Spyware, Table 2 shows the survey data details. As such the user groups who understood what Spyware is were unsurprisingly the Advanced Users and IT Professionals; however there does seem to be some misunderstanding of what Spyware is. From the data 18% (35 respondents out of the 196 respondents that know what Spyware is) misidentified Spyware as either a ‘Virus’ – 9 respondents, or ‘Adware’ – 7 respondents, or finally a ‘Pop-Up’ – 19 respondents. Survey respondents in general do seem to know from this survey what Spyware is, as can be seen in Table 2, where 79% or 161 respondents out of the total of 205 respondents correctly knew what Spyware is.

As can be seen from Table 3 of those who had heard of Spyware (as found in Table 1), a high proportion have had a Spyware infection, this being 64% of respondents; this is much more than those that have not had a Spyware infection. Of those that had acquired a Spyware infection most were able to remove the problem themselves, as shown by Table 3 where 53% of all survey respondents used their own Anti-Spyware

to remove the infection from their PC. Only a small proportion required 3rd party customer support help, this being 3% or 6 respondents. Furthermore only 8% of survey respondents had to re-image their home computer.

	Total	Removed it myself with my Home PC's Anti-Spyware software.	Required my home PC's customer support to help me remove Spyware using my home PC's Anti-Spyware software.	Called my home PC's customer support with problem, told to download Anti-Spyware, then walked through how to use it.	Unable to remove it from home PC, had to reinstall all software including operating system.
Total	196	104	3	3	15
No, I have never had a "Spyware" infection.	71	0	0	0	0
		0%	0%	0%	0%
Yes, up to 3 Months ago.	65	54	2	3	6
		52%	67%	100%	40%
Yes, up to 6 Months ago.	16	14	0	0	2
		14%	0%	0%	13%
Yes, up to 9 Months ago.	8	6	0	0	2
		6%	0%	0%	13%
Yes, up to 1 year ago.	16	14	0	0	2
		14%	0%	0%	13%
Yes, 1+ year(s) ago.	20	16	1	0	3
		15%	33%	0%	20%

Table 3: Past Spyware Infections on Respondents Home PC's and Resolution Methods Used.

In terms of defending themselves against Spyware, of the Anti-Spyware software used there is a clear winner as shown in Figure 3; furthermore it seems many consumers use multiple Anti-Spyware software, as the total of Anti-Spyware software used - 215, is larger than the total survey response of 205. Spybot comes top with 36% of respondents using the software, its high usage is probably down to being freeware, as well as having been around the longest (since at least October 2002), it is also updated regularly and is known to have a good scan engine. Next is Lavasofts Ad-Aware, again this software has a freeware version that is regularly updated but has not been around for as long as Spybot, it is also known to have a good scan engine. Next is the 20% of respondents who do not use any Anti-Spyware software at all, this equates to 39 respondents; comparing this answer to survey security questions on Anti-Virus, Firewall usage and security patching we find the

following respondents who do not use any Anti-Spyware software; where 13 respondents do not use Anti-Virus, 6 respondents do not use a Firewall and 3 respondents do not security patch the Operating System of their home PC. This information points to the glaring problem of why Spyware can spread quickly in this age of Broadband communications; the problem is that the onus is on the PC owner. With that comes a disparate level of computer security knowledge.

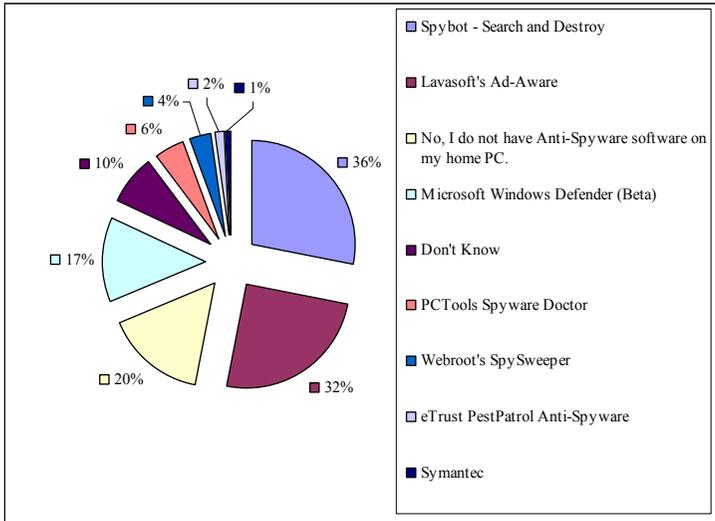


Figure 10: % Breakdown Of All Anti-Spyware Software Used On Each Respondents Home PC.

Subsequently, a question was asked in terms of the type of sites that could contain Spyware, the web site classifications covered most of the main sites people go to and the results can be seen in Table 4. From the raw data shown in Table 4 it can be seen that Adult Oriented sites and Pirate Software site are what respondents think contain the most Spyware; with the least likely being Online News sites followed by Kids oriented websites.

	Definitely/Yes	Possibly/Maybe	No	Do not Know
Adult Entertainment	72%	24%	1%	3%
Pirate Software/Warez	70%	21%	3%	6%
Screensaver/Wallpaper	37%	42%	10%	10%
Music Orientated	23%	61%	8%	8%
Games Oriented	22%	60%	11%	7%
Celebrity Oriented	18%	61%	8%	13%
Kids Orientated	11%	43%	32%	14%
Online News	5%	33%	45%	17%

Table 4: Home PC Users Opinion of Website Categories That Contain Spyware.

In terms of other forms of Malware infections, respondents were asked if they had received a Phishing email, just under half of respondents - 48%, had received a phishing email trying to obtain personal data, whilst 52% had not. From this avenue of questioning respondents were also asked as to what other Malware infections/devices they had been infected by, these are seen in Figure 4.

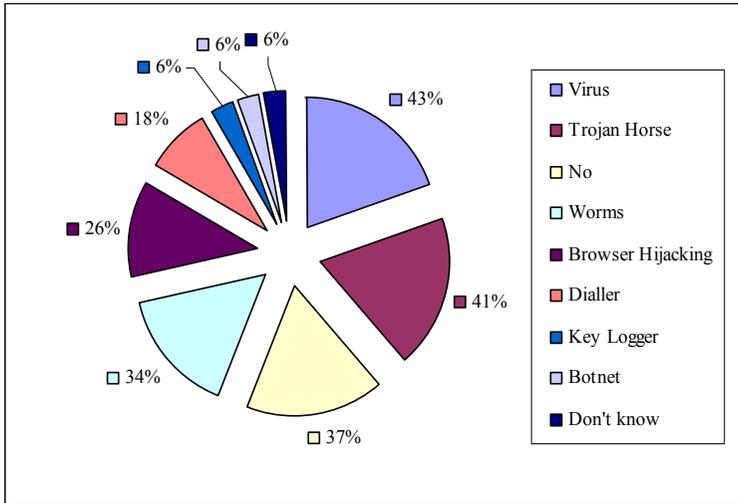


Figure 11: Other Malware Infections as % of Respondents.

What can be seen in Figure 4 is that the old sources of Malware infection are still evident: the Viruses, Trojan Horses and Worms. Even as more recent forms of infection vector are becoming more plentiful. This can be seen by the level of Browser Hijacker, Key Logger and BotNet infection. Additionally Dialler infections will probably reduce over time as Broadband becomes the mainstream, with less and less narrowband internet connections used.

	High Threat	Some Threat	Very little Threat	No Threat	Do not Know
Virus	56%	35%	8%	1%	1%
Worm	54%	28%	11%	1%	6%
Trojan Horse	55%	26%	9%	3%	7%
BotNet	27%	26%	18%	5%	24%
Key Logger	37%	23%	17%	6%	17%
Dialler	21%	27%	20%	15%	17%
Browser Hijacking	26%	31%	21%	7%	14%
Email Phishing	31%	28%	20%	13%	8%
Spyware	35%	40%	17%	3%	5%

Table 5: Threat Levels As Seen By Survey Respondents.

The survey respondents were then asked as to how much of a threat Spyware was to them on a scale from ‘High Threat’ to ‘No Threat’, the results are seen from Table 5. As can be seen the highest response to Spyware by 40% of home PC users was as

‘Some Threat’, the next highest response was from 35% of respondents who saw Spyware as a ‘High Threat’. This indicates that Spyware is seen as a serious threat by 75% of survey respondents.

On your home PC have you become more careful about what web sites you visit on the internet generally due to any of the below reasons?	
Yes, due to past infection/attack.	33%
Yes, due to information from 3rd party (e.g. TV News, Newspaper, etc.).	22%
Yes, due to past infection/attack and 3rd party information.	17%
No, I have not become more cautious.	28%

Table 6: Have You Changed Your Website Browsing Habits Due To A Specific Reason?

Furthermore, as can be seen from Table 6 a large proportion of the survey respondents have changed their website browsing habits, mainly due to past internet based infections and attacks i.e. Spyware. In total the respondents that changed their browsing pattern total is equal to 72% or 148 respondents.

5. Discussion and Evaluation

The results have shown that almost all respondents understand correctly what Spyware is, and of all the respondents a healthy amount use Anti-Spyware. Though from past history a high proportion of home PC users have had a Spyware infection, more than double those that have not. Furthermore most of these past infected respondents were able to remove the Spyware themselves with little outside help. This indicates that Anti-Spyware is becoming easy to use and effective. What is also of interest is that though software is seemingly able to remove Spyware, consumers seem to be doubling up with one to two Anti-Spyware software programs. It seems they may not be totally confident in the software yet, this may be due to the recent creation of most Anti-Spyware e.g. Microsoft’s Anti-Spyware software in January 2005. Only 39 respondents do not actively use Anti-Spyware, these respondents are not protected from Spyware and possibly will infect other people; they can be seen as a threat to the personal data security of others who do not use Anti-Spyware. However, Anti-Spyware users must keep their software regularly updated, as the Anti-Spyware users are only as well protected as their software is regularly updated.

Nevertheless it seems that even a proportion of the most allegedly knowledgeable of respondents, the ‘Advanced Users’ and ‘IT Professionals’, seem to not know correctly what Spyware is, as such there does seem to be a knowledge gap in terms of computer security and Spyware knowledge that needs to be bridged for a small proportion of respondents; and due to this lack of correct knowledge and possible scare mongering by the news media a high proportion of respondents have changed their internet browsing habits. This possibly causes a fear of unknown internet websites, as users may possibly get a Spyware/Malware infection from accessing them. Furthermore this fear, coupled with possible news media scare mongering, and accompanied by a lack of computer security knowledge to remove possible

infections (in case of detrimental damage to their home PC), may lead to increased levels of Spyware and Malware infection.

Of tremendous interest from the survey data, is the high proportion of respondents who use P2P software (a known Spyware infection route). Does this mean that Spyware monitoring in P2P software is ok for some consumers, where this type of Spyware used is seen as an “overt provider” by (Warkentin *et al.*, 2005). Does this then mean as (Warkentin *et al.*, 2005) states that some consumers see some Spyware as ok, as long as they get some kind of positive benefit from losing some of their online privacy. If this is the case does this not mean that not all Spyware is as bad as we think, as (Warkentin *et al.*, 2005) state in their research, legislation does not sufficiently cover the differential nature of Spyware, and that a comprehensive categorisation of Spyware must be undertaken to produce a far more democratic legislation process. This is needed, as can be seen in the grey area of client/server management agents like SNMP being used on employee’s corporate computers. The agent software could possibly be classed akin to Spyware dependent on how it is used, as the likelihood is the employee will not know about it, and that data may or may not be used remotely in the company by other departments.

In terms of evaluating the subjective Spyware infested website comparison research, the basis of the data analysis on Table 4 is from the research paper by (Moshchuk *et al.*, 2006), which was a web-crawl during 2005 across a number of website categories looking for Spyware. Figure 2 is a graphical explanation of Table 3 from (Moshchuk *et al.*, 2006) research, showing the most Spyware infested website category domains, in comparison survey respondents were asked to subjectively gauge which had the most Spyware. From the graphical extrapolation in Figure 2, the Top 4 website domains containing Spyware by percentage can be seen in descending order: Games websites (20%), Music websites (11.40%), Wallpaper/Screensaver websites (9.60%) and Celebrity websites (7.60%).

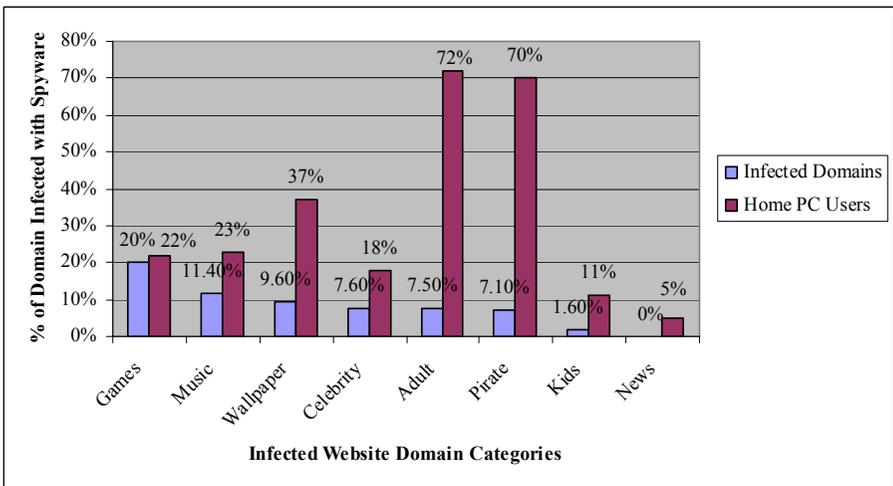


Figure 2: Executable Spyware Infections across Web Categories.

From the survey results in Table 4 and Figure 2, the respondents understanding of which websites categories that contain the most Spyware can be gauged. Furthermore it can be seen if knowledge is the basis of their assumptions or website prejudice. For the home PC users they subjectively felt that the following Top 4 website categories, in decreasing order, as shown previously in Table 4 have the most inherent Spyware: Adult Entertainment websites (72%), Pirate Software websites (70%), Wallpaper/Screensaver websites (37%) and Music Orientated websites (23%).

	Games	Music	Wallpaper	Celebrity	Adult	Pirate	Kids	News
Infected Domains	20%	11.40%	9.60%	7.60%	7.50%	7.10%	1.60%	0%
Home PC Users	22%	23%	37%	18%	72%	70%	11%	5%

Table 4 : Raw Data Showing Executable Spyware Infections across Web Categories.

Essentially the home PC users were mainly wrong about the Top 4 Spyware infested website categories. From this evaluation of respondent's subjective assumptions it can be safely said that survey respondents were wrong, and their assumptions are probably down to preconceived prejudices against certain website categories. Though respondents may not know where Spyware is, they do feel that it is a threat, as survey results point out that the majority of respondents see Spyware as a 'High/Some Threat'. This is a feasible reason why a high proportion of respondents have changed their internet browsing habits, as a consequence of primarily past infections/attacks from for example Spyware; and secondly, news/media articles on the threat of Spyware and the associated loss of personal data from their personal computer.

6. Conclusions

Even though it seems a high proportion of home PC users understand the need for security software i.e. Anti-Spyware, and information is being understood in terms of this threat, there is still a problem of low computer security knowledge in a proportion of respondents, as stated by both (Qing and Tamara, 2005; Zhang, 2005); it creates user reservations in doing anything about possible current or future Spyware infections. Conversely some respondents use multiple Anti-Spyware software from different vendors, seemingly showing a current low level of trust in current software. However most only use one vendor, with a small proportion using none; however the use of Anti-Spyware is moot, unless it is kept up to date with scanning engine/signature patches. Nonetheless most who have had a past infection were able to remove it themselves with their own Anti-Spyware, those that needed 3rd party help were also able to remove their Spyware infection, only a small proportion were unable to and had to re-image their PC. This end user ability points to current software being easy to use and effective against current Spyware. Furthermore a large proportion of respondents do seem to understand what Spyware is in terms of a current definition, whilst seeing Spyware as a 'High/Some Threat'. This correct understanding of Spyware is however at odds with their incorrect judgment of which website category's contain Spyware; their judgments here are down to preconceived website category prejudices. This coupled with past Spyware

infections and news media articles has probably changed survey respondent's internet browsing habits, making them more cautious in terms of what they look at. This 'hit-and-miss' approach to computer security knowledge is probably down to a lack of access to simple and good computer security education, jointly associated to possible apathy to understand what is already there. Somehow home PC users must be taught, or re-taught how to reduce their security risks via updating their operating system with security patches, as well as how to correctly configure, patch and use their security software i.e. Anti-Spyware software. Conversely legislation must be looked at accurately to remove the potential 'grey area' of Spyware usage. As it can be argued that Spyware can be in some part positive to home PC users and at other times negative as (Warkentin *et al.*, 2005) explains. This Spyware segmentation must be created to remove possible legal prosecution of possible positive 'grey area' Spyware software.

7. References

- AOL/NCSA (2005). *AOL/National Cyber Security Alliance (NCSA) online safety study*. National Cyber Security Alliance., pp1-11.
- Awad, N. F. and Fitzgerald, K. (2005). The deceptive behaviors that offend us most about spyware. *Communications of the ACM*, 48, (8) pp55-60.
- Freeman, L. A. and Urbaczewski, A. (2005). Why do people hate Spyware? *Communications of the ACM* 48, (8) 50-53.
- Moshchuk, A., Bragin, T., Gribble, S. D. and Levy, M. H. (2006). *A crawler-based study of Spyware on the web*. Proceedings of the 13th Annual Network and Distributed System Security Symposium (NDSS 2006). February 2006. p17.
- Qing, H. and Tamara, D. (2005). Is Spyware an internet nuisance or public menace? *Communications of the ACM*, 48, (8) pp61-66.
- Sariou, S., Gribble, S. D. and Levy, H. M. (2004). *Measurement and analysis of Spyware in a university environment*. ACM/USENIX Symposium on Networked Systems Design and Implementation. San Francisco, CA.
- Warkentin, M., Luo, X. and Templeton, G. F. (2005). A framework for spyware assessment. *Communications of the ACM*, 48, (8) pp79-84.
- Webroot Software Inc (2005). *State of Spyware Q3 2005*. Webroot Software Inc., pp1-91.
- Zhang, X. (2005). What do consumers really know about Spyware? *Communications of the ACM* 48, (8) pp44-48.

Investigate Placement of Relaying Node in Wireless Mesh Networks

D.Jing and X.Wang

Network Research Group, University of Plymouth, Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

Gupta and Kumar (2000) argued the throughput available per node decreases as the number of node increases in ad hoc networks. Therefore, the approach of adding relaying node was proposed to improve the network throughput which in turn improves the throughput available per node. The focus of this paper is to discover the functional area and determine the optimized position of relaying node when placed between two mesh routers in WMNs. Simple source to destination model and simple relaying node model were designed for analyzing the network capacity based on the computation. The main result is achieved by analyzing the capacity of network model and simulating network model with OPNET Modeler. The result shows the relaying node could operate to forward packet to receiver when the relaying node is deployed in the range of both transmitter and receiver when the receiver is deployed out of the range of the transmitter. It also shows that in the relaying node functional area, the maximum capacity could be achieved by deploying relaying node in the middle between transmitter and receiver.

Keywords

Wireless Mesh Networks, Relaying Node, Node placement

1. Introduction

Wireless mesh networks (WMNs) is becoming more and more popular in community wireless networks. Many advantages come with WMNs, such as low installation cost, easy network maintenance, robustness, reliable service etc. However, along with the increase of network size, the network capacity decreases dramatically. Roughly, the average available throughput per node decreases as the square root of the number of the nodes, in a static ad hoc networks(Gupta and Kumar, 2000).

This paper addresses the issue on the placement of relaying node when the approach of adding relaying node is adopted to improve the network capacity which in turn increases the capacity throughput available per node. Specifically, the aims of this paper are to discover the functional area of relaying node between transmitter and receiver, and find the optimized position of relaying node to achieve maximum network capacity. Functional area of relaying node is defined as the area that relaying node could operate to forward the packet received from transmitter to receiver. The optimized position of relaying node is defined as the position relaying node placed to achieve the maximum capacity.

2. Background

According to Gupta and Kumar (2000), the average available throughput per node decrease as the square root of the number of the nodes in ad hoc networks. Three approaches were suggested to improve the total network capacity which in turn improves the capacity throughput available per node, which are adding relaying node, adding mobility and grouping node into cluster respectively (Gupta and Kumar, 2000). This paper specifies the direction on adding relaying node to improve the network capacity which in turn improves the capacity throughput per node.

3. Methodology

Due to the similarity of WMNs and ad hoc networks, the capacity analysis result for ad hoc network can be used to analyze the network capacity in WMNs. A detailed literature review was carried out.

The network models with specific scheme were designed firstly for analyzing the network for specific aim. The MATLAB tool was used to plot the curve expressing the relationship between transmission distance and capacity for analyzing the network capacity. The simulation scenario was designed based on the network model analyzed. The result of simulation with OPNET Modeler was used to verify the result getting from the analysis.

Finally, the network capacity achieved from simulation was compared with network capacity result from analysis. The final result was drawn and discussed.

4. Network Capacity Analysis

The network model capacity analysis is carried out based on the network capacity computation, network model design, and the assumption.

4.1 Network Capacity Computation

Firstly, the wireless channel capacity is calculated by Shannon's formula which is given by (Goldsmith, 2005):

$$C = W \log_2(1 + P/N) \quad (2)$$

Where W denotes the channel bandwidth, P/N donates the signal to noise ratio. The P represents received power which can be calculated based on the free space propagation model and is given by (Goldsmith, 2005):

$$P_r(d) = P_t \left(\frac{\lambda}{4\pi d} \right)^2 \quad (3)$$

Where λ is the wavelength which can be calculated by c/f. P_t is the transmission power. N is the noise power which depends on the thermal noise, and is given by (Goldsmith, 2005):

$$N = T_{bk} B_{rx} K \quad (4)$$

Where T_{bk} is the background temperature and B_{rx} is the channel bandwidth. K is the Boltzmann's constant which is $1.379E-23$.

4.2 Assumption

IEEE 802.11b radio standard is assumed to be used as the radio for mesh backbone. Wireless IEEE 802.11b radio generates an unlicensed ISM 2.4 band radio signal with DSSS modulation and defines several raw data rates, 1, 2, 5.5 and 11 Mbps respectively. The transmission power of each node, which is regulated by FCC, is assume to be same. Furthermore, the network delay which caused by relaying node decode-and-retransmit mechanism will not considered in network capacity computation.

4.3 Network Model Design

In order to analyze the network model capacity, two models simple source to destination model and simple relaying node model are designed based on the aims of the project.

4.3.1 Simple Source to Destination Model

Simple source to destination model contains a transmitter i and a receiver k . This model is designed to discover the functional area of relaying node between transmitter and receiver. The functional area is defined as the relaying node could operate to forward the packet to receiver. The basic structure of simple source to destination model is shown in Figure 1.

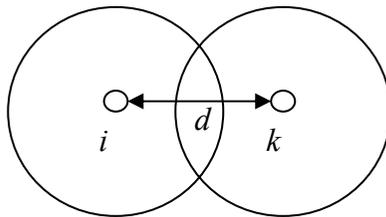


Figure 1: Simple Source to Destination Model

The network capacity of this model, which can be calculated based on Shannon's formula, free space path loss model and assumptions, is given by:

$$C = W \log_2 \left(1 + \frac{P_t (\lambda / 4\pi d)^2}{T_{rx} B_{rx} K} \right) \tag{5}$$

4.3.2 Simple Relaying Node Model

The simple relaying node model included a source node i , a destination node k which is deployed out range of source node, and a relaying node j which is deployed in the range of both transmitter and receiver. This model is designed for achieving the optimized position of relaying node between transmitter and receiver. The maximum

capacity would be achieved by deploying optimized position of relaying node in its functional area. Figure 2 illustrates the simple relaying node model structure.

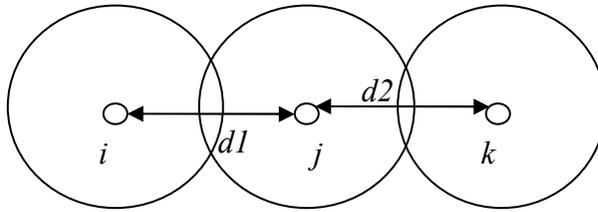


Figure 2: Simple Relaying Node Model

The network capacity of this model could be expressed by:

$$C = \text{Min}(C_{ij}, C_{jk}) \quad (6)$$

Where C_{ij} and C_{jk} can be calculated based on the capacity calculation in simple source to destination model.

4.4 Functional Area of Relaying Node Analysis

In order to find functional area of relaying node, the capacity achieved from simple source to destination model is treated as a reference capacity. The relaying node was deployed in the middle and in the range of both transmitter and receiver in simple source to destination model. Then, the reference capacity is compared with capacity achieved from the simple source to destination model with relaying node. For clarity, two transmission distance schemes are defined, which are short and long transmission distance scheme. Short transmission distance scheme is defined as the receiver is deployed in the range of transmitter. Long transmission distance is defined as the receiver is deployed out of range of transmitter and the relaying node is deployed in the range of both transmitter and receiver. In order to achieve and compare the relationship between transmission distance and capacity in the simple source to destination model and the model adding relaying node, following parameters are used (Table 1). The network capacity for simple source to destination model and model adding relaying node can be calculated using Shannon's formula, free space propagation model and noise computation (see equation 5 and 6) with parameters defined.

Parameters	Values	Parameters	Values
Pt	0.01 watts	f(central)	2.412E+9
Brx/W	22 MHz	K	1.379E-23
C	3.0E+8	Trx	290K
d	200m		

Table 1: Parameters for Capacity Calculation

The capacity comparison result achieved using MATLAB is shown in Figure 4-3.

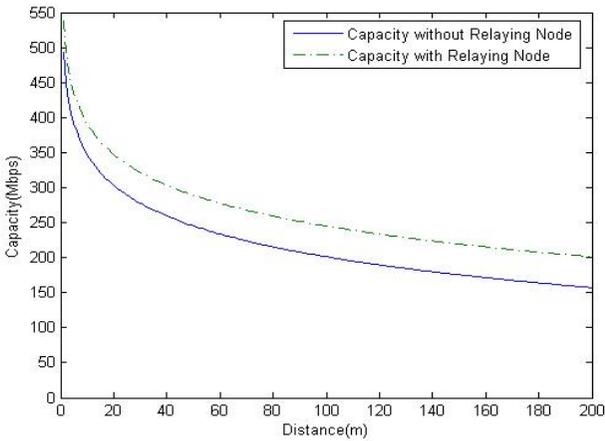


Figure 3: Network Capacity Comparison Result

It shows that adding relaying node could improve the network capacity when the relaying node is deployed in the middle and the receiver is deployed in the range of transmitter. Moreover, network with relaying node could improve the network capacity with both two transmission distance schemes. Therefore, conclusion for functional area of relaying node is drawn that the relaying node could be functional when relaying node is deployed in the range of both transmitter and receiver, and the receiver is deployed out of the range of transmitter.

4.5 Optimized Position of Relaying Node Analysis

In order to discover the optimized position of relaying node in its functional area, three relaying node position schemes are defined, which are relaying node close to source, relaying node in the middle and relaying node close to destination scheme. Relaying node close to source scheme is defined as the relaying node is deployed close to source and it still in the range of both transmitter and receiver. Relaying node in the middle scheme is defined as the relaying node is deployed in the middle between transmitter and receiver and it still in the range of both transmitter and receiver. Relaying node close to destination scheme is defined as the relaying node is deployed close to destination and it still in the range of both transmitter and receiver. The network capacity for simple relaying node model with three schemes is analyzed based on the Shannon’s formula; path loss model, assumptions, and analysis (see equation 5 and 6). The parameters in the table 2 are defined to calculate the network capacity.

Parameters	Values	Parameters	Values
Pt	0.01 watts	F(central)	2.412E+9
Brx/W	22 MHz	K	1.379E-23
C	3.0E+8	Trx	290K
D	300m		

Table 2: Parameters for Capacity Calculation

The transmission power which equals to 0.01 watts means that transmission range limit of each node is 199m based on the free space propagation loss model. Moreover, the distance between transmitter and receiver is 300m. Therefore, the relaying node should be deployed between 101m and 199m. The middle point between transmitter and receiver is 150m. The variable curve of network capacity which can be achieved based on the equation (equation 5 and 6) and the parameter defined is plotted in figure 4:

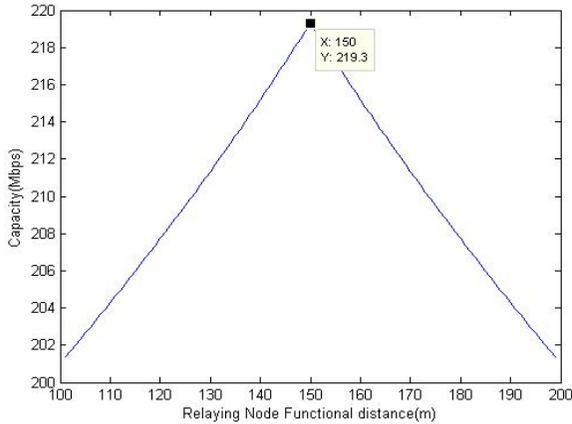


Figure 4: The capacity curve in relaying node functional area

As shown in above figure, in the relaying node functional area, the maximum network capacity is achieved by deploying the relaying node in the middle between transmitter and receiver. Therefore, it implies that the optimized position of relaying node is in the middle between transmitter and receiver when the relaying node is deployed in the range of both transmitter and receiver and the receiver is deployed out of range of transmitter.

It is noticeable that the capacity achieved based on the Shannon's formula is relatively larger than the maximum data rate, 11 Mbps, for IEEE 802.11b. Previous research (Wilson, 2004) which achieved similar results proves that the capacity achieved based on the Shannon's formula in the analysis is reasonable. The difference between data rate and maximum theoretical capacity is mostly due to channel impairments that are not incorporated into the theoretical model, for example, the relatively inefficient signaling methods used on today's wireless channels (Communication Systems, 2006).

5 Simulation Results

To validate the results, several network scenarios were simulated using the OPNET Modeler. For MAC layer, the 802.11b with RTS/CTS as it is well modeled in OPNET. The MANET traffic generation with OPNET to generate constant bit rate (CBR) User Datagram Protocol (UDP) flows.

5.1 Functional area of Relaying Node

Two scenarios were considered to discover the functional area of relaying node based on the simple source and destination model and the analysis. In the first scenario, the simple source to destination model was deployed with short transmission distance scheme and then, the relaying node was deployed in the middle between transmitter and receiver. In the second scenario, the simple source to destination model was deployed with long transmission distance scheme, and the relaying node was deployed in the middle between transmitter and receiver. The simulation result is shown in Figure 5.



**Figure 5: Left-Network Capacity Comparison with Short Distance Scheme
Right-Network Capacity Comparison with Long Distance Scheme**

The result shows that in network with short transmission distance scheme, adding relaying node could not improve the network capacity, which is different from the result achieved from analysis. The reason is that with RTS/CTS used in MAC layer, the radio signal transmits the signal directly to receiver without passing through the relaying node. Thus, the relaying node cannot operate to forward packet to receiver. Moreover, it also shows that with long transmission distance scheme, adding relaying node could improve the network capacity. Thus, the functional area is determined, that the relaying node is deployed in the range of both transmitter and receiver, and when the receiver is deployed out of the range of transmitter.

5.2 Optimized Position of Relaying Node

Three scenarios were designed to discover the optimized position of relaying node based on the simple relaying node model with three relaying node position schemes. The optimized position of relaying node could be discovered by comparing the network capacity for each scenario and achieve the maximum network capacity. The network capacity comparison for three relaying node position schemes is shown in Figure 6(Left). For clarity, another way to visualize the network comparison is given in Figure 6(Right), where the cumulative distributive function (CDF) of network throughput comparison is plotted.

The result of network capacity comparison with simulation verifies the analysis result. Although the capacity difference between each scheme is small, it shows that in the relaying node functional area, the maximum throughput achieved when the relaying node is placed in the middle between transmitter and the receiver. It also

shows that when relaying node is deployed in the functional area, it could improve the network capacity.

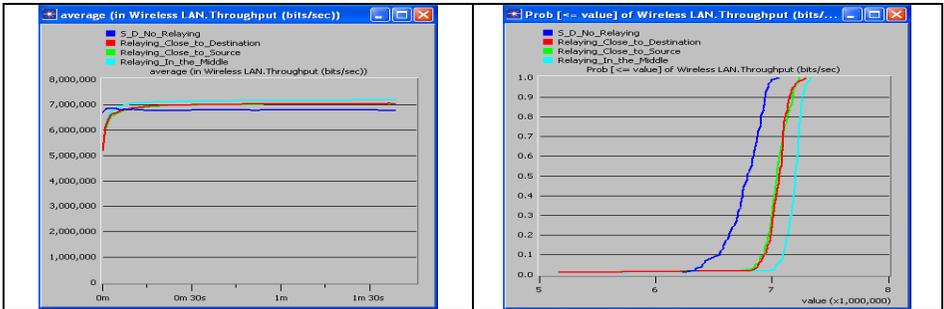


Figure 6: Left-Network Capacity Comparison for three relaying node position scheme. Right-Cumulative Distributive Function (CDF) Visualization

It is noticeable that there is a small difference between actual throughput and the source data rate traffic setup. Specifically, the source data rate was set as 11 Mbps based on the radio standard used, IEEE 802.11b, in this project; however, the throughput achieved is only around 7 Mbps. This difference is most due to the timing and framing information added which degrade the data rate (Communication Systems, 2006). Moreover, the inter-symbol interference (ISI) also would limit the throughput achieved.

6 Conclusion

The research question in this paper is specified on the functional area and optimized position of relaying node. The result shows the relaying node could operate to forward packet to receiver when the relaying node is deployed in the range of both transmitter and receiver and the receiver is deployed out of the range of the transmitter. It also shows that in relaying node functional area, the maximum capacity could be achieved by deploying relaying node in the middle between transmitter and receiver. This study involves the relaying node placement in single hop wireless network. The future work would focus on the placement of relaying node in multi hop wireless mesh networks.

7 References

Communication Systems (2006) “Basic Communications and Networks” <http://www.cs.ucl.ac.uk/staff/S.Bhatti/D51-notes/node18.html>, (Accessed 16th August 2006).

Goldsmith, A. (2005), *Wireless Communications*, New York, Cambridge University Press, ISBN: 978-0-521-83716-3

Gupta, P. & Kumar, P. R. (2000), “The Capacity of Wireless Networks”, *IEEE Transactions on Information Theory*, Vol.46, pp388-404.

Wilson, J. M. (2004). “The Next Generation of Wireless LAN Emerges with 802.11n”, *Technology@Intel Magazine*, 8: pp1-8.

Evaluation of Pervasive and Ubiquitous Healthcare Systems

K.A.Khan and X.Wang

Network Research Group, University of Plymouth, Plymouth, UK
e-mail: info@network-research-group.org

Abstract

Over the past few years wireless communication technology has changed people's lives with the newest products. This research paper is presenting an in-depth evaluation and investigation of Pervasive and Ubiquitous healthcare system in light of existing wireless sensor healthcare systems and their wireless devices. This investigation is also encompassing on most recent wireless sensor networks, wireless devices deployment issues, which involves in modern safety critical healthcare of patients. This research work presents future suggestions and strategies based on careful analysis and evaluation of existing healthcare system with wireless sensor networks scenarios. This research paper contains a brief discussion of wireless healthcare issues, which are much concern about the patient awareness of this modern healthcare technology through an analytical survey results. These survey results are obtained through online questionnaires form which acquire the patient's social, quality, safety and cost/finance problems during collaboration with wireless healthcare systems. Overall this research paper has future recommendation and open a new era for the researcher to overcome the existing wireless healthcare problems and improve the workability of the wireless devices in healthcare systems.

Keywords

Pervasive Wireless Networks, Healthcare Systems.

1. Introduction

Public health and safety is the major part of modern healthcare system and its work is to provide the safety and entire services to the patients when and where needed especially in case of emergencies. Most of the healthcare systems are trying to do their best in providing the maximum facility to the patients. But after the publishing a report from Institute of Medicine of the National Academy of Sciences, which claiming that between 44,000 and 98,000 American die each year as a result of preventable medical errors in 1999 (Ball *et al.* 2003), most of the healthcare systems has become more sensitive about the human health. They just want to reduce the medical errors and increase the patient health and that could only be done with the help of modern Information techniques and their devices. So, most of the healthcare system has adopted a new information technology techniques and devices in providing the entire medical facility to the patient. This research paper with the title of "***evaluation of pervasive and ubiquitous healthcare systems***" is the evaluation of wireless technologies i.e. wireless sensor network and wireless devices with their contemporary methodologies which are currently functioning with healthcare systems. These wireless technologies are providing dissimilar wireless facilities in healthcare systems with reliability, flexibility, QoS, mobility, security and cost

effectively and how much a patient is fully aware of these modern wireless healthcare systems.

2. Wireless Healthcare Systems

The wireless networks have improved the communication among patients and physician through its wide range of availability and delivered the medical treatment any time and any-where, which has reduced the medical errors in healthcare of the patients. It also increases the wireless technologies effect with intelligent and wearable wireless devices like mobile phone with help of wireless network. The Wireless Healthcare System is that “which provides the tools required for screening, monitoring, and managing of General Consumer Health, Disease Management and Fitness in any healthcare systems” (Card Guard Website). The wireless healthcare systems have some characteristics and issues related to medical healthcare.

2.1. Characteristics of Wireless Healthcare Systems

Here are some most common characteristics of wireless healthcare systems which are very briefly described here:

- *Telemedicine* is the main feature of wireless healthcare system and has the ability to integrate with wireless healthcare devices by supporting the telecommunication and wireless technologies collectively.
- In pervasive wireless healthcare systems *e-Health* is providing facilities to their patients to communicate with their physicians by getting email access. This e-Health concept is the same as E-Commerce.
- The *mobile Health* is an important feature of wireless healthcare system as this providing the medical facilities with the help of latest mobile device or PDA's and is also quite accommodating in emergency situations where immediate diagnoses and response entail.
- *Location based services* also enhanced the medical facilities by tracking the location of the victims and presents the immediate response to the patients while using wireless devices in healthcare systems.
- The *intelligent system* is support the wireless healthcare system intelligently and remotely by adopting automatic managing approaches like biometrics and robotics. This intelligent system will use for location tracking technology and to filter emergency calls by matching different health reports of the victims.
- Wearable devices can be failed during online monitoring due to intrusion of electrical and mechanical home appliances. These intrusions can be the failure cause of wireless devices, networks and application services. The fault tolerance of these devices should be improved by adopting fault detection and monitoring mechanisms (Shiva *et al.*).
- Wireless healthcare system also has a very common feature which is the *remote monitoring* of patients their own homes.

The above defined features are the most common features available in mostly healthcare systems and also facilitating the users with their impressive features.

2.2 Issues in Wireless Healthcare Systems

No doubt, the newly developed and implemented wireless healthcare systems are performing well and serving the humanity with their technological advancement, but still some of the issues related to medical data confidentiality, wireless devices availability and wireless medium security is in under process for fully implementation with no fears. Here are some issues highlighted which are normally occurs in majority of the already implemented wireless healthcare systems. The very brief description of each of them is as follow:

- *Confidentiality* and privacy of medical patient's records is one of the big issues in wireless healthcare systems while using internet as communication protocol. Most of the wireless healthcare systems are currently working to reduce the privacy theft by using the latest designed systems available in the form of software.
- *Consent* means that how much awareness, patients has to this modern wireless healthcare technology treatments. The basic function of the consent is to explain the legal and other risk issues to the patient's health while using the modern telemedicine techniques and also fully informed approval during patient's physical treatment or examinations, to access the patient's electronic data in healthcare systems.
- *Liability* in healthcare system mean, the physicians or healthcare managing authorities should fully aware of the most up-to-date and modern technologies and their standards before going to be launched due to the sophistication in new developed technology and is not easy to use for a common person.
- *Security* of medical data and medical assessment through wireless is one of the biggest issues in healthcare industry. In present modern era, the wireless technique has a lot of security threats and external accessibility media like viruses, hackers and other more security threats. The security threats can be reduced by using the closed system, virtual private network and increase the usage of firewall during communications (Kelly *et al.* 2002).
- *Dependability* and *access control* of medical data in healthcare system are also the important issues mean what are the common factors which have to be subsequent in every healthcare systems and access control mean to define the rules or policies to access the patient data with correlated to the privacy, integrity and confidentiality issues in healthcare systems. Both these issues can be manageable by implementing the authentication and access security techniques.
- The *cost deployment* of medical devices and wearable sensor network could effects the overall cost of the treatment in wireless healthcare systems. Some of the existing implemented wireless healthcare systems are also charging services charges from the patients in wireless healthcare system. So, the cost of deployment and extra charges should be negotiable.

The above mentioned are very common issues which are normally exist in each of the existing wireless healthcare systems.

3. Wireless Sensor Networks

The wireless sensor networks allow the hardware devices to wider acceptance in the physiological monitoring of the patients in crucial stages of their health. These sensor networks have the ability to increase the processing, computational, energy consumptions and wireless communication through their devices between the patients and medical healthcare services providers using a link or communication protocol. This part of paper is describing a very concise picture of existing or implemented wireless sensor networks and their benefits, limitations and evaluation matching with present requirements of wireless healthcare systems. There are three main categories of sensor networks are currently available and are wearable, mobile and wireless sensor which are based on personal area networks respectively.

There are some example of existing wearable wireless sensor networks like CodeBlue, AMON, HealthGear, and BSN which are using the wearable devices like in the form of a wrist watch, designed chips which can be directly attached with human body, and interactive home where the wireless healthcare devices can communicate with wireless healthcare facilities. The mobile wireless sensors normally worked with mobile devices and they can communicate with wireless healthcare systems with modern mobile phones. The examples of mobile existing wireless sensor networks are mPCA and MobiCare healthcare systems. The third category of wireless sensor healthcare system is personal area sensor network which interact with wireless healthcare system through WLAN i.e. the Wireless LAN. The example of these personal area sensor networks is Vision – the Future Hospital which is only designed only for testing bases and still under consideration in medical Labs. These wireless sensors networks are serving the humanity will tangible facilities but the problem is that how, but now finally the author moving to analysis and investigation phase of this research area.

During research of this subject area, the author has found that what are the main operating entities of these wireless healthcare systems? These are wireless medium, physicians and the patients which are directly attached with wireless healthcare system. The author has decided to conduct a small scale of investigation related to wireless healthcare awareness to the physicians and patients has and what are their thinking about these wireless sensor networks. The investigation involved the distribution of a small size of survey in the form of questionnaires. The rest of the part of this paper is describing the survey methods and discussion related to the obtained results from survey questionnaires.

4. Methodology used for survey

Initially the literature related to this specific survey is searched and collected from all available sources like libraries, internet, newspaper, books, magazines and electronic and published journals. The data collection is carried out by adopting the research strategies defined by the Cresswell (Cresswell, 2003). According to Cresswell, there are many types of technique or approaches can be used to collect data from the available sources, but in case of author he used evaluation approach defined by the Cresswell.

4.1 Evaluation Approach

The evaluation approach is another research category for analysis of the data. This approach is a combination of different approaches including the experimental analysis and practical analysis, but in case of my evaluation and investigation, I used a Survey approach for one of my part of evaluation related to the awareness of a common people about the modern healthcare systems.

In this survey I intended to look at wireless healthcare devices and networks with medical healthcare as “Patient Prospective” and the intention is that the results from this work should be of interest for all those vendors and wireless healthcare systems for their future works. I designed and conduct this survey using three approaches i.e. quantitative, qualitative and mixed approaches (Cresswell, 2003).

In the start of this survey I was not be able to construct questionnaires to be used in a quantitative survey about the measurements of awareness, attitudes, and behavior. I simply did not know that what type of questions to ask. So, I choose the “Qualitative research is exploratory and is useful when the researched does not know the important variables to examine. This type of approached may be needed because the topic is new. The topic has never been addressed with a certain sample of group of people, or existing theories do not apply with the particular sample or group under study” (Cresswell, 2003).

5. Discussion and Survey Results

Through this survey, I accomplished the desired results and now be able presently longing to underline those issues of the wireless and their devices which have to face a common person in wireless medical healthcare systems and to collect the literature related to wireless healthcare systems. 37 % of people show interest in wireless healthcare system and the rest of 63% have aborted from the survey due to less technological knowledge. According to the collected results, it can be presenting that some patient’s uncertain behaviour against this wireless healthcare systems and they still crave to carry on with present and currently running healthcare systems.

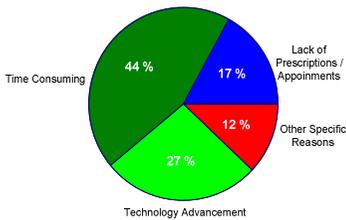


Figure 1: Graph Presenting Factors and Healthcare System

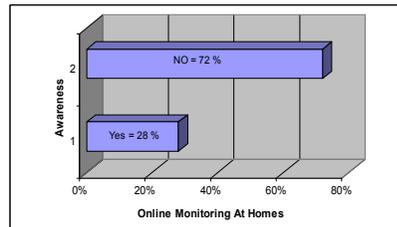


Figure 2: Online monitoring at home

Almost 63 % of people are not happy with the present environment in healthcare systems or hospitals and they recommend improving the healthcare system with more technological advancements due to reduced the timing wastage standing in queues (Figure 1). On the other hand, they do not have greatly knowledge on

wireless devices used in healthcare systems and the newest techniques i.e. online monitoring and online medical prescription through wireless communication links but average of 55 % results that they prefer to visit personally to the physician wait for long appointments and then collect the prescription by hand (Figure 2). This is another example of patient’s uncertainty behaviour against these modern wireless healthcare systems.

More than 55 % percent people are thinking that internet is a suitable wireless medium which can be used in wireless healthcare systems but it must maintain the security, reliability, interpretability, and reusability issues during the whole communication between client-serve communications. The newly established healthcare systems should be cost effective, less consumption of time and also provide the extreme level of treatment and training.

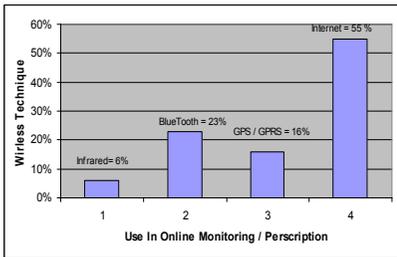


Figure 3: Most preferable wireless medium

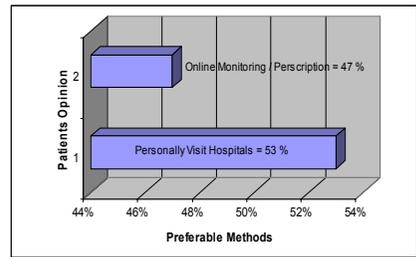


Figure 4: Preferable prescription methods

The results obtained from survey are also showing the death rate increasing due to the un-satisfactory and careless behaviour of physicians.

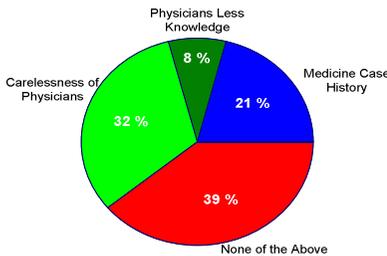


Figure 5: Physicians related Issues

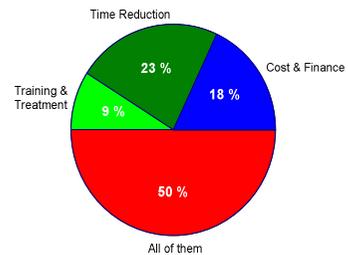


Figure 6: Negotiable issues by the patients

The wireless technology is boost up the healthcare systems with its steady, fast, computerized and location based characteristics but these wireless medium still has security issues like confidentiality, availability and integrity of the wireless data (figure 8). The majority of the participants are preferring electronic data accessed techniques and also suggesting that the administration should adopt recently skills defined access policies, encryption, and up-to-date the users for authentication update to maintain the electronic records. These results are also suggesting that wireless usage can be increase in healthcare systems if there should be an increase in

wireless access points, use 2.4 GHz or more powerful range cordless phones, WAP coverage, use Wi-Fi channels, reduced environmental issues and protect the wireless systems from security thefts like hackers and viruses.

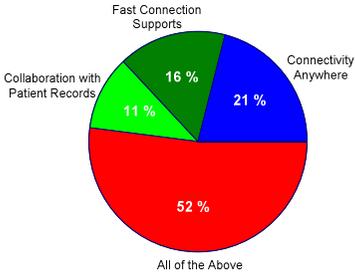


Figure 7: Wireless can provide

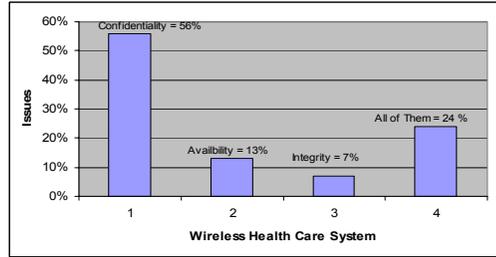


Figure 8: Wireless Issues in healthcare Systems

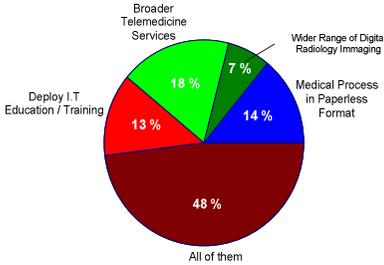


Figure 9: Wireless deployment will help healthcare system

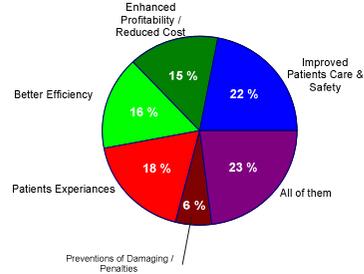


Figure 10: Patients Suggestion to Improve the Wireless healthcare system

The survey results represents that wireless healthcare systems will be acceptable for a common person if these devices and wireless healthcare systems will compatible with existing approaches, fit in cultural and social modes, user friendly, cost effective and provide the extensive healthcare facilities in a single environment.

Over all the survey results are very impressive and this survey accomplish its desired approach. I suggested in the light of survey results that people intend to adopt the new wireless technology in healthcare system but they are still in indecisive situation either they have to assume these new technology or not?.

6. Conclusion and Future Recommendations

Wireless technology is broadening the entire world networks scenario with their wireless devices. These technologies are also working with every aspect of human life even in medical sciences. The medical science is using wireless technologies for medical treatments, facilitating and monitoring patients at their own homes. The next 15 to 20 years of future will be the essential time for medical industry and medical staff due to wireless devices and wireless sensor networks. The manufacturing organizations or vendor are functioning to developed new cognitive wireless systems

with distributed network capabilities. They are functioning to shaped new control system which can get the widely used public attention any-where and will remain the hidden from the users. Here are some future recommendation which I think that would help out to the physicians and patients while interacting with wireless healthcare. These are:

- *Broader HealthCare Services* – wireless services could be improved in healthcare systems especially facilitating the users at their homes. According to survey results (Figure 9), the people suggested that deploy the IT services with broader telemedicine ranges and medical records in paperless format would improve the wireless healthcare services and the wireless can provides the fast connectivity any-where with collaboration of medical patients records (Figure 7).
- *Provide Fault Tolerance* – In pervasive computing, the term fault tolerance is implies on the wearable healthcare devices at home, there is an intrusion can be occur between the medical sensor wireless devices and the electrical, mechanical and automobile appliances. So, the newly developed sensor networks and device should be capable of showing non-intrusive behaviour. According to Shiva Chetan, “the fault tolerance issue have not yet properly addressed in pervasive and ubiquitous healthcare systems” (Shiva *et al.*, 2004).
- *Awareness and Training* – According to survey results, the people are not well aware off the online monitoring and online prescription wireless healthcare systems techniques (Figure 2) and they are prefer to visit the physicians and prescribed from him at hospitals instead of modern healthcare systems (Figure 4). In future, we would have to take some solid step to enhance the awareness of such type of sensor network and devices among customers or patients. If they will properly aware of all these devices then they will adopt them, and it is only possible with adopting some motivation techniques like survey, newsletter and seminars.
- *Improve Data Confidentiality* – wireless should improve the data security and confidentiality. According to survey results, most of the survey participants prefer Internet as the most suitable wireless medium for wireless healthcare systems (Figure 3) but they still have to face confidentiality of medical data is one of their big issues (Figure 8). So, the newly developed healthcare system would follow the data privacy and confidentiality standards define in Data Protection Act 1998 and secure the medical data from third party like viruses and hackers by defining some authentication and authorization policies.

In research paper, I locate that the both patients and physician are the most interactive entities, which involved with wireless healthcare systems and wireless devices. After careful analysis of my designed survey results, I found that only 37 % of people are conscious of wireless healthcare technology and these 37 % participants still have ambiguity in their minds, either they would accept and use them or not? Which is directly associated with wireless sensor networks and their devices. Majority of them are willing to adopt these new wireless technology but they do not know how to manage it? Another main cause, I have illustrated that the cost of new wireless technology to the patients is the cost of wireless sensor facilities

and their devices is too high and could not be affordable for an ordinary patient, and the third cause is that there is a very rare research work available on wireless fault tolerance of wireless devices in healthcare systems. In wireless healthcare systems, most of the wireless devices are wearable and normally attached with the human body. So, while in wireless communication with the patients, the wireless devices could be interrupted with the home mechanical and electrical appliances due to magnetic interference between two devices, and the wireless signals could be dropped. So, if the wireless healthcare systems want to improve these technologies then they would have to reduce the cost, produce and distribute the appropriate understanding to the patients about the new wireless healthcare technology. Hence, from this research paper evaluation it is clear that the existing wireless healthcare system still has some ambiguity to their implementation and deployment issues and their architecture. They still desire for further research before fully deployed in wireless healthcare industry because the patients do not ready to take risk with their lives.

7. References

Ball, M.J., Weaver, C.A. and Kiel J.M., (2003), "*Healthcare Information Management Systems*", cases, strategies and solutions, 3rd edition, Springer, ISBN 0-387-40805-3.

Card Guard: "*PMP wireless healthcare system*", PMP4 web based platform, Website. [Available online] <http://www.cardguard.com/site/products-list.asp?id=17>, (accessed 12/08/06).

Cresswell, J.W., (2003), "*Research Design. Qualitative, Quantitative, and Mixed Methods Approaches*". Second Edition. SAGE publications. ISBN: 0-7619-2442-6.

Kelly, G. and McKenzie, B., (2002), "*Security, Privacy and Confidentiality issues on the Internet*", Journal of Medical Internet Research, Jmir, 2002. [Available Online] <http://www.jmir.org/2002/2/e12/>, (accessed 17/07/06).

Shankar, S.C., Ranganathan, A. and Campbell, R., (2004), "Towards Fault Tolerant Pervasive Computing", White Paper, Pervasive 2004 Workshop on Sustainable Pervasive Computing, Linz/Vienna, Austria, April 2004. [Available Online] <http://choices.cs.uiuc.edu/~chetan/cpapers/tfpc.pdf>, (accessed 17/07/06).

Web-Based Risk Analysis and Education for Home Users

J.Marston and N.L.Clarke

Network Research Group, Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

Broadband Internet access is now widely available to home users, providing better data transfer rates than dial-up Internet access. However this improvement in technology comes at a price with home users at an increased risk of unauthorised access to their resources and information as a result of the ‘always on’ nature of Broadband. These new risks mean that there is a need to provide a web-based risk analysis tool specifically for home users that overcome the problems associated with the current standards and techniques available which require expertise in security and risk analysis are oriented towards organisations and may be prohibitively expensive.

The tool proposed in this paper is designed around the relevant sections of the ISO17799 standard. The advantage of this is that it is based on an industry wide standard that has been adapted specifically to the home user. The standard indicates why various security controls are required but does not indicate how to implement them. The tool provides this implementation information along with advice on secure working practices and also serves as an educational tool.

Keywords

Risk Assessment, Broadband, ISO17799, Home User, Probability, Threat, Potential Impact.

1. Introduction

Until recently residential users generally accessed the Internet using a dial-up connection. The main limitation of a dial-up service is the data rate available for information transfer. This can be anything from 28.8kbs⁻¹ to 33.6kbs⁻¹ (Bates, 2000). This technology is now being replaced by Broadband access which can be defined as a high capacity bi-directional connection (France, 2004).

The technology being considered is Asymmetric Digital Subscriber Line (ADSL) in which the home user to Internet Service Provider (ISP) - (upstream) - data transfer rate is lower than the ISP to home user - (downstream) - data transfer rate. Upstream rates are typically 16 to 640kbs⁻¹ and downstream rates are typically 1.5 to 8.192Mbs⁻¹ (Bates, 2000).

In addition to the increased data transfer rates there are two other contributing factors to the significant growth of Broadband Internet access. First the British Telecom (BT) exchange upgrade and second the fall in subscription rates due to increased competition. As a direct result of these drivers the number of residential Broadband connections in the UK reached 6.2 million by the end of December 2004 which

equates to approximately 25% of all households (Ofcom, 2005). The success of Broadband has meant that there are more Broadband accounts than there are Narrowband or dial-up accounts (Ofcom, 2005).

A flat fee is charged for Broadband subscription therefore connections are often left on for long periods of time making home users computer resources and information an easy target for intruders (Kuhn *et al*, 2002). Surveys conducted have highlighted that home user's lack knowledge in terms of computer security. One survey conducted showed that 52% of the people interviewed had little or no knowledge of computer security and 53% did not know how to improve their computer security (HM Government, 2005). Another survey showed that 81% of home users did not have at least one of the minimum protection mechanisms installed, namely updated Antivirus software, spyware protection software, and a secure firewall (AOL/NCSA, 2005). These problems are compounded by the fact that vulnerabilities exist in the operating systems themselves and the change in Internet usage habits as a direct result of its accessibility (Orvis *et al*, 2005).

Standards and techniques are available that guide users through risk analysis but these are costly, require expertise and are aimed at organisations rather than home users (Mash, 2002). As a result there is a need to develop a suitable alternative and this paper proposes a web-based risk analysis tool, based on the ISO17799 standard, aimed specifically at home users.

The rest of the paper is organised as follows. Section two introduces the new risks associated with Broadband Internet access. In section three the risk management concept is described and how it will be captured by the web-based tool. Some current standards and techniques are identified along with reasons why they do not match the requirements of the home user. The paper then outlines the structure of the proposed web-based tool in section four and section five concludes the paper.

2 The new risks associated with Broadband access

Unfortunately the improvements in data transfer rates provided by a Broadband connection come at the cost of an increased risk for unauthorised access to the home user's computer. For Broadband connections a computer may be left on all day therefore the Internet Protocol (IP) address will remain fixed for longer periods of time thus providing hackers more time to compromise the computer (Kuhn *et al*, 2002). The main reasons for hackers targeting a Broadband connection are as follows (Glass, 2001):

- **Bandwidth** – in general a home user will only use a small amount of the available bandwidth. Hackers can therefore use the spare bandwidth without having a detrimental affect on the home user's data transfer rates. This means they go unnoticed and can utilise this bandwidth for Distributed Denial of Service (DDoS) attacks.
- **Resources** – the Central Processing Unit (CPU) capabilities and memory are useful to hackers. They can use home computers as servers for chat rooms or for storing illegal copies of software. Due to the amount of processing power and

memory available on modern computers the loss of performance and storage would go unnoticed by the home user.

- **Information** – personal and financial information is generally stored on the home user's computer. If online transactions have been made, bank account details stored and so on then sufficient information may be available for identity theft.

It is evident that there are new risks associated with Broadband Internet access but there are relatively simple ways for home users to counter these risks. Problems caused by the shortfall in the home user's lack of computer security knowledge, their Internet usage habits and system vulnerabilities can be overcome by providing a web-based risk analysis tool that educates the home user, informs them how to implement the protection measures available and provides advice on secure computing practices.

3. Information security and risk analysis

At home people will invest time and effort using their computer to create documents and store information. The computer and its applications will be used as tools to help with various tasks and to track different types of information i.e. finances. This type of information has a value to the home user and can be considered as an asset that needs to be protected. The web-based risk analysis tool will provide information security using a risk management strategy consisting of the following three separate processes (Stoneburner *et al*, 2002):

- **Risk assessment or risk analysis** – identifying assets and then considering what the threats are to those assets. This will be the primary aim of the web-based tool. Answers to a security questionnaire will be used to determine the level of risk to the home user.
- **Risk mitigation** – putting controls in place to protect the assets from the identified threats. This will be the secondary aim of the web-based tool. Guidance will be provided with regard to the protection measures available and how to implement them.
- **Evaluation** – risks and mitigation tools should be continually reviewed to ensure that security is maintained. The web-based tool will indicate the importance of this and will be updated by the developer as new threats and protection methods are realised.

Conducting a risk assessment is not an easy exercise, particularly if trying to comply with one of the security standards as well as catering for the varying degrees of computer literacy. For small-to-medium enterprises (SME) the expertise may not be available internally and the cost of employing a consultant may prove too expensive. These problems are also applicable to the home user. However the risks are the same as those faced by larger organisations therefore it is important that an assessment is conducted (Mash, 2002).

A suitable alternative for an SME is to purchase an off-the-shelf product, which guides a user through the risk assessment process identifying the security threats and

vulnerabilities and the controls required to mitigate these risks (Mash, 2002). However for the home user the cost will again be too high and they will be largely oriented towards organisations.

This reinforces the need for designing a web-based risk analysis tool specifically aimed at the home user. The cost, the expertise required and the relevance of existing methods illustrating the need to provide an effective alternative.

3.1 Risk assessment standards and techniques

To aid the development of a suitable alternative existing risk assessment methodologies were reviewed to identify one that provided a useful framework that could be adapted to suit the home user. There are numerous standards and techniques available. Some can be downloaded free from the Internet whilst others are available at a cost. The standards and techniques available that have been reviewed include:

- *Operationally Critical Threat, Asset And Vulnerability Evaluation (OCTAVE®)*; (Alberts *et al*, 2003)
- *Control Objectives for Information Technology (COBIT)*; (ISACA, 2005)
- *Site Security Handbook, RFC2196*; (Fraser, 1997)
- *Common Criteria (CC), ISO15408*; (ISO, 2005)
- *Risk Management Guide for Information Technology Systems, SP800-30*; (Stoneburner *et al*, 2002)
- *Information Technology – Code of practice for information security management, ISO17799*. (ISO, 2005)

It was decided that the ISO17799 standard, which is becoming the de facto standard in Europe (Walsh, 2002) would provide a good framework around which the web-based tool could be designed. The output from the risk assessment would identify the risks and provide recommendations on controls that would reduce these risks to an acceptably low level. Using ISO17799 as a guide for identifying these controls provides a consistent strategy, which is in compliance with a widely used and recognised standard (Mash, 2002).

It is intended that the web-based tool will be hosted on a web server for access to home users at no cost. It will use a simple questionnaire oriented specifically towards the home user environment with supporting information to help users answer the questions and it will assume no prior knowledge, hence overcoming the disadvantages of existing techniques.

In addition to the standards and techniques there are a number of websites available from authoritative organisations, such as Microsoft that give advice and guidance on home user security. However the disadvantage is that they are biased towards their products, which often need to be purchased. The advantage of the web-based tool is that it is unbiased and will provide links to resources that are free and equally as effective.

4. Application of the ISO17799 standard to home users

The ISO17799 is a high level standard identifying what should be done not actually how to implement the security measures (Kairab, 2005) or even how to determine the level of risk but it does provide details on various controls that are required to provide information security. The controls that are applicable to the home user environment have been used to create a number of questions, which together make up the risk assessment.

The web-based tool combines these questions with an extended version of the Jacobson's Window risk model, which has quantitative values assigned as shown in Table 1. The values assigned to the potential impacts LOW, MEDIUM and HIGH are 0.1, 0.5 and 1.0 respectively. The values assigned to the probabilities LOW, MEDIUM and HIGH are 10, 50 and 100 respectively. The risk score based on these values will range from 1 to 100.

Probability	Potential Impact		
	LOW (10)	MEDIUM (50)	HIGH (100)
HIGH (1.0)	LOW $1.0 \times 10 = 10$	MEDIUM $1.0 \times 50 = 50$	HIGH $1.0 \times 100 = 100$
MEDIUM (0.5)	LOW $0.5 \times 10 = 5$	MEDIUM $0.5 \times 50 = 25$	MEDIUM $0.5 \times 100 = 50$
LOW (0.1)	LOW $0.1 \times 10 = 1$	LOW $0.1 \times 50 = 5$	LOW $0.1 \times 100 = 10$

Risk scale: HIGH (>50 to 100); MEDIUM (>10 to 50); LOW (1 to 10)

Table 1: Risk level matrix (Stoneburner *et al*, 2002)

Assigning the values in this way means there will only be one case of high risk, which is when both the potential impact and probability are high. This is done intentionally so that the importance and immediacy of a high risk can be seen and acted upon urgently (Kairab, 2005).

Having determined a suitable risk scale it is important that the potential impact and probability of threats are described in order to allow consistency in the risk assessment. The potential impact to home users for each category is shown in Table 2.

HIGH RISK = 100
Potential Impact
Total loss of personal data
Total loss of system and application software
Intruder access to personal information
Intruder access to financial information
Intruder access through 'back door'
Intruder access to hardware i.e. memory, CPU, bandwidth
Monitoring of home users Internet activities
All contacts adversely affected
Computer system unavailable for an extended period

MEDIUM RISK = 50
Potential Impact
Personal data recoverable with significant effort
System and application software recoverable with significant effort
Intruder access to personal information only
Abnormal display and computer activity
Internet activity monitoring detected and addressed
Parts of the computer system unavailable for significant period
LOW RISK = 10
Potential Impact
No loss of personal data
No loss of system or application software
No intruder access to personal information
No intruder access to financial information
No intruder access to hardware i.e. memory, CPU, bandwidth
No monitoring of home users Internet activities
Contacts not affected
Computer system remains available

Table 2: Potential impact to a home user

For each category of potential impact not necessarily all listed items will or indeed need to occur. For example if all personal information is lost then this is certainly as a consequence of being at a high risk, even if none of the other impacts listed occur.

When a home user is connected to the Internet using a Broadband connection they are susceptible to a number of threats and vulnerabilities. The probability of these threats occurring and vulnerabilities being exploited will relate to the level of protection already in place and to the way in which the Internet is used. Table 3 identifies the common threats and quantifies them according to both the home user’s level of protection and their Internet usage.

The questions that make up the risk assessment and the associated ISO17799 section are shown in Table 4. The web-based tool splits the questions into two sets, the first set, questions 1-14, relate to the controls that can be implemented to protect the user and their computer and the second set, questions 15-28, relate to best practices.

The home user’s answers to questions 1-14 will be quantified using appropriate values from Tables 2 and 3 to calculate a risk score which will be displayed to the home user as LOW, MEDIUM or HIGH at the end of the risk assessment. If their score is either MEDIUM or HIGH a hyperlink will be provided to a web page indicating ways in which the risk can be reduced and providing links to the resources required i.e. Antivirus software. If their score is LOW they will be informed that their configuration is optimal and no changes are required.

Threat	Probability	Comment
Virus	1.0	If no Antivirus software installed
	0.5	If Antivirus software installed but not up to date
	0.1	If Antivirus software installed and up to date
Trojan Horse	1.0	If no Antivirus software installed
	0.5	If Antivirus software installed but not up to date
	0.1	If Antivirus software installed and up to date
Worm	1.0	If no Antivirus software installed
	0.5	If Antivirus software installed but not up to date
	0.1	If Antivirus software installed and up to date
Spyware	1.0	If no anti-spyware installed and regular freeware downloads
	0.5	If no anti-spyware installed and some freeware downloaded
	0.1	If anti-spyware installed and/or no freeware downloaded
Adware	1.0	If no anti-spyware installed and regular freeware downloads
	0.5	If no anti-spyware installed and some freeware downloaded
	0.1	If anti-spyware installed and/or no freeware downloaded
Back door	1.0	If Windows automatic updates disabled
	0.5	If Windows automatic updates done manually
	0.1	If Windows automatic updates enabled
Identity theft	1.0	If personal information is stored on the computer without password protection
	0.5	If all personal information stored is protected using strong passwords
	0.1	If no personal information is stored
Financial loss	1.0	If financial transactions are made without checking for a secure connection
	0.5	If checks are made sometimes to ensure a secure connection during financial transactions
	0.1	If either checks are always made to ensure a secure connection during financial transactions or no financial transactions are carried out
Phishing	1.0	If you respond to emails asking for personal account details
	0.5	If you respond to some emails asking for personal account details
	0.1	If you delete all emails allegedly from banks, eBay, etc asking for personal account details
Spam	1.0	If no Spam filtering installed
	0.5	If Spam filtering done manually by the user
	0.1	If Spam filtering done automatically

Table 3: The probability of a threat occurring or a vulnerability being exploited

As an example consider questions 1 and 2 which relate to Antivirus software. There are three possible scenarios. First if a home user answers yes to both questions then their risk score will be 100 (potential impact – Table 2) x 0.1 (probability – Table 3) = 10, which according to the risk level matrix in Figure 1 is LOW. Based on this they will be told that they do not need any additional controls. Second if a home user answers yes to the first question but no to the second question then their risk score will be 100 (potential impact) x 0.5 (probability) = 50, which according to the risk

level matrix is MEDIUM. Based on this they will be advised to check that the automatic update facility of their Antivirus software is enabled. Third if the home user answers no to question 1 then question 2 will not be displayed and their risk score will be 100 (potential impact) x 1.0 (probability – by default) = 100, which according to the risk level matrix is HIGH. Based on this they will be guided towards suitable Antivirus software and advised to install it as soon as possible. The same principle will be used for the other questions and where necessary, guidance will be provided on the controls that can be implemented to reduce the risk score.

Question number	Question	ISO17799 section
1	Do you have Antivirus software installed?	8.3.1
2	Do you regularly update the Antivirus software?	
3	Do you have Anti-spyware software installed?	
4	Do you download freeware or shareware from the Internet?	
5	Do you have a software firewall installed?	9.4.2
6	Do you have a hardware firewall installed?	
7	Do you have automatic updates for Microsoft Windows enabled?	10.4
8	Do you store personal information on the computer i.e. address?	9.6.1
9	Do you store financial information on your computer i.e. bank account details?	
10	Do you use strong passwords to protect personal and financial information?	9.3.1
11	Do you use online banking?	8.5.1
12	Do you purchase goods from Internet websites?	
13	Do you respond to emails asking for user account and password details?	8.7.4
14	Do you open all emails, even those from unknown sources?	
15	Do you leave your Internet connection on when not in use?	9.5.8
16	Do you regularly back-up your files to a suitable storage device i.e. CD?	8.4.1
17	Have you recently upgraded any software or hardware?	4.1.4 & 10.1.1
18	Have you recently installed any new software or hardware?	
19	Have you registered your system and application software?	10.4
20	Do you change your passwords at regular intervals i.e. monthly?	9.3.1
21	Do you use the same password for all accounts?	
22	Do you store your passwords on the computer?	
23	Do you always scan email attachments received for viruses?	8.7.4
24	Do you check that digital signatures are valid?	10.3
25	Do you scan files on all storage media for viruses before accessing them?	8.6.1
26	Do you only accept media from authorised or trusted sources?	
27	Do you regularly check the firewall log?	9.7.2
28	Do you know how the security settings on your Internet browser are configured?	9.4.1

Table 4: Risk assessment questions mapped from the ISO17799 standard (ISO, 2005)

The home user's answers to questions 15-28 will be recorded for each user. When the user has completed the questionnaire a hyperlink will be provided to a web page that will identify best practice based on any weaknesses highlighted in their answers i.e. scanning all email attachments for viruses.

As an example consider question 15 relating to the physical Internet connection. In this instance if the home user leaves their Broadband connected when not in use it will be recommended that they disconnect it as soon as they have finished with the Internet to minimise the risk because the longer they are connected the less the IP address will change and the more chance an intruder has of identifying it and gaining unauthorised access. In addition it will also be highlighted that it is good practice to physically disconnect the telephone line from the computer for protection against lightning strikes during any storms.

Questions 15-28 are very important. For example, consider a home user with up to date Antivirus software installed. Although the control is installed the user may regularly open email attachments without first scanning them for viruses. Their actions put them at a risk of a virus attack and this has not been captured in questions 1-14 but it is clear that the protection software and the user behaviour are both important in terms of information security.

Questions 1-28 will be the core part of the risk analysis tool. As a minimum the home user should be able to work their way through these questions and be presented with a set of risk scores along with the support needed to minimise them and a set of best practices which together with the controls will provide the best possible security solution.

5. Conclusion

The widespread use of Broadband Internet access has led to the increased likelihood of hackers gaining unauthorised access to the home users computer resources and information. This problem is exacerbated by the users' lack of knowledge and the vulnerabilities of the computer software itself. Standards and techniques are available that provide a risk analysis. However these require expertise, are oriented towards organisations and may be expensive. Organisations offering online advice on home user security are generally biased towards their own products and often assume a certain level of knowledge.

The proposed web-based risk analysis tool overcomes these problems. It assumes no prior knowledge of risk management, it is aimed specifically at the home users and will be hosted on a web server for access free of charge. It uses a simple questionnaire with supporting information, to highlight the areas of risk and provides links to the resources required to reduce these risks to an acceptable level, educating the users in the process. The risk assessment questions focus on the home user environment with the controls being selected in accordance with those identified in the ISO17799 standard. In addition to this the importance of secure working practice has been emphasised.

The advantage of using the ISO17799 standard as a framework for developing the tool is that it provides a consistent strategy, which is in compliance with a widely used and recognised standard. The next phase is user testing of a prototype tool that has been developed. The feedback will be used to improve the functionality of the tool as necessary.

6. References

Alberts, C., Dorofee, A., Stevens, J. and Woody, C. (2003), '*Introduction to the OCTAVE® approach*', http://www.cert.org/octave/approach_intro.pdf, (Accessed 6 September 2005)

America Online and the National Cyber Security Alliance, AOL/NCSA (2005), '*AOL/NCSA Online Safety Study*', http://www.staysafeonline.info/pdf/safety_study_2005.pdf, (Accessed 2 August 2006)

Bates, R.J. (2000), '*Broadband Telecommunications Handbook*', McGraw-Hill

France, P. (2004), '*Local Access Network Technologies*', The Institution of Electrical Engineers

Fraser, B. (1997), '*Site Security Handbook*', <http://www.ietf.org/rfc/rfc2196.txt>, (Accessed 16 September 2005)

Glass, B. (2001), '*Got Broadband? You're under attack*', <http://www.extremetech.com/article2/0,1558,23886,00.asp> (Accessed 18 August 2005)

HM Government (2005), '*Get Safe Online Report*', <http://www.egovmonitor.com/reports/rep12338.pdf> (Accessed 21 August 2006)

ISACA (2005), '*COBIT and related products - Guidance material for IT Governance*', http://www.isaca.org/Content/NavigationMenu/Members_and_Leaders/COBIT6/Obtain_COBIT/CobiT.pdf, (Accessed 5 September 2005)

ISO (2005), 'Information technology – Security techniques – Code of practice for information security management', *ISO/IEC 17799:2005 International Organisation for Standardisation*

ISO (2005), 'Information technology – Security techniques – Evaluation criteria for IT security', *ISO/IEC 15408:2005 International Organisation for Standardisation*

Kairab, S. (2005), '*A Practical Guide to Security Assessments*', Auerbach Publications

Kuhn, R., Tracy, C.M. & Frankel, S.E. (2002) 'Security for Telecommuting and Broadband Communications', *Special Publication 800-46, Recommendations of the National Institute of Standards and Technology*

Mash, S. (2002), 'Risk Assessment for Dummies', *Computer Fraud & Security Journal*, Vol. 2002, Iss. 12, pp. 11-13

Ofcom (2005), '*The Communications Market 2005 – 3. Telecommunications*' <http://www.ofcom.org.uk/research/cm/cm05/telecommunications.pdf> (Accessed 16 August 2005)

Orvis, W.J., Krystosek, P. and Smith, J. (2005), '*Connecting to the Internet Securely; Protecting Home Networks*', Computer Incident Advisory Capability CIAC-2324, www.ciac.org/ciac/documents/CIAC-2324_Connecting_to_the_Internet_Securely_Protecting_Home_Networks.pdf, (Accessed 22 August 2005)

Stoneburner, G., Goguen, A. and Feringa, A. (2002), 'Risk Management Guide for Information Technology Systems', *Special Publication 800-30, Recommendations of the National Institute of Standards and Technology*

Walsh, L.M. (2002), '*Security Standards*', <http://infosecuritymag.techtarget.com/2002/mar/iso17799.shtml>, (Accessed 2 September 2005)

The Art of Network Monitoring

A.Mohyuddin and P.S.Dowland

Network Research Group, University of Plymouth, Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

This research paper focuses on different types of Network Monitoring techniques and putting micro level details on various elements that contribute to a good network monitoring platform. There are thousand of network monitoring systems available in the market; it is hard to conclude which system is best to requirements and what elements needs consideration when making a choice, some good monitoring systems has been discussed in this research paper.

Keywords

Network Monitoring, Network Monitoring Tools

1. Introduction

Network monitoring systems can be seen as a complete solution to constantly monitor the network performance against any failures, bottlenecks or unusual activities that can result in slowdowns or breakdowns of computer networks. Today network monitoring systems are working along with security applications to prevent computer network from outside world and any vulnerability within the organization. A recent study conducted by computer security institute (Flukenetworks.com, n.d.) and FBI revealed that out of 264 companies surveyed 53% of the companies detected un-authorized usage of the company's network and approximately 50% of un-authorized usage was reported within the organization. These figures are very alarming as companies now just do not have to secure themselves from the outside world but also within the organization.



Figure 1: Problem solving with network monitoring (Network Probe, 2006)

Modern network monitoring systems are more responsive then they were ever before, they analyze network usage and study different behavior on the network at all the times and solve problems which they can or have ability to alert it to network administrator immediately incase of any security breach. It is also important to understand that intruders have access to complicated technology, so if any company or an organization wants there networks up and running they have to be secure in a better way; it is also important to point out that company employees always find a

way to breach security policies, any malicious software being installed by any employee can leak out company's data to outside world.

2. Types of Network Monitoring

2.1 Passive Network Monitoring

Passive network (Ciuffoletti, 2006) Monitoring technique examine network traffic by scanning individual packet, this process allows to study patterns on network and helps to determine packet flows. The advantage of using passive network monitoring is that there is no need to insert additional packets, hence keeps the traffic on the network low.

Passive network monitoring (Timm, 2003) is helpful when network administrator need to know very low level detail about the network such as network topology, services, operating system, and application being used at different nodes. This is achieved by scanning TCP and IP headers using various packet sniffers such as Tcpdump or an Ethereal. A variety of information can be gathered just by analyzing a packet such as host logical location can be determined by just determining the TTL field of the IP header. Filters can be created to gather information from the packets and this information can help to determine if there are any vulnerabilities to the network.

2.2 Active Network Monitoring

Active network monitoring works by injecting packets into the network or send it to workstations, servers, applications etc to measure network performance. The problem lies in sending extra packets which sometimes create an extra traffic, but usually little amount of packets can be used to attain desired information. In addition active network monitoring allows a full control over additional packets that are required to be sent over the network, these can be sent whenever required by any specific monitoring application hence are more flexible.

2.3 Hybrid Network Monitoring

Hybrid Network Monitoring (Landfeldt, 2000) is an emerging technique to monitor large number of growing wireless Networks, As the name suggest Hybrid network monitoring make use of active monitoring where passive network data is unavailable and vice versa. The passive monitoring on a wireless network can only be used in case of an open connection; if there is no open connection active monitoring techniques will be used. Imagine two segments of a wireless network; which are wired and wireless.

3. Core of Network Monitoring

Network monitoring covers an extensive range of features; its dimension goes from monitoring different operating systems to checking memory usage or downtimes of devices attached to the network and there are many more potential features offered

by good monitoring packages. There are different types of network monitoring which are as following:

3.1 Bandwidth and Traffic Monitoring

Bandwidth and traffic monitoring helps network administrators to determine any vulnerabilities to the network. Traffic and bandwidth monitoring allows:

- Avoiding any bottlenecks on networks
- Worms entering into the network can be tracked down by looking at the traffic trends
- Nodes with high data transfer rates can be determined for any further investigation
- Bandwidth monitoring can make it easier to avoid any extra cost or quality constraints

Bandwidth and Traffic (Paessler.com, n.d.) monitoring works by recording all outgoing and incoming packets and maintaining a record of how many packets has been transferred and how many packets has been received. Usually traffic monitor maintains its own database for this purpose, however traffic and bandwidth monitor make use of standard protocols such as SNMP, Net flow and various packet sniffer record network usage.

3.2 Performance Monitoring

Performance monitoring collects data at various points from where the traffic is being passed; it monitors the packets flow, packets being successfully transferred and packet loss, availability, CPU load, memory and disk space utilization. This would allow network administrator to look for any slow node or any point where network performance is not up to mark. Network performance monitor software can interface with SNMP and supply information about nodes that are on network.

3.3 Security Monitoring

Network security monitoring (Ferraro, 2003) works closely with Intrusion detection system (IDS) and collects event logs, session logs and historical data and identifies any intrusion. Network security monitoring is usually event driven, and alerts when any event occurs to breach security,

3.4 Application Monitoring

Application monitoring can help network administrators to solve any problems well before time by looking at each application behavior, and how application is performing technically. Application monitoring can help to distinguish nature of the problem caused by applications on the network, can help to restart the application if they are causing any problems. Application monitoring (Polozoff, 2003) works by analyzing large amount of system and event logs and its frequency of occurrences;

this enables to analyze problems at very earlier stage before things start getting to worst.

3.5 Packet Capturing and Protocol Analyzer

Packet capturing and (Packet sniffer, n.d.) protocol analyzers are the software or hardware that has ability to intercept the traffic that is passing through a particular network point, this enables to study network behavior including any problems solving, knowing more about network, network usage etc. Capturing packet allows working on many more application of network monitoring, there are various implementation being used by various applications to transfer packets (approved by RFC) which can help to analyze what applications client are using but it is however considered as less secure and data integrity is damaged by any such of the monitoring device or software.

3.6 Database Monitoring

Database monitoring (Monitoring, n.d.) works by observing a database application on the server, functionality includes querying database after regular interval to see the query response time, disk space, database availability, database access, usage, data creation change or deletion etc, since a database is really critical to business database monitoring also monitors the server machine by checking machine performance, CPU usage, or by studying background processes.

3.7 Web and Email Server Monitoring

Web site monitoring includes accessing a web page (Network Monitoring Tools, n.d.) and domain name servers (DNS) resolution after specific interval of time. A query is made to resolve an internet address, incase of a no response administrator are alerted. Email server use SMTP to send and receive emails, mail server monitoring includes SMTP handshaking with specified mail server by sending an email and receiving an automated response. In case if there is a no response of handshake network administrators are alerted about the problem.

Third party web and email monitoring solutions make use of various check points around the globe and they use various methods to ensure that your network is accessible around the world by testing it from various places.

4. Reviewing network monitoring tools

There are thousand of network monitoring tools available in the market equipped with latest features and technologies that allows network administrators to take control over network even from remote location. Good network monitoring systems are capable of monitoring large number of different devices, compatible with various platforms, analyze network resources and filter very micro level network details but what really makes them a good choice is features that gives network administrator a facility to get indication before worse happen. This is usually achieved by analytical engine present in network monitoring systems. There are many issues to consider

when choosing a network monitoring system such as level of detail being analyzed for resource discovery, alert time, number of devices being supported by and number of networks that can be monitored over a large geographical area; some of the good networks monitoring systems equipped with such technologies are discussed below:

ManageEngine OpManager is one of the complete networks monitoring system; it monitors a very micro level detail of the network devices over a large geographical area. Backed by a good customer service this system cost really high and any organization with critical network can afford to keep it running. There are some freely available network monitoring tools available such as Nagios and Kismet; these systems are capable of reporting network faults via email and text messages and available as an open source free to implement and distribute under public license. The only problem lie is lack of customer support and hard implementation process, expertise are required to implement these monitoring system and look after.

5. Conclusion

Network monitoring tools are key elements for survival of any computer network, although there are lots of network monitoring tools available but there is a further research available on various methods such as Hybrid network monitoring. Also Hybrid network monitoring is gaining momentum as the new generation of networks is a combination of wired and wireless clients. This particular area needs researcher's attention and new hybrid monitoring platforms are needed to be developed for local and remote networks. It is also important to mention here that the next generation of computer networks will involve VoIP applications thus current network monitoring tools has to expand there functionalities to VoIP applications monitoring.

6. References

Ciuffoletti, A. (2006). *Architecture of Network Monitoring Elements*. Retrieved August 1, 2006, from web site: www.coregrid.net/mambo/images/stories/TechnicalReports/tr-0033.pdf

Cottrell, L. (2001). *Passive vs. Active Monitoring*. Retrieved August 1, 2006, from web site: <http://www.slac.stanford.edu/comp/net/wan-mon/passive-vs-active.html>

Ferraro, C. (2003). *Network security monitoring*. Retrieved August 1, 2006, from web site: http://searchsecurity.techtarget.com/tip/1,289483,sid14_gci922007,00.html

Flukenetworks.com (n.d.). *The cost of network security failure*. Retrieved August 1, 2006, from web site: <http://www.flukenetworks.com/fnet/en-us/findit?Document=2422673>

Landfeldt, B. (2000). *The case for a Hybrid Passive/Active network monitoring scheme in wireless internet*. Retrieved August 1, 2006, from web site: http://www.cs.usyd.edu.au/~bjornl/research/papers/icon2000_landfeldt.pdf

Monitoring. (n.d.) Retrieved August 1, 2006, from web site: <http://database.ittoolbox.com/topics/t.asp?t=331&p=343&h1=331&h2=343>

Network Monitoring Tools. (n.d.) Retrieved August 1, 2006, from web site: <http://www.dotcom-monitor.com/network-monitoring.asp>

Network Probe. (n.d.) Retrieved August 1, 2006, from web site: <http://www.objectplanet.com/probe/>

Packet sniffer. (n.d.) Retrieved August 1, 2006, from web site: http://en.wikipedia.org/wiki/Packet_sniffer

Paessler.com. (n.d.). *Bandwidth and Network Usage Monitoring Made Easy*. Retrieved August 1, 2006, from web site: <http://www.paessler.com/prtg>

Polozoff, A. (2003). *Proactive Application Monitoring*. Retrieved August 1, 2006, from web site: http://www.ibm.com/developerworks/websphere/library/techarticles/0304_polozoff/polozoff.html

Remote Monitoring. (2002). Retrieved August 1, 2006, from web site: http://searchmobilecomputing.techtarget.com/sDefinition/0,290660,sid40_gci214268,00.html

RMON. (n.d.) Retrieved August 1, 2006, from web site: http://searchmobilecomputing.techtarget.com/sDefinition/0,290660,sid40_gci214268,00.html

TCP/IP Remote Network Monitoring. (2005) Retrieved August 1, 2006, from web site: http://www.tcpipguide.com/free/t_TCPIPRemoteNetworkMonitoringRMON.htm

Timm, K. (2003). *Passive Network Traffic Analysis*. Retrieved August 1, 2006, from web site: <http://www.securityfocus.com/infocus/1696>

Security Considerations for a Wireless Local Area Network

O.I.Nwobodo and X.Wang

Network Research Group, University of Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

The adoption of wireless networks technology is growing very rapidly in the recent years. Wireless Local Area Networks (WLAN) is enjoying wide deployment in all types of enterprise, small offices and residential homes as it is expected to leverage communication in terms of freedom of mobility, flexibility and low cost. These advantages come with security risks which are worth assessing for efficient WLAN. This research seeks to evaluate some security mechanisms in securing WLAN. The strengths and weaknesses inherent in these mechanisms were discussed. A survey questionnaire was used to sample users' perception on WLAN security. The feedback from the survey was analysed in line with the discussed literature to make effective recommendations which will alleviate the lingering security problems with WLAN.

Key words

Wireless Networks, Communication, Security, Association, Station

1. Introduction

The wireless industry has evolved phenomenally over the past few years. Wireless transmission is now a popular means of data communication for Wireless Local Area Network (WLAN), cellular phones, wireless Personal Digital Assistants (PDA) and text pagers. The most obvious advantages of wireless networking are mobility and flexibility. Wireless network users can connect to existing networks and then are allowed to roam freely. The flexibility of wireless networks can translate into rapid development.

Security on any network is a prime concern. On wireless networks, it is often a critical concern because the network transmissions are available to anyone within range with the appropriate antenna. Furthermore, wireless networks tend to have unclear boundaries. Maxim and Pollino (2003) explained that whereas a security breach in a wired network requires some form of cooperation from an insider, an attacker sitting in a parking lot can easily capture packets from a wireless network. Hence a number of attacks can be easily launched on a wireless network.

To address the problem of security with wireless networks requires an elaborate research and evaluation of some mechanisms used to secure wireless networks. This research seeks to evaluate the strengths and weaknesses inherent in some of the wireless networks

security mechanisms relative to WLANs. A survey feedback was analysed followed by recommendations based on the findings.

2. Keys to real WLAN security

Securing a WLAN involves the task of ensuring a reliable, interoperable, scalable and cost-effective method. Miller (2003) and Reynolds (2003) identified three keys to WLAN security as authentication, privacy and access control.

2.1. Authentication

Gast (2002) explained that authentication is the process of attempting to verify the digital identity of the sender of a communication such as a request to log in. The principle need is for a system that authenticates users via an existing user ID and password. According to Reynolds (2003), WLAN end-users are given enough initial network access to pass credentials to a web-based authentication server and if the process is successful, they are given extended network access.

Miller (2003) explained that a sender (station) must authenticate itself to an access point (AP) when it wants to associate with the AP. During the association, the station and the AP exchanges a common type of authentication acceptable to them.

2.2. Privacy

“Privacy (commonly called encryption) is the process of hiding the meaning of a message” (Cater and Shumway, 2002). This involves the combination of the original information (plain text) and a defined algorithm (cipher) to produce encrypted information (the cyphertext). Encryption is helpful when a sender of a message cannot be absolutely sure that the message will be delivered only to the intended recipient. In addition to viewing private data files, an attacker is potentially able to sniff usernames, passwords and other private information to gain access into a network that does not guarantee privacy.

2.3. Access Control

According to O’Hara and Petrick (1999), access control is the process of allowing or denying a mobile device to communicate with the network. This is accomplished through an access control list (ACL) which is simply a look-up table based on some Identity criteria. Access control mechanisms work very closely with authentication as they rely on a valid identity to make decisions concerning access. Controlling user and group access to specific servers and applications based on credentials is an important element of many networks especially WLAN.

3. WLAN Security Mechanisms

The security risks to users of WLAN technology have risen exponentially as the service became more popular. Attackers have learned that there is much vulnerability in the wireless protocols, encryption methods and in the awareness of

users. Different methods exist to secure a WLAN although no one method guarantees absolute security.

3.1. WEP

Wired Equivalent Privacy (WEP) was designed in order to eliminate the lack of privacy in WLAN. Using the Rivert Cipher 4 (RC4) encryption algorithm, WEP encrypts key transmissions between two communicating devices. WEP protocol is part of the IEEE 802.11 specifications. It comes in different key sizes and the common key lengths are 128-bits and 256-bits. The longer the keys the better as this will hinder the chances of cracking the encryption. According to Edney and Arbaugh (2004), “there are two parts to WEP’s security system viz: authentication phase and encryption (privacy) phase”. The authentication phase proves to a legitimate AP that the mobile device it is associating with knows their 40-bit pre-shared secret key. Privacy phase which is central to WEP’s objectives is intended to prevent strangers from intercepting and understanding the transmitted data.

3.2. Wi-Fi Protected Access (WPA)

Wi-Fi Protected Access (WPA) was developed by the Wi-Fi Alliance to address the flaws with WEP the original native security mechanisms for WLAN.

According to the Wi-Fi Alliance website (2003), WPA implements the hardened Temporal Key Integrity Protocol (TKIP) encryption technique to provide such encryption not found in WEP. TKIP uses key hashing (KeyMix) and nonlinear Message Integrity Check (MIC) with a rapid-rekeying (ReKey) protocol that changes the encryption key about every 10000 packets. TKIP increased the size of the WEP’s 40-bit key to 128-bits thus enabling the authentication server to dynamically generate and distribute keys. The MIC provides a strong mathematical function in which the communicating devices each compute and then compare. A mismatch of the output means that the data have been modified and the packet is subsequently dropped. This prevents an attacker from capturing data packets, altering them and resending them. WPA guarantees secure authentication by implementing 802.1X and the standard Extensible Authentication Protocol (EAP). By greatly expanding the size of keys, the number of key in use and by implementing an integrity checking mechanism, WAP increases the complexity and difficulty involved in decoding data on a WLAN.

The Wi-Fi Alliance has upgraded WPA to increase its efficiency. WPA 2 is an enhanced version of WPA with the inclusion of the Advanced Encryption Standard/Counter Mode CBC MAC Protocol (AES/CCMP) block cipher. The AES/CCMP provides for confidentiality, integrity and origin authentication.

3.3. Remote Authentication Dial in User Service (RADIUS)

Remote Authentication Dial in User Service (RADIUS) is an Internet Engineering Task Force (IETF) security management protocol that is designed for remote authentication. It is a widely deployed protocol for network access authentication, authorisation and accounting (AAA). Reynolds (2003) explained that RADIUS

operates by authenticating dial-in users on a network. When a user logs into a favourite Internet Service Provider (ISP) with user name and password, the information is passed to a Network Access Server (NAS) and then to a RADIUS server over the RADIUS protocol. The RADIUS server checks that the information is correct using authentication schemes like Password Authentication Protocol (PAP) or Challenge Handshake Authentication Protocol (CHAP). If the information matches, the RADIUS server then authorises access to the ISP. Ohrtman and Roeder (2003) argued that RADIUS still has some flaws in its security mechanism. This weakness revolves around the shallowness of the protocol, the clear text PAP and the response/message authenticator's vulnerability to dictionary attack.

3.4. Kerberos

Kerberos is an authenticating protocol developed by the Massachusetts Institute of Technology (MIT). It allows individuals communicating over an insecure network (like WLAN) to prove their identity to one another in a secure manner. Ohrtman and Roeder (2003) explained that Kerberos is based on the key distribution model developed by Needham and Shroeder which involves four processes viz: authentication exchange, ticket-granting service exchange, user/server exchange and secure communications between user and server. These provide both user authentication and encryption key management which guards the network against attacks on data. Kerberos is fairly complex but provides confidentiality, integrity, access control and network availability.

3.5. Virtual Private Network (VPN)

Virtual Private Network (VPN) is a private communication network implemented within organisations to ensure data transmission across a public accessible network. Kosiur (1998) explained that the tunnelling and security functionalities of VPN can be extended to WLAN. Tunnelling allows streams of data and associated user information to be transmitted over a shared network within a virtual pipe. VPN implements the IETF's Internet Protocol Security (IPSec) to provide authentication, access control, confidentiality and data integrity. IPSec is a framework of open standards that ensures private communication over IP.

The Wi-Fi Alliance website (2003) explained that when a WLAN client uses VPN tunnelling, communications data remains encrypted until it reaches the VPN gateway which sits behind the wireless AP. This effectively prevents intruders from intercepting all network communications. Since the VPN encrypts the entire link from the PC to the VPN gateway at the central network, the WLAN segment between the PC and the AP is also encrypted. Although VPN provides excellent security, they present challenges to the overall network management and throughput.

3.6. Firewalls

“A firewall is a security agent that protects a private network from the public network to which it is connected” (Gregoire et al, 2006). It can take the form of a hardware or software in a WLAN to prevent some communications forbidden by a network access policy. They provide controlled access in a network between zones

of different trust levels as defined by the security policy. Firewall provides WLAN with benefit such as: protection from vulnerable services, controlled access to site systems, log in and statistics on network use and abuse, and policy enforcement.

Despite these benefits from the firewall approach, it is by no means a panacea for WLAN security problems. Liska (2003) identified the following problems with firewalls: Restricted access to desirable services, large potential for back doors, little protection from insider attacker and a bottle-neck to the overall network throughput.

4. Research Method

The previous sections of this research have discussed some fundamental mechanisms in WLAN security. This research evaluated the different WLAN security mechanisms while highlighting their strengths and weaknesses. A survey questionnaire has been integrated into this research to sample the opinions of experienced professionals in the field of wireless networking. The questionnaire was distributed among network engineers and administrators in some selected network engineering community. The questionnaire identified the level of awareness of these professionals (respondents) on some WLAN security mechanisms. The respondents also rated the efficiency of the mechanisms. The survey feedback was examined and analysed to make a logical recommendation on how to resolve the lingering security problems with WLAN.

4.1. Survey Results

Which security mechanism do you implement as your main security check? (you can choose more than one option)	Number of responses	Response ratio
RADIUS	17	19%
Kerberos	5	6%
WEP	62	72%
WPA	47	55%
WPA 2	26	30%
VPN	13	15%
Firewalls	77	90%
None	2	2%

4.2. Survey Analysis and Recommendations

The survey feedback displayed the over reliance on firewalls for WLAN security solutions. A significant 90% of the respondents rely on firewall to protect their WLAN. According to Chapman and Zwicky (1995) firewalls are good at blocking ports and IP blocks or addresses, detecting and dropping malformed packets. Even though they are part of WLAN security strategy, it should not be the only security precaution to secure a WLAN. A firewall will be more effective when used as part of an overall network security policy than the network security policy.

72% of the respondents implement WEP for the security of their networks. WEP was developed by the IEEE 802.11 Working Group to provide a level of security that conforms to the difficulty of tapping Ethernet network traffic. While WEP's minimal security met this level of protection, Arbaugh, Mishra, et al (2004) explained that it suffers from weak encryption algorithm and key management issues. The attachment of Initialisation Vector (IV) to the 40-bit secret key weakens the keys making it easier for an attacker to decrypt. Other weaknesses include bad packet integrity checking, lack of session replay and the requirement that all users on a WLAN use the same keys which must be manually entered. This research recommends that WEP should not be relied upon for securing WLANs. Instead, the 802.1X and TKIP based WPA should be implemented to replace WEP. According to the Wi-Fi Alliance website (2003), "WPA greatly increases the level of over-the-air data protection and access control while addressing all known weaknesses of WEP". Users are also advised to imbibe the following configuration if they must use WEP: disable Dynamic Host Configuration Protocol (DHCP), change the administrator's password at regular intervals, disable Service Set Identifier (SSID) and select the highest encryption key algorithm.

An average number of the respondents rely on WPA and upgraded version WPA 2 to secure their WLANs. WPA was developed by the Wi-Fi Alliance to replace the unreliable WEP. WPA uses 802.1X and TKIP while WPA 2 uses AES-CCMP block cipher to provide data confidentiality, integrity and network availability. One of the major concerns of WPA as Reynolds (2003) identified is on the reliability of the Extensible Authentication Protocol (EAP) which it supports. EAP which is the prevalent authentication method in use today sends a clear text message. However, Edney and Arbaugh (2004) suggested that this will be fixed when work on the tunnelled EAP-TLS (Transport Layer Security) is complete. Other problems with WPA/WPA 2 according to the Jupitermedia website (2006) are the interface problem that allows a user to enter weak keys that can be cracked with offline dictionary attacks and compatibility issues with existing WEP architecture.

In general, a good WLAN security system should integrate the highlighted mechanisms into a comprehensive wireless security policy. "A security policy is the set of decisions that collectively determine the limits of acceptable behaviour, and what the response to violations should be" (Bellare and Cheswick, 1994). It encompasses elaborate risk assessment (preparation), building and implementing adequate security control (prevention), monitoring the network for security violations and responding swiftly to any security breach.

5. Conclusion

Wireless networks are a key technology in the actualisation of ubiquitous communication. Security is a paramount issue in this evolving technology and should be addressed in order to realise the benefits in this seemingly growing means of communication.

This research has discussed the strengths and weaknesses inherent in some popular WLAN security mechanisms. A research questionnaire was used to sample the

perception of users on WLAN security. The survey feedback indicated that a good number of the respondents rely on firewall and WEP for securing their WLAN. The survey feedback was analysed in line with the discussed literature to make recommendations on how to alleviate the flaws in WLAN security system.

6. References

Arbaugh, W.; Fraser, T.; Mishra, A. and Petroni, N. (2004), "Security Issues in IEEE 802.11 Wireless Local Area Networks: A Survey, *Wireless Communications and Mobile Computing*, Vol. 4, pp821-833

Bellovin, S. and Cheswick, W. (1994), *Firewall and Internet Security: Repelling the Wily Hacker*, Addison-Wesley, Massachusetts, ISBN: 0-201-63357-4

Carter, B. and Shumway, R. (2002), *802.11 Wireless Security End to End™*, Wiley Publishing, Indianapolis, ISBN: 0-7645-4886-7

Chapman, B. and Zwicky, E. (1995), *Building Internet Firewalls*, O'Reilly and Associate, California, ISBN: 1-56592-124-0

Edney, J. and Arbaugh, W. (2004), *Real 802.11 Security; Wi-Fi Protected Access and 802.11i*, Pearson Education, Boston, ISBN: 0-321-13620-9

Gast, M. (2002), *802.11 Wireless Networks: The Definitive Guide*, O'Reilly & Associates, Sebastopol, ISBN: 0-596-00183-5

Gregoire, J.; Khlifi, H. and Phillips, J. (2006), "VoIP and NAT/Firewalls: Issues, Traversal Techniques, and a Real-World Solution", *The IEEE Communications Magazine*, Vol. 44, No.7, pp93

Jupiter Corporation (2006), "Weakness Found in Wi-Fi Security Protocol", [online]. Available: <http://www.wi-fiplanet.com/news/article.php/3105271>, (Accessed 4 September 2006)

Kosiur, D. (1998), *Building and Managing Virtual Private Networks*, John Wiley & Sons, New York, ISBN: 0-471-29526-4

Liska, A. (2003), *The Practice of Network Security: Deploying Strategies for Production Environments*, Pearson Education, New Jersey, ISBN: 0-13-046223-3

Maxim, M. and Pollino, D. (2003), *Wireless Security*, McGraw-Hill, California, ISBN: 0-07-222286-7

Miller, S. (2003), *Wi-Fi Security*, McGraw-Hill, New York, ISBN: 0-07-141073-2

Ohrman, F. and Roeder, K. (2003), *Wi-Fi Handbook: Building 802.11b Wireless Networks*, McGraw-Hill, New York, ISBN: 0-07-141251-4

O'Hara, B. and Petroni, A. (1999), *IEEE 802.11 Handbook: A designer's Companion*, IEEE Press, New York, ISBN: 0-7381-1855-9 SP1118

Reynolds, J. (2003), *Going Wi-Fi: A Practical Guide to Planning and Building an 802.11 Network*, CMP Books, San Francisco, ISBN: 1-57820-301-5

Wi-Fi Alliance (2003), “Wi-Fi Protected Access: Strong, Standards Based, Interoperable Security for Today’s Wi-Fi Networks”, [online]. Available: http://www.wi-fi.org/files/uploaded_files/wp_8_WPA%20Security_4-29-03.pdf, (Accessed 4 September 2006)

VoIP Security Threats and Vulnerabilities

S.M.A.Rizvi and P.S.Dowland

Network Research Group, University of Plymouth, Plymouth, UK
e-mail: info@network-research-group.org

Abstract

This paper presents the assessment of Voice over Internet Protocol (VoIP) security threats and vulnerabilities along with current security technologies and security patterns. Although these threats and vulnerabilities are mentioned in many research papers but they are still need to be acknowledged for future. The convergence of voice and data in one simplified network brings both benefits and constraints to users. Among the several issues that need to be addressed when deploying this technology, security is one of the most critical. This paper presents recommendations with security patterns for VoIP that are specific to four attacks. The paper is helpful in giving a different perspective of vulnerabilities and risks of VoIP.

Keywords

Security, Technologies, Patterns, Threats, Vulnerabilities

1. Introduction

The VoIP technology being deployed as major infrastructure of organizations and companies have since remained in a cloud of doubt. The question would this technology prevail, while prone to a number of vulnerabilities and exposed to threats, is answered by increasing usage in Western Europe and recent heavy investments by British Telecom in United Kingdom. With the popularity and benefits of VoIP, there are number of increasing security threats taking place. Considering the fact that VoIP systems have security concerns, it is important to continue researches on any existing risks and presenting appropriate solutions.

This paper discusses security in terms of VoIP with threats and vulnerabilities identified. Existing technologies are discussed followed by recommendations and VoIP patterns.

2. Security

The VoIP technology is based on the previously threatened IP network and adds telephony threats as well. As the VoIP technology is evolving, it is collecting vulnerabilities and threats of both Internet and Telecom technologies. Although there have been many articles on security issues but the organizations are still lacking any implementation of security infrastructure steps for VoIP. According to (Schwartau, 2005) the communication world that is moving towards VoIP technology have no security expertise available. The reason is a little amount of budget is set aside for the security. Security needs to be classified in terms of VoIP. The security concept

related to VoIP has many different aspects but there are three main fundamentals Confidentiality, Integrity and Availability (CIA) (Pfleeger and Pfleeger, 2002).

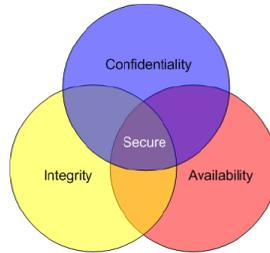


Figure 1: CIA relationship

2.1 Confidentiality

Confidentiality refers to mechanisms ensuring that only intended recipient have access to the VoIP call. Man-in-the-middle attacks are considered to be confidentiality breaches including eavesdropping, sniffing and application attacks. Many free sniffing tools such as dsniff, ethereal, and tcpdump (uses vomit) are available (Porter, 2006). Sniffing packet headers could lead to network and infrastructure disclosure, while sniffing packets leads to leak of private data. ARP monitoring, encryption, VPN are some techniques to mitigate such attacks.

2.2 Integrity

Integrity refers to the prevention of any unauthorized modification in voice packets. Any unauthorized activities must be checked upon. Password breaches are common when a switch reactivates and boots with default settings (Kuhn et.al, 2005). Further attacks include IP spoofing, quality-degradation, registration/session hijacking and server insertion attacks (Ransom and Rittinghouse, 2005). Any rouge packets must be blocked by using VLAN(segmentation), Caller ID verification and fixed routing mechanisms (Green, 2002) should be applied.

2.3 Availability

Availability refers that the VoIP services is always available when needed. Denial of Service which is a threat to availability could have an adverse effect if the VoIP call centre network is hit by such attack. Other attacks include TCP SYN, SIP INVITE flood (Goode B, 2002) and Spam over Internet Telephony (SPIT). Actions needed are using state-full firewalls, Intrusion detection and spam filters on servers (Eyeball, 2006).

3. Threats and vulnerabilities

There are a number of risks associated to VoIP network. Different threats and vulnerabilities are classified in attack categories. The technology needs to be secured as the packets take an unspecified route while traversing from source to destination

end. Analyzing the vulnerabilities and threats while implementing the security measures, is known as 'Risk Identification'.

3.1 Registration attacks

These are such type of attacks where the attacker tries to hack into the system or could be defined as those in which an attacker takes advantage of vulnerabilities in registration injecting themselves into the signal path of the VoIP network. Various type of registration attacks include IP Spoofing, Theft of Service, Reflection Attack, Brute Force Attack

3.2 During a call attacks

These attacks that are carried out mainly when a person is making or receiving a call. The attacker intercepts the route where voice/data packets are being sent. Call Hijacking, Eavesdropping, ARP spoofing (Porter, 2006), Connection Hijacking, Signal Protocol Tampering are some of the attacks in this classification.

3.3 Denial of Service attacks

These attacks have no concern about gaining any valuable information. This simply isolates the endpoint of network from rest of the world by jamming the switches and IP PBX with loads of rouge requests. Different categories include SIP INVITE Flood, TCP SYN Flood, and Malicious RTP Streams (Reynolds and Ghosal, 2002)

3.4 Attacks on VoIP components

These attacks are primarily on the devices, as they seem to be affected easily. The most common attacks are on IP PBX, Soft phones and IP phones

Further attacks include Application layer and SPIT attacks.

4. Security Technologies

It is fundamentally important to establish a security policy to design a secure VoIP system which can guarantee confidential delivery of services to subscribers. Some existing best practises are discussed as follows.

4.1 Virtual LAN

They allow the network administrators to logically divide a LAN in to a number of VLANs. This method provides security if any other VLAN is attacked other remains safe and secure. VLANs use Segmentation which separates voice from data VLAN.

4.2 Virtual Private Network

This technology establishes a private network within the public network. Mainly there are three subsets of VPN of technology, LAN VPN services, Dial-up VPN

services and Extranet VPN services (Venkateswaran, 2001) VPN is based on tunneling. The most popular technologies in VPN are Point-to-Point Tunneling Protocol (PPTP), Layer 2 Tunneling Protocol (L2TP) and IP Security (IPSec).

4.3 Encryption

Voice packet encryption is the best defence against call interception. The major benefit of VoIP based telephony is the ability to encrypt the digital signals representing the voice stream. VoIP designers can also perform encryption at routers.

4.4 Authentication

It is the best to counterattack against registration problems. Authentication is based on cryptography using common secret or public and private key based methods along with signatures and certificates.

4.5 Firewall and NAT Issue

They are handled by a number of methods such as Stateful Firewall that checks each connection present at on any interface of the firewall to make sure they are legitimate Other method is Application Level Gateway (ALG) which intercepts the packet headers and perform modification if appropriate so they correspond to the correct source or destination IP address. Next approach is Demilitarized Zone (DMZ) which is a zone created between a trusted internal network, such as a private LAN, and an un-trusted external network, such as the Internet. (Tanenbaum, 2003). Placing an Intrusion Detection System (IDS) on firewall is also effective. An IDS detects intrusions and malicious behaviour on networks and hosts.

5. Recommendations

Based on the different threats and technologies that were discussed some recommendations might be given as follows

- The endpoint at user's premises shouldn't need any access to the Internet. A connection to the call manager and other phones is feasible.
- The phone or endpoint should not have any access to the normal data, a possible virus outbreak or DoS could result in spread of the virus on the data systems.
- When there is a confidential conversation traversing in the public network a very highly encryption is needed. The information can contain secret key as well.
- Just as the data networks the phones should also be protected by a firewall remotely. The firewall may deny any unencrypted traffic to the phone from Internet.
- The internal data network should be implemented separately with the voice network.
- Any upgrades or configuration required on any devices must need authentication.

- Any ports opened must be closed after call disconnection.

6. VoIP related patterns

Security patterns are better solution to recurring information security problems. (Schumacher and Roedig, 2001). It consists of an overview, problem description and solution with consequences. In this paper it is reduced to two headings. The four patterns are described in this section are Voice/data Segmentation, Tunneling, Call authentication and Call confidentiality.

6.1 Voice/Data Segmentation

The Voice/Data Segmentation pattern separates the voice and data services in order to counter against the threats to voice VLAN from an attacker in data VLAN. The converged services provide the ability to implement the telephony on the existing IP based data network. An economical factor for moving towards VoIP is the ability to use a single network to run both voice and data services. The problem domain consists of finding methods to prevent any attacks from data networks to the voice traffic in a VoIP environment. If an Accountancy company has implemented voice services for example starting a VoIP based customer call centre for marketing and sales purpose. If there is an attack on the data system, there are backup services available to retrieve the data, but the management is doubtful what if an attack leads to the VoIP call centre. The disruption in service would be very costly. How to prevent the voice network from such attacks?

6.1.1 Solution

Two different VLANs for voice and data could be isolated, by using layer three segmentation. All the inter VLAN traffic has to pass through the routing device that filters traffic using access control lists. The deployment of IP telephony services and IP data services should be segregated on two logically separate VLANs (DISA, 2004).

The terminal devices such as IP phones should reside in VLANs that support only IP telephony services. Similarly, the VoIP servers must be protected by a VoIP aware firewall residing on a separate segment. The packet filtering could be easily configured on routing device e.g. routing switches etc. that connects voice and data VLANs. Implementing a state full firewall at Voice VLAN could provide better protection from data VLAN.

6.2 Tunneling

The Tunneling pattern ensures the provision of confidentiality and integrity of voice packets in IP telephony. A voice link has to be established between two or more VoIP end user at remote location on different intranets. The communication link either could be established through a private Metropolitan Area Network (MAN), a Wide Area Network (WAN) or a public medium such as Internet. The Voice traffic is suspected of exposure to hackers while passing through a public network like the Internet. The traffic running on public medium is visible to other private networks.

The problem domain consists of finding methods to counter man-in-the-middle attacks and similar attacks on voice packets running on VoIP network. If an organization that is spanning round the globe wants to connect all its branch offices with the head office so that the communication is better and faster. But the problem is the organization has to choose public medium as the main path because having special leased lines is too much expensive. How would be the confidential information of a company be secure on the public network?

6.2.1 Solution

Virtual Private Networks (VPN) technology provides a tunneling mechanism through the public network to carry any confidential traffic from the private networks. The two locations can communicate securely over these end to end tunnels. One of the end points initiates the connection to establish a secure channel. Appropriate network nodes form the starting and termination points of the intermediated transport network.

VPN technique consists of encapsulation technique. Voice traffic is secured by encapsulating it inside an IPSec or similar tunneling standard. The fundamental mechanism behind tunneling is encryption that ensures confidentiality and data integrity in VoIP networks. Prior to establishing a connection tunneling makes use of Authentication Protocol to set up a trust relationship between the network terminal devices. Encryption could slow down the performance and it's a big issue to quality of service. A symmetric encryption algorithm should be preferred for the voice transportation that would help up in speeding up the process while providing confidentiality. VPN could use public key cryptography.

6.3 Confidential Call

In confidential call pattern security mechanisms such as encryption is provided at the hardware level such as IP phones. When two or more subscribers are engaged in a confidential voice call over a public channel, end users need to be sure that their message delivered from one end to other regains its secrecy. The voice conversation might be intercepted in between the originating and terminating points of VoIP network. The public network such as Internet is not a secure medium; therefore network administrators should apply cryptographic algorithms and techniques in order to ensure security of voice packets. A voice stream on the Internet is vulnerable to eavesdropping. The problem domain consists of deducing techniques to prevent attacks from sniffers while making or receiving a call on the public network. If general home users or a small company which haven't got a big infrastructure such as number of servers and gateways for the VoIP network, like to communicate securely. Applying encryption without tunneling, will that provide similar results?

6.3.1 Solution

To address confidentiality issue, the Secure VoIP Call pattern uses encryption and decryption techniques for VoIP calls. As mentioned earlier latency is an important issue in many converged services, symmetric encryption algorithms are preferred.

This algorithm generates a common cryptographic key i.e. shared secret key passed on both sides of the channel.

Preferably IPSec standard can be used, if so, then it is mandatory for the caller and callee participating in a voice conversation to agree previously on a data encryption mechanisms that must be included in IPSec i.e. DES, MD5, SHA along with a shared secret key. The originator encrypts the call using a common secret key at his end and sends it separately to the person at other end. The receiver decrypts the voice call using the key and playback the information.

Public key cryptography could be used as the other encryption mechanism where latency is not a big issue. This is regarded as the most secure method. In this scenario the receiver must obtain the senders public key before any voice connection is established. The sender encrypts the voice/data with his private key, callee must obtain caller's public key before establishing a connection. Caller encrypts the voice call with callee public key and sends it to him. Callee decrypts the voice call and recovers the original voice packets.

Properties of both symmetric and asymmetric cryptography could be fused together. The symmetric key that needs to be distributed among the end terminals and transported along the same medium could make use of asymmetric cryptography which is feasible for small amount of data. In this way both symmetric and asymmetric cryptography techniques are combined to provide fast and secure results.

6.4 Call Authentication

In Call Authentication pattern user authentication along with the dice authentication is verified when a VoIP call is made on the public network. A voice conversation which is using the public access as a medium could give rise to confidentiality and authentication issues. Subscriber who is imitating a VoIP call could be in doubt whether he is talking to intended recipient or an attacker. Similarly A person at the end of VoIP conversation could not prove the authenticity of Caller, as Caller can decline the authorship of any calls made by him. On top of that, as Public keys are widely available, any attacker could intercept the encrypted data, although he cannot read it but can append any false information or send entirely a new packet encrypted in receivers public key. The receiver may not authenticate the message integrity. The problem domain consists of any methods ensuring attackers are not able to masquerade the call. Solution is needed on how to prove the callers and message authentication so that caller is not able to deny a call made to callee. While making a payment over the IP phone a customer is not sure that the details he is giving to a legitimate party or not, on the other hand the company need to know whatever the customer gives detail, if proved wrong he must not be able to deny it. What should be the mechanism addressing both issues?

6.4.1 Solution

Authentication could be provided by different means. The subscribers can make use of public and private keys to produce digital signatures technique. The public keys need to be exchanged before the voice conversation starts. Sender can sign the voice

packets by using his private key and re-encrypt the result with the receiver's public key. In this scenario only receiver could decrypt the information with his private key and then repeat the same process by using sender's public key. This would provide authentication of the call that has been generated from the legitimate caller.

Another option could be used by applying using MD5. The caller takes the hash of voice signal and encrypt the original message and the hash with callee public key. When the callee receives the call he decrypts using his private key and if the hash of the voice signal is the same as the hash received it means the message is original without any modifications.

Both of the above mentioned techniques could be applied to make the voice conversation more secure but the limiting factor is QoS.

7. Conclusions

The VoIP security issues and solutions play a significant role in the success of VoIP services. Most of the threats discussed above are the threats from public networks. The above discussed recommendations and patterns would be helpful in determining an exact solution to the most common issues in VoIP patterns etc. These VoIP patterns make use of various techniques that are readily available. More information on related patterns is given in (Braga, 1998).

This paper gives an overview of some of the main threats and vulnerabilities posed to VoIP networks. We have also covered some technologies which are incorporated to mitigate such risks. The solutions for VoIP would continue to be researched on as it is a long process, but it would help the end users be aware of security implications and know how they can protect themselves from the VoIP related security threats. Finally in summary VOIP is still an emerging technology, so it is important to counter the emerging and unforeseen threats and vulnerabilities associated with the converged services. As this unique environment of VoIP develops and increase at rapid pace, new challenges and problems would arise.

8. References

Braga, A.M., Rubira, C.M. & Dahab, R. (1998) "Tropyc: A Pattern Language for Cryptographic Software", http://hillside.net/plop/plop98/final_submissions/P25.pdf (accessed August 2006).

Defense Information Systems Agency (DISA) (2004), "Voice over Internet Protocol (VOIP) Security Technical Implementation Guide", iase.disa.mil/stigs/stig/voip_stig_v1r1.pdf, (accessed July 2006)

EyeBall Networks Inc. (2006) "Eyeball Anti-SPIT™ Technology", http://www.eyeball.com/technology/anti_spit.html (accessed August 2006)

Goode, B. (2002) "Voice over Internet protocol (VoIP)", Proceedings of the IEEE Volume 90, Issue 9, Sept. 2002 Page(s):1495 – 1517.

Green, J. (2002) "Voice and Video over IP", McGraw-Hill Professional, USA, ISBN: 0071382488.

Kuhn, R., Walsh T. and Fries, S. (2005) "Challenges in securing voice over IP" ; Security & Privacy Magazine, IEEE Volume 3, Issue 3, May-June 2005 Page(s):44 – 49.

Pfleeger, C. and Pfleeger, S. (2002) "Security in Computing" (3rd ed.) Prentice Hall, USA, ISBN: 0130355488.

Porter, T. (2006) "Practical VoIP Security", Syngress Publishing, USA, ISBN: 1597490601.

Ransom J. & Rittinghouse J. (2005) "VoIP Security", Elsevier Digital Press, USA.

Reynolds, B. and Ghosal, D. (2002); "STEM: Secure Telephony Enabled Middlebox", Communications Magazine, IEEE, Volume 40, Issue 10, Oct. 2002 Page(s):52 – 58.

Schwartau, W. (2005) "With VoIP, it's all over again", <http://www.networkworld.com/columnists/2005/111405schwartau.html?page=2> (accessed June 2006).

Schumacher, M. and Roedig, U. (2001) "Security Engineering with Pattern", <http://www.cs.ucc.ie/misl/publications/files/plop01schumacher.pdf>, (accessed August 2006).

Tanenbaum, A.S. (2003) "Computer Networks" (4th ed.) Prentice Hall, USA, ISBN: 0130661023.

Venkateswaran, R. (2001) "Virtual private networks"; Potentials, IEEE, Volume 20, Issue 1, Feb-Mar 2001 Page(s):11 – 15.

Evaluation of Grid Computing Security

E.Vahedi-Sarrigani and X.Wang

Network Research Group, University of Plymouth, Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

Distributed computing has become extensively prevalent in recent times. Today there are many different ways of using grid computing to improve existing IT infrastructure and distribute data and resources. Grid computing is indeed a very powerful remedy to the problem of sharing diverse resources such as data storage, computational resources and other services provided by different entities. In this work, different applications that utilise or could utilise proper use of distributed computing such as grid is studied.

Unfortunately during the vast development of distributed computing, security risks have not been fully taken into account because of the desire to implement a high quality distributed computational system. But currently the increasing size and profile of the grid require sophisticated and inclusive security measures since they are vital for the success of any undertaking in distributed computing. Attacks and breaches may have dire consequences for both users of the grid computing and resource providers. Thus, a proper security solution is needed that is able to offset any attack on grid resources.

Keywords

Grid, Grid Security, Grid Security Infrastructure

1. Introduction

Grid computing is a type of distributed computing that has become ever more important in recent times. Distributed computing, involves resource aggregation of many computing entities in order to allow them to collectively and collaboratively operate a single computational undertaking in an articulate and clear way.

There are several ways in which workload is distributed between existing IT infrastructure to gain from available shared resources and data. Grid technology is believed to be the most powerful method. This innovative approach improves existing IT infrastructure to use computing resources in best way possible and manage data most efficiently. Grid computing is indeed a useful process utilizing a number of resources such as data and computational abilities of shared entities to collaboratively and simultaneously perform a common large task.

2. Grid Computing

Generally in grid computing resources are categorised into four main types. These are: (Qin and Jiang, 2003)

- **Computation:** Computing resource provided by the machines' processors is the most commonly used of all resources in grids. This resource is varied in nature and this is because computation is a combination of processors with different speeds, structural designs and software platforms. A submission by an end user can utilize computation resources by working on a single machine. This submission may well be broken into several sub-tasks which will be performed at the same time on multiple machines in the grid.
- **Storage:** Storage is thought to be the second most commonly used resource in grids. The grid that has an integrated storage is formed as "Data Grid". Storage resources are generally associated with memory, hard disks, and other lasting storage media. Many file systems used on networks which have proved to be reliable and secure have been largely applied to grids. These file systems are as follows:
 - Network File System (NFS)(Sandberg *et al.* 1985)
 - Distributed File System (DFS)
 - General Parallel File System (GDFS)
 - Parallel Virtual File System (PVFS)(Ching *et al.* 2002; Zhu *et al.* 2003)
 - Andrew File System (AFS)(Howard *et al.* 1988)
- **Communication:** There has been a rapid growth in communication capacity of grids. The high performance of data grids highly depends on the available bandwidth for moving resources and tasks among the trust domains in the grid. To ameliorate the reliability of the grid, it would be advantageous to use redundant communication paths that have the capability to lighten the weight of network burden resulted from too much data traffic. Sometimes a monitoring system is developed for the management and discovering of communication gridlocks in a grid.
- **Software:** Where there is not, same software on every machine in the grid, the grid scheduler, along with balancing the grid load, may allocate jobs requiring the specific software to a machine that has that software installed on it and therefore is able to perform that job. The total scientific data created by simulations or gathered from major experiments is generally big, and such data is usually geographically stored over wide-area networks for the sake of large-scale teamwork.

3. Security Issues on the Grid

Grid applications are characterised by the coordinated use of resources from different administrative domains. Figure 1 illustrates this co-ordination by showing the policies and platforms in each domain.

Each site in the VO is independently administered and has its own local security solutions such as Kerberos and PKI. These solutions are built on top of different

platforms such as UNIX (UNIX, 2006), Windows (Microsoft, 2006) and OS2 (IBM, 2006).

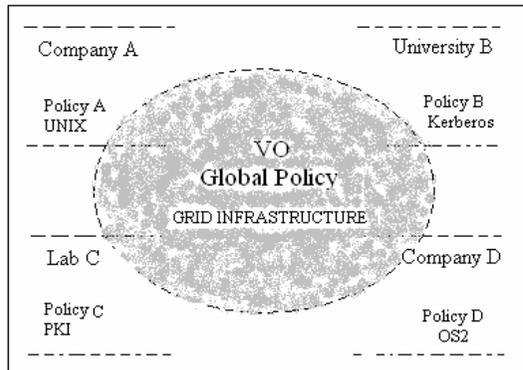


Figure 1: Reconcile local policies with Global policy (Haidar, 2002)

When these organisations are collectively collaborating on a common project in these diverse conditions many security issues emerge:

Interoperability: Because of technical, political and financial reasons it is impossible to change the security mechanisms at each site in the VO. Therefore, the Grid's security must interoperate with the local security solutions at different levels

Policy level: Each institution in the VO maintains its own security policy (Figure 3.1) which is vigilantly adopted to maximise the protection of its resource. The major issues are:

- How to reconcile global security policy with local security policy
- What are the solutions to conflicts between local and global policy, in other words which policy will apply, local or global

Authentication level: There have to be particular mechanisms in VO sites to identify users from one security domain to another. For instance the identity of a user from company (U.A) and its credential as articulated in policy are worthless in another VO sites. Therefore, how does U.A authenticate (i.e. UNIX login) to site B to access resource (R.B) (i.e. Kerberos)?

Authorisation level: Access control systems also differ from one VO site to another according to the type and value of the resource contained. As an example site A may have an Access Control List (ACL) system or a Role Based Access Control (RBAC) system to grant access. The first problem is how to determine whether a user, U.A that has been authenticated in site B is granted access to resource, R.B in B. The second is who decides what the access rights of U.A are.

Confidentiality and integrity issues: Grids allow users to transmit data over the Internet and also allow them to access remote data resources and run programs on remote sites. This brings about the issue of confidentiality and integrity that should:

- Protect transmitted data over internet
- Guarantee the privacy and accuracy of the results of programs executed on remote sites.
- Ensure the privacy and precision of the data resources shared between the users.

Firewall: A commonly encountered problem on the grid is firewalls. VO collaborators wish to share their resources with other members. However, they also want to keep some of their other resources confidential. Members of a grid must therefore allow requests from and replies to jobs initiated from other sites to pass through their firewall to access their resources. This surely is another vulnerability to the local security of the VO members' organisation. As for businesses it is very unlikely to compromise local security and because of this companies may actually end up without collaboration.

4. Grid Security Infrastructure

Security is an imperative factor in grid environments. For the users who intend to execute a task on a remote grid site, it is very important that the remote system is secure so others cannot gain access to their data. On the other hand security is equally important for the resource provider, who permits tasks to be performed on their systems. They need to be sure that those tasks would not and cannot corrupt, interrupt or access other confidential data on their system.

Apart from these two concerns, the grid environment is also exposed to all other security issues that exist in distributed computing environments. The Globus Toolkit, at its core holds the Grid Security Infrastructure (GSI), which offers a lot of services to facilitate management of the security conditions within the grid environment. Security thus, has to be taken into account when developing applications targeted for a grid. Authentication, authorisation, data encryption and secure communication in grids are absolutely dependant on the security functions within the grid environment.

The GSI has an OpenSSL implementation. It also maintains a single sign on method so when a user is authenticated, a proxy certificate is made and used when executing actions on the grid. Before designing a grid system a GSI sign in can be used to provide access to the portal or otherwise a personal security for the portal can be obtained. The portal would then be accountable for signing into the grid, either by the user's credentials or by a common set of credentials for all authorised users of the portal.

Building grid environments by the GSI components require a number of keys for public key cryptography as well as the certificate from the CA including a copy of the public key of the CA. Figure 2 demonstrates the procedures to create the GSI communication:

- Copying the CA's public key to grid host on which the GSI is established.
- Creation of private key and a certificate request

- Diverting the certificate request to CA by an email or other more secure means if running a production system and require proper identification of the sender.
- CA will then sign the request and sends it back to the grid host.

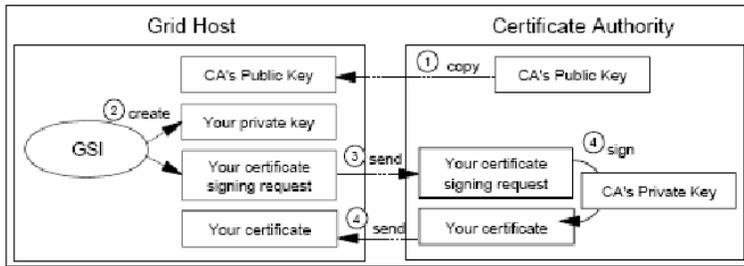


Figure 2: Preparation Procedure for GSI (Ferreira *et al.* 2003)

Once, the preparation procedures have been completed and a signed digital certificate is received, there will be 3 essential file on the grid host namely:

- The digital certificate of the grid host
- The private key of the grid host
- The CA's public key

Provision of secure authentication and communication in grid computing, emphasises on not letting others get access to the private key of the grid computer. An extra level of security has been added to the private key, which consist of a secret password that has to be used while using private key with the digital certificate. This effectively prevents others from stealing the digital certificate. The grid host's public key is made secure by the local operating system privilege inside the grid server.

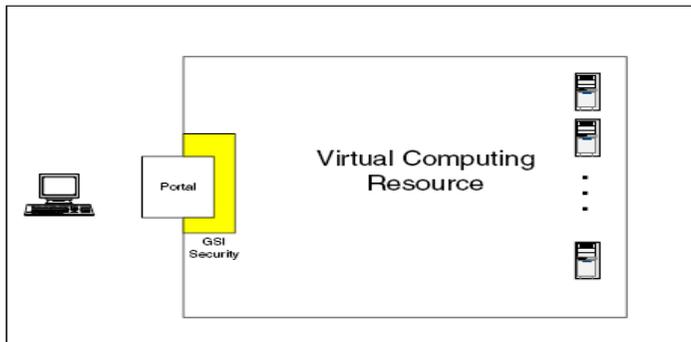


Figure 3: Security in a grid environment (Ferreira *et al.* 2003)

5. Recommendations

Having a secure grid involves significant challenges in relation to its design, operation, formation and deployment and because of rather recent emergence and prevalence of grid computing and its relative newness there is a shortage of useful

guidance on this matter. As it was established in this paper the general and basic security components for grids are:

- Grid project must provide contact information of a person responsible for the grid's security.
- The system that runs the grid must undertake a risk and vulnerability check/assessment every two years.
- System logging must be permitted on the host system and duplicated on a hub log server.
- The system log/audit track must contain for every application: application name, remote host address, remote user, authentication success or failure, user identity that is authenticated and authentication token category (password, X. 509 certificate, Kerberos, etc...).
- Login and office data resources must be saved for at least 90 days.

6. Future expectation of grid systems

- Broad and continuing investment in computational grid environments
- Rapid development of technology (e.g. new versions of Globus, Permis, Condor, etc.)
- One probable trend is: Continuing convergence of grid and web services
- Another trend is: discontinuation of Globus and comparable grid toolkits and their replacement by new generic commercial web services operations. However, the fundamental elements will remain pretty much the same.
- Grid security may continue its existence but with rather some specialist requirements; or it may be subsumed into industry paradigm security.
 - Probably relies on employment of thin-clients
 - Growing use of peer-to-peer systems, sensors and impediment of tolerant networks
- Powerful drive towards the computer science research
- Huge new number of grids can take shape which require promotion

7. Conclusion

During the last decade the grid concept has quickly developed from a limited non-scalable series of informal point to point connections with inefficient functionalities toward a scalable dynamic VO that offers much better set of functionalities. The potential advantages of the grid and generally VOs are abundant. However, the biggest difficulty for their broader implementation is security. At present only an inadequate concept of VO is recognised by the Grid regarding collaborations for scientific purposes. Nonetheless, the domains of potential applications are massive: VO adoption varies from e-learning, e-government and health to military and coordination of multinational forces. Grid security has considerably progressed forward in recent times in many aspects including authentication and confidentiality. However, still there are many inefficiencies and deficiencies and grid security is yet to convince the commercial world. The new pioneering applications of PKI and cryptography have resulted in significant enhancements in the development of security solutions for such aspects.

8. References

Ching, A., Choudhary, A., Liao, W., Ross, R. and Gropp, W. (2002) ‘Noncontiguous I/O through PVFS’, *Proceedings of 2002 IEEE International Conference on Cluster Computing*.

Ferreira, L., Berstis, V., Armstrong, J., Kendzierski, Mike., Neukoetter, A., Takagi, M., Bingo, R., Amir, A., Murakawa, R., Hernandez, O., Magowan, J. and Bieberstein, N. (2003) ‘Security’ in: *Introduction to Grid Computing with Globus*, International Business Machines Corporation, USA: 51-69

Haidar, A. N. (2002) ‘Critical Evaluation of Current Approaches to Grid Security’, [online]. Available HTTP: myweb.lsbu.ac.uk/~haidaran/Grid%20Security-Ali-N-Haidar. Pdf. (Accessed 24 July 2006)

Howard, J., Kazar, M., Menees, S., Nichols, D., Satyanarayanan, M., Sidebotham, R. and West, M. (1988) ‘Scale and Performance in a Distributed File System’, *ACM Transactions of Computer Systems*, 6(1): 51-81.

IBM (2006) www.ibm.com/os2 (Accessed 08 June 2006)

Microsoft (2006) www.microsoft.com/windows (Accessed 21 March 2006)

Qin, X. and Jiang, H. (2003) ‘Data Grid: Supporting Data-Intensive applications in Wide-Area Networks’, [online]. Available HTTP: www.cs.nmt.edu/~xqin/pubs/tr03-05-01.pdf. (Accessed 28 July 2006)

Sandberg, R., Goldberg, D., Kleiman, S., Walsh, D. and Lyon, B. (1985) ‘Design and Implementation of the Sun Network Filesystem’, [online]. Available HTTP: www.citeseer.ist.psu.edu/sandberg85design.html. (Accessed 11 May 2006)

UNIX (2006) www.unix.org (Accessed 29 June 2006)

Zhu, Y., Jiang, H., Qin, X., Feng, D. and Swanson, D. (2003) ‘Improved Read Performance in CEFT-PVFS: Cost Efficient, Fault-Tolerant Parallel Virtual File System’, in proceeding of *IEEE/ACM CCGRID Workshop on Parallel I/O in Cluster Computing and Computational Grids*, Japan.

A Performance of network coding in randomised settings

I.C.Tjhai and L.Mued

School of Computing, Communications and Electronics,
University of Plymouth, Plymouth, United Kingdom

Abstract

The scope of network coding has reached many area of networking systems with more promising benefits for the future of communication networks. This paper presents the performance of randomised network coding in both static and ad hoc environments. It demonstrates the advantages of randomised network coding versus randomised flooding algorithm. Results show that within a sufficiently large network size, randomised network coding outperforms randomised flooding algorithm. These results also show that the superior performance of network coding is achieved within a higher value of maximum degree for its incoming and outgoing links. It proves that the use of field size as small as four is adequate to support the encoding process.

Keywords

Ad hoc network, Flooding algorithm, Network coding, Static network

1. Introduction

The issue of transmission in communication systems has emerged the concept of network coding introduced by Ahlswede et al. (2000). Network coding has seen a superior opportunity to recombine several input packets into one or multiple output packets (multicasting) by treating the information packets as mathematical entities which can be combined linearly (Li et al., 2003). The principle is to transmit the information as a vector of bits (symbols). Thus, this is contrary to the traditional routing system where the packets is simply copied, forwarded or routed to other receivers.

Figure 1 gives a simple example of traditional routing systems and new systems with network coding. In Figure 1(a), the intermediate node C received data X and Y at a time but node C can only transmit either data X or Y to both receivers. Unlike this, the intermediate node C on Figure 1(b) performs coding to transmit a function of $X+Y$ carrying both data and transmits to node D and to both receivers.

From this idea on, network coding has opened up more interesting researches with many kinds of network environments and topologies such as in static, wireless or sensor networks (Fragouli et al., 2006). Optimality of network coding solution both in static and ad hoc networks have been proved as polynomial time which is against a traditional routing solution, the NP-hard Steiner tree-packing problem (Li et al., 2005).

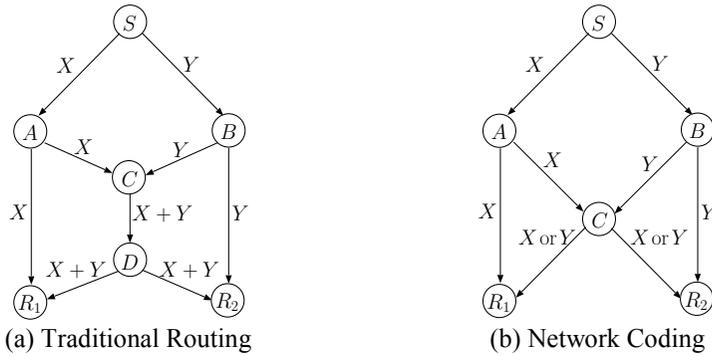


Figure 1: Comparison of traditional routing system and network coding approach

This paper presents the analysis of distributed randomised network using randomised network coding and randomised flooding algorithm. The randomised network considered here is a randomly generated network graph over a finite geographical field. In addition, the contribution of linear algebraic approach (Koetter and Medard, 2003) and distributed randomised approach (Ho et al., 2004) are also included as the concepts for simulation and comparison with the flooding algorithm.

2. Background and Related Work

The term of network coding was first introduced by Ahlswede et al. (2000). They discussed the encoding and decoding at intermediate nodes and the information rate to each terminal as the minimum cut between a source to each terminal. Their results demonstrated that network coding could improve optimal throughput over routing. Li et al. (2003) showed that optimality of multicasting could be gained via linear coding and gave practical encoding and decoding algorithms. Another research was conducted by Koetter and Medard (2003) who had extended the work of Li et al. (2003) by presenting an algebraic linear concept for reaching optimal capacity of a network. They also proved the effectiveness of time-invariant of the Min-Cut Max-Flow theorem which maximised the robustness of networks, including arbitrary networks, networks with delay and non-delay and also cyclic networks.

In randomised networks, Ho et al. (2003) had found a randomised coding approach for creating more robust and distributed transmission with compression of information. The optimum capacity of this approach depended asymptotically on the code length. Ho et al. (2004) had also come out with another work on ad hoc network coding in dynamic environments. A number of papers had also discussed the potential habitat of network coding in wireless or dynamic settings such as Wu et al. (2005) on minimal energy per bit used the total cost of multicasting. For undirected and also directed networks, Li et al. (2005) had presented the optimality of network coding was closely related to polynomial time and was contrary to routing concept as the NP-hard problems.

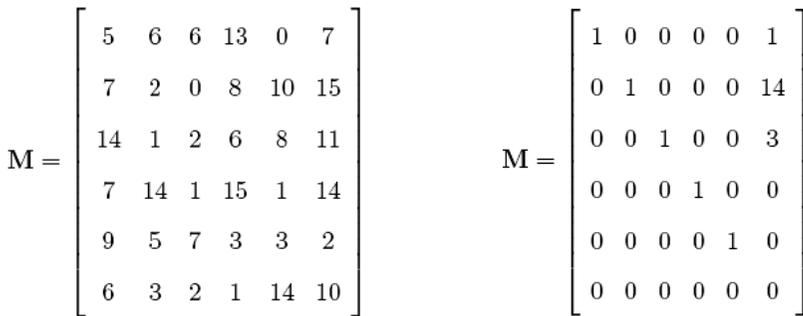
3. Model Formulation

The network model is a directed acyclic network represented by a graph $G = (V, E)$ with V as the vertex set and E as the edges set. A vertex (v) depicts a node in networks. An edge (e) is a communication channel or a link from which information can be transmitted to other nodes. The node or vertex (v) has its own in-degree $\gamma_{in}(v)$ and out-degree $\gamma_{out}(v)$. The network has a set of discrete random processes X_n and $Y(e_i)$. These random processes consist of two portions, coefficients portion and data portion. X_n represents the source packets and $Y(e_i)$ represents the packets carried by the edges. $Y(e_i)$ is also a linear combination of random process $Y(e'_i)$. Thus, the encoding equation (Koetter and Medard 2003) can be written as:

$$Y(e_i) = \sum \alpha_{n,e_i} X_n + \sum \beta_{e'_i,e_i} Y(e'_i) \tag{1}$$

The coefficients (α_{n,e_i} and $\beta_{e'_i,e_i}$) on the above equation are in polynomial form and are randomly chosen over the finite field, F_2^m . With this equation, the outgoing packets are linear combinations of original packets where multiplication and addition are also performed over F_2^m .

At receiver end, the receiver has to decode a set of linear equations which is the received random processes or packets. If it receives some packets which are stacked into a matrix M , as illustrated on Figure 2(a), the original sent packets can be decoded by using Gauss Jordan Elimination process to find the rank of matrix M . The Gauss Jordan Elimination process (Simmons, 2006) is a method of solving linear equations or finding a matrix inverse (M^{-1}). The existence of this inverse can be found by transforming matrix M into a reduced row-echelon form or an identity matrix, as shown on Figure 2(b). Hence, the reduced row-echelon form performs also the rank of matrix M or the number of the decodable packets which is five decodable packets.



(a) Before Gauss Jordan Elimination (b) After Gauss Jordan Elimination

Figure 2: A matrix M contains a set of received packets which has rank five or five decodable packets after Gauss Jordan Elimination process

On simulation, it assumed that the packet has only the coefficient portion. The reason is that running Gauss Jordan Elimination process on the entire packet (coefficients

and data) will give the same result as running this process on the coefficient portions only.

4. Simulation Setup

The simulations are run on a randomly generated network graph over a finite geographical field. This field is a defined area within x and y axes. Figure 3(a) gives an example of randomised network model created on the simulation and Figure 3(b) shows the encoding process based on the equation 1 above. By setting the coordinate x and y , each node of the network can be randomly formed. This same network graph is used for both randomised network coding and randomised flooding algorithm in static and ad hoc settings. In order to model a realistic ad hoc network, the simulation has been set to give a probability for some of the intermediate nodes to turn off. In other words, the network has some missing nodes which are not able to receive and transmit data. Initially, the network model is generated with these following parameters: number of nodes n , number of sender nodes, number of receiver nodes, maximum degree, and seed. The maximum degree is a value used to limit the number of edges incoming into and outgoing from a node. The seed is a value used to generate the random number of coordinate x and y . Then, each node can be assigned an identification number from 0 to 10. Since it is an acyclic network, the sender nodes have to be in lowest identification number and the receivers have to be in the highest number. Thus, only nodes with smaller identification number can link to nodes with higher identification number.

In the construction of node edges e_i for all types of node, it has to be within the limit of maximum node degree. For sender nodes, they are allowed to have both incoming and outgoing edges or only the outgoing edges. For intermediate nodes, they must have both the incoming and outgoing edges. Then, the receiver nodes can also have both the incoming and outgoing edges but they must have the incoming edges.

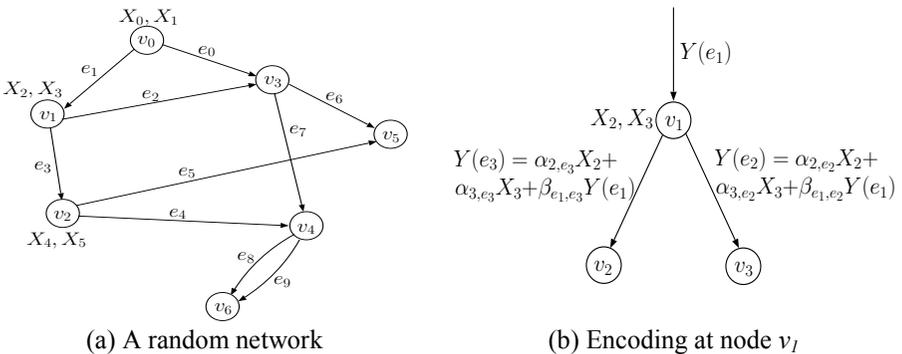


Figure 3: A sample of random network model with the encoding process at a node

The transmission is assumed as a simultaneous transmission. It means that in a specific time or in one hop, all the nodes are transmitting packets to other nodes. The capacity of each edge is assumed to be the same. The number of packets or random

processes at a sender node depends on the number of outgoing edges of that sender node. This implies that every sender node may have different number of packets.

For randomised network coding, the random processes carried by edges have been linearly encoded with the coefficient generated from F_2^m . Each node is assumed to have its own incoming buffer and outgoing buffer. A packet carried by an edge is retrieved from the edge's source node outgoing buffer to the edge's destination node incoming buffer. The packet is pushed into the buffer row by row with the last coming at the last row. For flooding algorithm, it has also the same buffers but no encoding and decoding processes. It just simply passes the packet to other nodes within the capacity of the edges. Thus, if the capacity is assumed as capacity one, only one packet can be transmitted by each edge. If a node has parallel outgoing edges which is destined to the same node, it is allowed to transmit different packets for each edge. In contrast, it transmits the same packet via each edge if no parallel edges.

The same procedure is continuing again until some specific hop when the outgoing buffer of each node of network coding has nothing to send. Then for flooding, it stops when the receivers have received all the packets.

5. Results and Discussion

n	s	r	d	Seed	Number of Missing Nodes	Average Throughput (packets/receivers)			
						No Coding	No Coding (Missing Nodes)	Coding	Coding (Missing Nodes)
8	3	2	2	464	1	3.900	4.000	2.000	2.000
8	3	2	3	4646	1	6.025	5.425	5.775	3.925
10	3	3	2	65	1	4.033	3.650	2.650	1.967
10	3	3	5	986	1	9.333	7.717	10.967	8.467
15	3	3	2	132	3	4.550	3.167	3.950	1.667
15	3	3	4	54	1	9.067	9.450	11.533	9.583
20	6	5	6	32	1	16.910	15.840	20.200	17.470
20	6	6	6	896	1	19.533	18.333	26.117	24.108
22	7	6	6	197	4	17.617	16.167	26.408	18.592

Table 1: A sample of simulation results with different average throughput gained which is generated with the following parameters: number of nodes n , number of sender nodes s , number of receiver nodes r , maximum degree d , and seed.

The simulations are run on a number of networks with different parameter combinations. Some of the simulation results are given on Table 1. Note that, the coding is for network coding and no coding is for flooding algorithm. It is clearly seen that some of the average throughput (packet/receiver) on network coding are having lower throughput than the results of no coding, especially when the number of nodes are small. However, some of the results also show that even if the number of nodes is small but if they have greater values on maximum degree, network coding still has better performance than flooding algorithm. These results imply that

network coding is more beneficial for networks with bigger number of nodes but the maximum degree value has to be sufficient enough to improve the performance. Moreover, the simulation has also come out with two important results.

Firstly, the choice of galois field (GF) or finite field size has its effect on the average throughput. As featured on Figure 4, the simulations are run on the same network parameters but different $GF(2^m)$ choices, from $GF(2)$ to $GF(2^8)$, for network coding with missing nodes and without the missing nodes. The results show that the throughput gained from $GF(2^2)$ to $GF(2^8)$ has been in a stable condition or gained the same throughput but on $GF(2)$, it is very low. This result has proved that using the sufficiently small $GF(2^m)$ for the code length has been efficient to achieve better performance as stated also by Ho et al. (2004). The reason not to opt for large $GF(2^m)$ is because it causes complexity of coding and requiring large memory for processing (Lehman, 2005). Thus, $GF(2^2)$ is the lower bound for the code length of network coding. In addition, comparing these two graphs, the throughput gained by network without missing nodes is higher than the network with missing nodes. This suggests that a network with more nodes (large network) will give better probability to obtain more linearly independent packets (decodable packets).

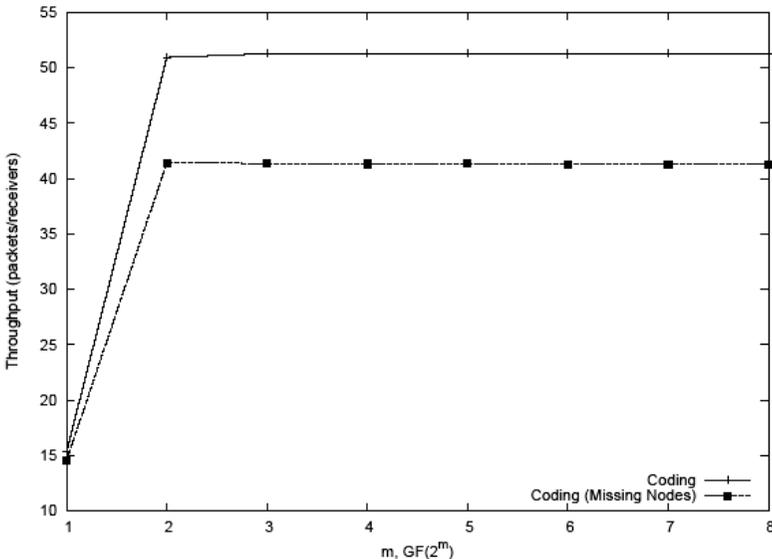


Figure 4: Effect of galois field on throughput

Secondly, as illustrated on Figure 5, the average throughput of network coding does not only depend on the number of nodes it has but also the maximum degree of the number of edges incoming into and outgoing from a node. This result has proved that a higher maximum degree parameter in network coding has effect to its performance about six times better in the throughput. In contrast, no coding network both with and without missing nodes has shown not much being effected by the increased of maximum degree. It is only about twice increased in its throughput. From these results of two different networks, it can be concluded that network coding performance depends on the value of its maximum degree. The smaller value seems

to cause the performance of network coding is more similar to no coding network. The reason is the more edges in and out of a node in network coding are giving more probability of linearly combined packets to be received by the receivers.

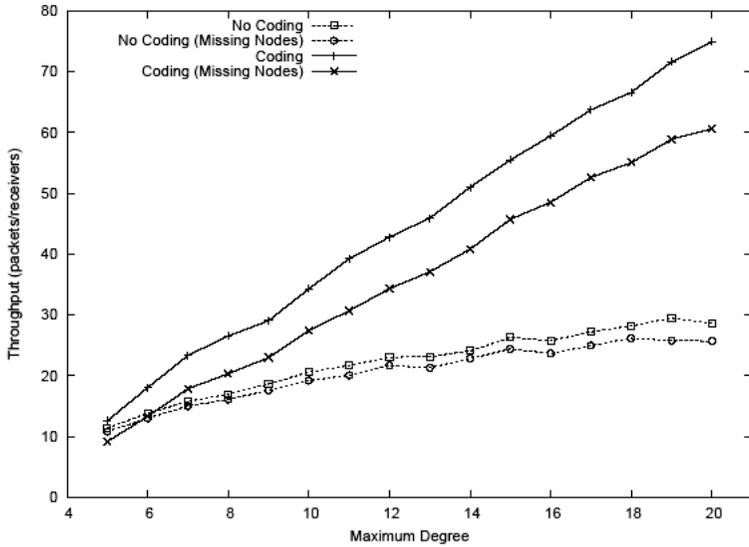


Figure 5: Effect of maximum degree on throughput

6. Conclusion

The paper investigates the performance of randomised network coding concept and randomised flooding algorithm concept. The introduced network graph has been designed to model the real world static and ad hoc networks. These two different network settings have been compared and analysed based on the two networking concepts. Results show that the optimal throughputs of randomised network coding both in static or ad hoc networks are achieved within a large network and with the higher maximum degree for the incoming and outgoing edges of a node. This higher maximum degree has given more probability for more edges' function to be transmitted over the network. This research also shows that the existence of minimal code length used for network coding which is as small as 4.

Future work includes the improvement of this simulation model and consideration of network coding with delay and with cyclic network. Extending the simulation into finding the existence of lower bounds for maximum degree to allow the network coding to achieve optimality within a small or large network size is also an interesting area for investigation. Moreover, researching the capacity of network coding to support in transmission is also one problem to solve. The issues of network security, protocol and management are also needed to be considered.

7. References

Ahlsweide, R., Cai, N., Li, S.-Y. R. and Yeung, R. W. (2000), "Network Information Flow", *IEEE Transactions on Information Theory* 46, pp. 1204-1216.

- Fragouli, C., Boudec, J.-Y. L. and Widmer, J. (2006), “Network Coding: An Instant Primer”, *ACM SIGCOMM Computer Communication Review*, Vol. 36, No. 1.
- Ho, T., Koetter, R., Medard, M., Karger, D. R. and Effros, M. (2003), “The Benefits of Coding over Routing in a Randomized Setting”, in *Proc. IEEE International Symposium on Information Theory (ISIT)* p. 442.
- Ho, T., Leong, B., Medard, M., Koetter, R., Chang, Y.-H. and Effros, M. (2004), “On the Utility of Network Coding in Dynamic Environments”, *International Workshop on Wireless Ad-Hoc Networks (IWWAN)* pp. 196-200.
- Koetter, R. and Medard, M. (2003), “An Algebraic Approach to Network Coding”, *IEEE/ACM Transactions on Networking* 11, pp. 782-795.
- Li, S.-Y. R., Yeung, R. W. and Cai, N. (2003), “Linear Network Coding”, *IEEE Transactions on Information Theory*, Vol. 49, pp. 371-381.
- Li, Z., Li, B., Jiang, D. and Lau, L. C. (2005), “On Achieving Optimal Throughput with Network Coding”, in *Proc. IEEE INFOCOM 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, Vol. 3, pp. 2184-2194.
- Simmons, B. (2006), “Gauss-Jordan Elimination”, Available from: http://www.mathwords.com/g/gauss-jordan_elimination.htm, last updated: 14 April 2006. Date visited: 10 July 2006.
- Wu, Y., Chou, P. A. and Kung, S.-Y. (2005), “Minimum-Energy Multicast in Mobile Ad Hoc Networks Using Network Coding”, *IEEE Transactions on Communications*, Vol. 53, pp. 1906-1918.

Investigation on Static Network with Network Coding

L.Yang and L.Mued

School of Computing, Communications and Electronics,
University of Plymouth, Plymouth, United Kingdom

Abstract

In this paper, we propose to give a general overview of network coding. We will start by explaining the basic ideas behind the concept and Max-Flow Min-Cut Theorem which determines the maximum capacity of the network. Then we will present the advantages and weakness about network coding. At last a simulation will be introduced about network coding in P2P file sharing system. Finally we will discuss about the results with conclusions and future work.

1. Introduction

Network coding as a new technique in coding theory is widely spread to help achieve the optimal throughput in multicast transmission. Network coding is firstly proposed by (Ahlsweide *et al.* 2000), with a network modeled by a directed graph (V, E) with edge capacities. They showed that a sender $s \in V$ can communicate common information to a set of receivers $T \subseteq V$ at rate achieving the capacity h which is calculated by Max-Flow Min-Cut Theorem. In order to achieve the optimal throughput, network coding is used to replace the traditional multicast method, by allowing encoding operation at the intermediate nodes of the network. Li, Yeung, and Cai (Li *et al.* 2003) proposed that it is sufficient for the encoding functions at the intermediate nodes to be linear. Koetter and Medard (Koetter and Medard, 2002) showed how to find the coefficients of the linear encoding and decoding functions by finding values for the indeterminate of a polynomial for which the polynomial is non-zero. They also showed that such values can always be found in a field of size $h|T|$, where $|T|$ is the number of receivers. And Jaggi from (Jaggi *et al.* 2003), Sander from (Sander *et al.* 2003) showed that for acyclic networks, how to find the encoding and decoding coefficients in polynomial time, and showed that field size $|T|$ suffices. They also showed that the linear encoding functions can be randomly designed. And if the field size is at least $|E|/\delta$, the encoding will be invertible at any given receiver with probability at least $1-\delta$, while if the field size is at least $|E||T|/\delta$, the encoding will be invertible simultaneously at all receivers with probability at least $1-\delta$. Section 2 introduces the basic concept of network coding, Max-Flow Min-Cut Theorem. Section 3 is about the benefits and weakness of network coding. Section 4 is about the simulation work in P2P file sharing and its comparing results. Section 5 is about conclusions and future work.

2. Overview of Network Coding

2.1 Basic concept of Network Coding

At beginning, we need to define a series of terminology and notations to represent the network coding issue. Firstly, we represent a point to point communication network by a directed graph $G=(V, E)$, where E is the set of edges such that information can be sent noiselessly from node i to node j for all $(i, j) \in E$ (eg: an edge represents a link between two nodes). Let V be the set of nodes of a network. Now we focus on a particular case called the single source problem, here we only have a single node as the sole source, so we present the source node as s . and it transmits packets to a set of nodes called t_1, t_2, \dots, t_L is called sinks. So for a specific L , the problem is referred to the one-source L -sink problem. And the capacity of an edge $(i, j) \in E$ is given by R_{ij} . We define $F = [F_{ij}, (i, j) \in E]$ as a flow in G from s to t_L for all $(i, j) \in E$, $0 \leq F_{ij} \leq R_{ij}$. Such that for all $i \in V$ except for s and t_L , we have

$$\sum_{i:(i,i) \in E} F_{i'i} = \sum_{j:(i,j) \in E} F_{ij}$$

It means that the total flow into node i is equal to the total flow out of the node. F_{ij} is referred to as the value of F in the edge (i, j) .

2.2 Max-Flow Min-Cut

The Max-Flow Min-Cut theorem is the theoretical method to determine the maximum throughput of the network and used to characterize the admissible coding rate region of the given network. Based on the above basic network coding concept, we have the definition for the value of F as follows:

$$\sum_{j:(s,j) \in E} F_{sj} - \sum_{i:(i,s) \in E} F_{is} = \sum_{i:(i,t_l)} F_{it_l} - \sum_{j:(t_l,j)} F_{t_lj}$$

Such that F is considered as a max-flow from s to t_L in G if F is a flow from s to t_L whose value is greater than or equal to any other flow from s to t_L . For a graph with one source and one sink, the value of a max-flow from source to sink is called the capacity of the graph (Ahlsweide *et al.* 2000).

In multicast communication system, there could be multiple receivers with different max-flows as a max-flow is defined for a point-to-point communication. In order to evaluate a unique capacity of the specific multicast communication, we define the max-flow of a point-to-multipoint communication as:

$$F_{multicast} = \min_{i \in \{t_1, \dots, t_L\}} F(i)$$

where $F(i)$ is the max-flow from s to i , and $i \in \{t_1, \dots, t_L\}$ the set of multicast receivers. According to the well known Max-Flow Min-Cut Theorem in graph theory, we have the multicast capacity of a network represented by a graph $G = (V, E)$ (it means the maximum number of bits that can be transmitted from the source to all the receivers simultaneously) is equal to the minimum value of the max-flows of

all the receivers. Such that for a multicast capacity of a network is calculated by computing the minimum value of all the max-flows individually calculated for each receiver. By defining the capacity of the multicast network, we will present an example on why network coding is required, why the current traditional multicast method can not satisfy the requirement of optimal max-flow.

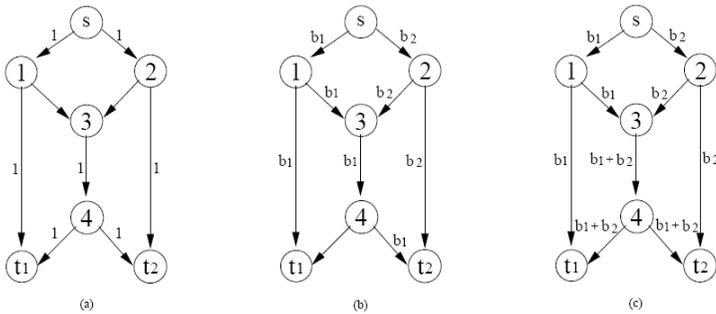


Figure 1: Butterfly One-Source-Two-Sinks Network Topology

In Figure 1, there shows a one source-two sinks graph. Figure 1(a) shows the capacity of each edge in the network with 1 unit. According to the principle, we can easily check that the value of the max-flow from s to t_L is 2, $L = 1, 2$. The Max-flow Min-cut states that we can send 2 bits to t_1 and t_2 simultaneously because of the value of the max-flows in the graph. In Figure 1(b), it shows an actual network based on the same capacity in Figure 1(a), there are two bits b_1 and b_2 generated at node s . At node 3, as the capacity between node 3 and node 4 is 1 bit, so it means every unit time, only 1 bit is allowed to send through the edge. So once b_1 and b_2 arrives at node 3, node 3 is only capable to send b_1 and b_2 at 2 unit times. As the delay of transmission, it causes the sinks could not receive the two bits simultaneously. So the capacity is actually less than the expected capacity value. So in order to achieve the optimal throughput, network coding is required to complement the weakness. As shown in Figure 1(c), it is also with the same network topology and same capacity with 1 bit. At node 3, once receiving the two bits, node 3 implements a kind of encoding method (eg: as shown is modulo 2 addition or exclusive-or) to combine the two bits into $b_1 + b_2$ or $b_1 \oplus b_2$ instead of directly store and forward. Then the combined bit $b_1 \oplus b_2$ is send to the sinks. At sink t_1 , it receives b_1 and $b_1 \oplus b_2$, and capable decode the received data into actual bits b_1 and b_2 . The same decoding occurs at sink t_2 to get the original data b_1 and b_2 .

3. Benefits and Weakness of Network coding

3.1 Advantages of Network Coding

The main idea of network coding is supposed to achieve the max-flow bound on the information transmission rate in a multicast network. In (Ahlsweide *et al.* 2000), it is

found that the traditional multicast method by simply store the information and forwarded through the intermediate nodes is not optimal to achieve the expected capacity as shown in graph. So there would be a technique (eg: network coding) to employ to achieve the optimality. In Figure 1, we can clear see that the scheme with network coding (Figure 1(c)) is capable to send faster than the traditional multicast technique (Figure 1(b)).

Another benefit of network coding that can be easily seen from Figure 1 is the saving in bandwidth when network coding is utilised. It is by combining a set of incoming data into a single one packet and send to its sinks. Instead, in traditional multicast method, the node is only simply performing the function of replicating and sending to the outgoing links. Hence, network coding is capable to save in bandwidth, as shown in Figure 1. We can see that in the traditional multicasting, a total of 10 bits are sent in the network, while only 9 bits are sent with network coding method. So we can easily see that network coding brings a 10% save in bandwidth.

Finally, network coding also could be a helpful tool in network management and network robustness. Coding is not only applicable to networks in order to achieve the capacity, but also can be used to recover from link failures in networks. The failure is considered are the long-term failures due to a link cut, or the permanent removal of an edge, or certain disconnection. At present, such kind of failures is dealt by the use of rerouting, like link or path protection. From the result in (Koetter and Medard, 2002), it shows that the network codes operating under a certain failure scenarios can be designed for recovery.

3.2 Weakness of Network Coding

The major problem with network coding is that the loss of one packet could affect a series of related packets and renders some useless information at the receivers. Such as in Figure 1(c), node t_2 requires either the bits b_2 , $b_1 \oplus b_2$ to recover b_1 . If b_2 is lost during the network transmission, node t_2 will not be able to decode b_1 even if $b_1 \oplus b_2$ received correctly. Hence, in such kind of situation, the correctly received encoded information is regarded as a loss, since the encoded information itself is considered as invalid. In other words, in network coding, one bit loss in the network could probably result in several bits losses for the receivers.

4. Simulation

4.1 Simulation Implementation

During the simulation, we concentrate our research on the application layer of P2P file sharing system. For comparison, we use a kind of transmission method used in BT application, called *Flooding* to help present the benefits of network coding. Flooding as its original idea, when a node receives a new chunk and after getting access to the medium, it blindly forwards the chunk to its neighbours even if none requires. Given that neighbours share a static network resource by randomised generation, a node may receive multiple chunks before it could perform one single

transmission. In that way, although the medium is free, the node schedules the chunks that have not been transmitted in the same order of reception, such as first received first transmitted like queue algorithm. Hence, every node could transmit as many chunks as it receives. For flooding with network coding, every node checks continually with its table to look for its neighbours who require the data held by the node. Once it finds out at least one neighbour, the node generates and transmits a combination of all what it posses. In our simulator, we split the time space into rounds each of $1/c$ unit of time. A node can transmit only at the beginning of a round. During each round there are four steps:

- 1) We identify the candidate transmitters, which are the nodes holding data to send. These candidate transmitters are placed in a candidate list.
- 2) We pick up at random one candidate transmitter, node A, and move it into a neighbour list. Then we prohibit all A's neighbours from transmitting during the current round. As a consequence, A's neighbours are deleted from the candidates list. We also delete the neighbours of A's neighbours from candidates list. The purpose is to avoid the situation where a node receives from more than one neighbour at once. Then the candidates list is accessed and a new node is moved to the neighbourlist. Such process continues until the candidates list becomes empty. In such way of transmitting, we ensure each neighbours has the same chance to access the medium.
- 3) We scan the neighbour list and all selected transmitters send one chunk each.
- 4) We update the list of chunks at all nodes.

And the four steps are performed recursively until all the nodes receive the entire file.

4.2 Results

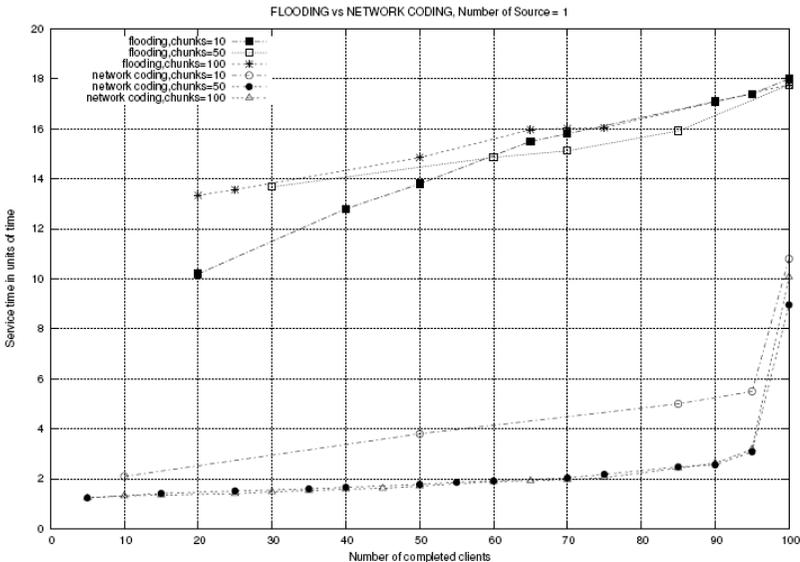


Figure 2: The results of comparison between Flooding and Network Coding with single source

By implementing the simulation, with condition of chunks = 10, 50, 100, we have the results using network coding comparing with flooding as shown in Figure 2. By implementing the simulation, with condition of flooding with multisource and network coding with single source, we have the comparing results as shown in Figure 3.

From Figure 2, the results between one source multicast in Flooding and Network coding, we can conclude that:

- The more the number of chunks is, the faster the multicasting can be established, it works in both flooding and network coding.
- Multicasting with network coding, the performance is increased by at least 5 times than the traditional way.

From Figure 3, the results between Flooding with multisource and network coding with single source, we conclude that:

- For flooding with multisource, the more the number of source is, only can slightly increase the performance of multicasting.
- The results of flooding with network coding and single source, could achieve the optimal throughput by almost 6 times than flooding with multisource.
- The increment rise in network coding during the completion of more than 90 nodes is due to the less neighbouring nodes connected with the nodes which are located at the edge of the map.

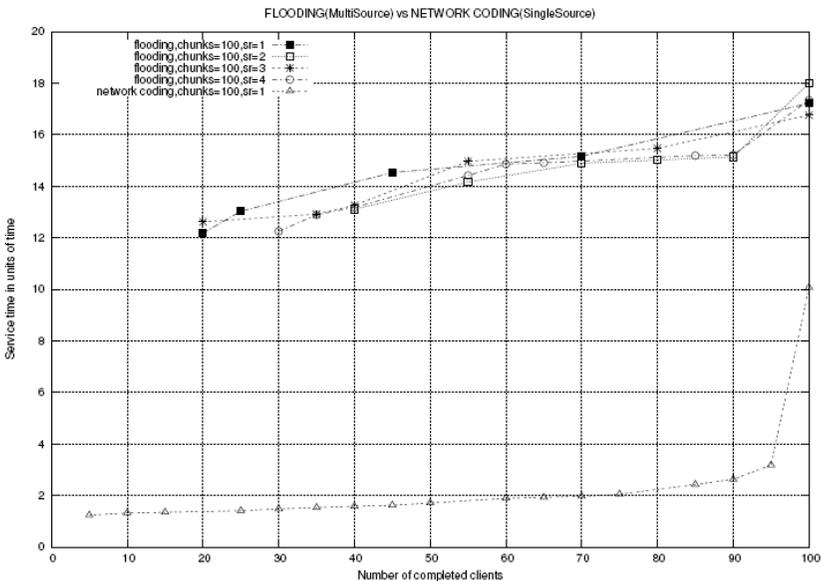


Figure3: The results of comparison between Flooding with multisource and Network Coding with single source

From the above results of simulation, we can conclude that:

- Network coding speeds up the distribution of the data information. It achieves optimal throughput comparing with flooding.
- Network coding could become more efficient when many neighbours benefit from the same transmission. And when the number of nodes in the system gets larger, node would probably have more neighbours from which to get the chunks and information. And the potential gain from each transmission would also increase.
- Increasing the number of chunks would improve the performance either in flooding or network coding. When we divide the file into multiple chunks, a node can start exchanging its chunks as soon as it finishes downloading it instead of downloading the whole file data. So we have the larger the number of chunks is, the faster the nodes could distribute its information with its neighbours and increase the system performance.

And we also can see that the transmission with network coding achieves a significantly improvement, even though a multisource method is allowed for flooding transmission, it also could not achieve the optimal throughput as network coding does.

5. Conclusions

By the study on network coding, its encoding and decoding method on nodes provides the right way of achieving the optimal throughput which is theoretically calculated. The related research work render a new view on multicasting in network, it breaks the traditional multicasting idea during a network. The deployment of network coding could significantly help improve the performance of network transmission such as in resource sharing. Comparing to the traditional routing techniques, network coding offers additional benefits, including fewer network resource consumption, ease of network management, and robustness feature. With the development of network coding, more and more areas such as ad-hoc sensor networks and P2P file sharing, will be affected to achieve an improved performance from network coding.

By our simulation work, we successfully show that network coding is capable to help improve the throughput in a butterfly network multicast with bandwidth saving in 10%. Furthermore, the results from simulation on P2P file sharing clearly indicates that network coding has the potential on improving the speed of download and the network resource usage between the sharing system. It may significantly change the method of data sharing between the individual computer users. Network coding may bring a new generation of data transmission. Fewer time expenses, fewer network resource consumption, optimal throughput performance will be its representative. Even though the downloading client is dial-up connection, network coding could help to save the time on downloading less number of chunks which have already contained the sufficient linear combination data for decoding instead of downloading the whole completed chunks.

The future work is concentrating on the selective flooding method with network coding to enable the node has the ability to identify which neighbours request the

specific chunks compared with their holding chunks. In selective method, the chance of sending the useless packets will be reduced. With network coding, the node will effectively transmit the encoded linear combinations to the requested nodes.

6. References

Ahlsvede, R., Cai, N., Li, S.-Y.R. and Yeung, R.W. (2002), “Network information flow”, *IEEE Trans. Information Theory*, IT-46(4):1204-1216.

Jaggi, S., Chou, P.A. and Jain, K. (2003), “Low complexity optimal algebraic multicast nodes”, *In Proc. Int’l Symp. Information Theory*, Yokohama, Japan.

Koetter, R. and Medard, M. (2002), “An algebraic approach to network coding”, *INFOCOM*.

Li, S.-Y.R., Yeung, R.W. and Cai, N. (2003), “Linear network coding”, *IEEE Trans. Information Theory*, IT-49(2):371-381.

Sander, P., Egner, S. and Tolhuizen, L. (2003), “Polynomial time algorithms for network information flow” *In Symposium on Parallel Algorithms and Architectures (SPAA)*, pages 286-294, San Diego, CA.

A guide for small and medium enterprise of implementing security and firewall system

R.Zhang and P.S.Dowland

Network Research Group, University of Plymouth, Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

The aims of project were try to produce training tools to assist network administrators with limited specialist security knowledge in the selection, installation and configuration of firewall systems. The project collected customer's expectation data by questionnaire and real companies' interview methodology, with analysis current arts of firewall technology and available commercial firewall products, the project chose SmoothWall Express 2.0 as firewall system for SME and produced a step by step training guide including installation and configuration parts. The project introduced basic terminology about security and firewall in order to end users getting start of the training guide. Before the training guide producing, the project did firewall functions and security test in order to make sure the SmoothWall can fully satisfy customers' requirement, and the final of the project reviewed real companies expectations and pointed out general weakness of firewall systems and give extra suggestion about firewall and business application. The project compared and discussed similar functions techniques, in order to help SME administrator does not confused between those technologies, and gave some suggestion about how to choose the appropriate techniques for real business environment.

Keywords

Network, Security, Firewall, Service

1. Introduction

In a recent security survey reported by Penn, Scheon and Berland Associates, around half of small and medium enterprise (SME) respondents said the security of their IT systems been threatened in the past year, another research group Gartner reported SMEs are often with limited budget for in-house IT manpower, they can not afford staff who has sophisticated security knowledge and experience, the report pointed out the security fact, half of SMEs are likely targets for attackers and 60% of them will be unaware of the attacks.

From above information, we learned that the internet for SME is so dangerous, if they just subscribed the internet connection from an ISP (Internet Service Provide) without considering any protection device/method between your LAN (Local Area Network) and the internet, that situation is very horrible dangerous, all of your devices connected to the internet are public for all of the internet users, they only need download a very small program, then they can control the LAN in few minutes. They can do anything as you can, or even they will do something but you do not know how to do. When connect the internal LAN to the internet, you are putting three things at risk:

- Your data: Your computers stored data and outbound and inbound data
- Your resource: the computers themselves and its peripherals
- Your reputation

2. How to protect

Above context, we have introduced firewall can be considered as a guard to check inbound and outbound information to protect your network, however, network firewall can not be perfect unless you stop all of the data transferring, Zwicky et al. (2000) said “a firewall can not fully protect against viruses” and Operating System’s bugs can be used by attackers. Virus always pretend like a part of programme or language code, they can cross the firewall with reasonable service requirement, detecting a virus in a random packet of data passing through a firewall is very difficult, it requires:

- Recognizing that the packet is part of a program
- Determining that a change in the program is because of a virus
- Determining what the program should look like (Source: Zwicky et al., 2000)

For example, virus can be compressed into RAR/ZIP file, we can not expect firewall can detect such type files from each transmitted packet, even they can, how about other kind of compression. Also the virus can be from green zone, a piece of infected CD or floppy disk, such as spy virus or Trojan horse virus. We have to install Anti-virus software for each computer/server, and the Anti-virus must have automatic update function. And Windows has a lot of security holes, without hole-fix program, like Windows XP Service Pack 2 (sp2), Firewall + Anti-Virus will become adornments.

Firewall can be a guard, but firewall is not automatic protection. Pohlmann et al. (2002) described “A firewall system does not provide automatic protection; rather, protection is possible only if a firewall system is correctly operated.” They suggest that a security policy must be developed and implemented before a firewall system can be used. To produce the security policy, we need to know who the users are, what should be prohibited, what network protocol will be used, different privileges for different group users, etc.

3. Different types of Firewall

Home PC Firewall Guide divided security defence lines into three layers, first, Choose an Internet Service Provider (ISP) or an email service that offers online (server side) firewall or virus email filters. This will block infections before downloading them. However, this defence line is not very stable and we do not have widely choices, and normally, the ISPs only can offer very limited performance or uniform configuration. Second, to install a wired or wireless hardware router with a built-in firewall between your modem and your computer or network. For this line, we also can use dedicated firewall server, like SmoothWall. Third, Personal firewall

software on every computer on your network, commercial products like Microsoft XP built-in Firewall, Norton™ Personal Firewall and ZoneAlarm or BlackICE.

From technologies aspect to approach firewall, Proctor et al. (2002) divided firewalls into three fundamental technologies, Network layer, Application layer and Hybrid. Or other terms describe these technologies include packet filters, application gateways, and stateful packet inspection. Network layer firewall works at three layer of Open Systems Interconnection (OSI) model, it led by Check Point, looking for faster technology and greater flexibility, it operates as a packet filtering, check the packets traffic based on source and destination information, for example rule as figure 1:

Action	Source	Port	Destination	Port
Allow	Any	TCP 80	Web Server	TCP 80

Figure 1 An example of firewall security rule

However, packet-filter firewall does not check detail information, such as the state of communications or application information. Another problem is the “rule” table maybe very huge, without carefully configuration it easy to make mistake. Proctor et al. (2002) concluded “Network layer firewalls are very fast, relatively inexpensive, and the least secure of all firewalls.”

Application layer firewalls are also known as application gateways and proxies. Pohlmann et al. (2002) described application gateway does not just check addresses of inbound deliveries, it opens every packet, examines its contents, and checks the shipping documents prepared by the originator against a clearly defined set of evaluation criteria. The security check at this point is significantly more reliable than packet filtering. However, the check takes longer that packet-filter firewall, for those willing to trade some performance for enhanced security, application firewall may be the good choice.

Hybrid firewalls typically combine characteristics of both network level and application level firewalls to give an improved balance between performance and security. Proctor et al. (2002) described hybrid firewall is a state- or session-aware and performs packet filtering but does not act as proxy. It can provide adequate security and the performance between packet filters and application gateways.

4. Firewall questionnaire and interview

In order to understand SME expectation for firewall, the project did two types of data collection, 1. Sending firewall questionnaire to random people but not the students/staff who from University Technology/Computer department. 2. Interview two real small companies and discuss with their IT administrator to evaluate real company’s requirement.

4.1 questionnaire results

The questionnaire includes 17 questions, and be asked 20 people to fill the questionnaire with face-to-face mode. The questions were designed into three main parts: General part, network and firewall part and firewall training expectation parts.

General part provides information about the fact of average security knowledge level, the average results show people have some computer knowledge, but not network security knowledge, they normally do not understand basic network and firewall terminology. Some people familiar with words: firewall of security. But they normally do not know what type firewall they are using, they also do not know the difference between firewall products. The most of people are using Windows based operating system, only a few people have tried Linux based before.

Network and firewall part shows information about what kind of service are expected, TCP based service more popular than UDP service. The most of them have no idea about services which out though the firewall. The most of people like to spend less than 50 pounds per year for firewall system. They expect a secure, cheap, easy installation and controlled firewall product.

Firewall training part provides information about, the most of people expect the training Guide can provide step-by-step installation and configuration with useful functions guide. They do not mind about presentation media but they expect rich pictures of screen shot, and they hope they can do practice during training.

4.2 Interview

The project interviewed two China's companies, ShanXi DaRen Education and ShenZhen HuaYi digital. They have around 20 staffs for each company, DaRen prepare to use broadband in recently, HuaYi has already connected to the internet. We are going to discuss their situation.

Case 1: DaRen is a company offering foreign language training. The project interviewed DaRen's Manager, he introduced they have around 20 computers want to connect the internet, all computers are Windows based operating system, he expect all of computers can share the internet for HTTP and MAIL Service, and the company needs a firewall to protect their network. However, they do not have in-house IT staff to conduct this task. The company like to invest around 300 pounds to purchase all of networking devices, 150 pounds for each year to maintain those devices. He answered about firewall requirement: fully protection, LAN to internet access, Voice Over IP (VoIP), he also interest for a firewall training program, he hope the training could based on non-computer background people.

Case 2: HuaYi is a company for doing digital video and audio, they are using ADSL as internet connection. Due to ADSL router setting, outside of company can not access company's LAN. Their staffs want to remotely access company's database when they away from company. The company only have an IT staff charging all IT staff, but he is specialist in Web-Design. He likes to join in a firewall training, and expect the training should not too longer, and training should include some basic

terminologies could be used in the firewall system. The company would like to spend 200 -300 pounds to improve the networking situation.

Above two interviews reflect the fact of SME requirements and budget ranging for the whole devices. With questionnaire results, we can divide internet application into two main groups: LAN to internet, internet to LAN

5. SmoothWall test and results

The project chose SmoothWall for SME, before the Guide we need to do some test in order to make sure SmoothWall can fully satisfy customer requirements. The project will use VMware software to simulate internet environment to test the firewall. The test bed contains 5 virtual machine based on two real network cards, and each card connect to different switch. The test bed structure likes figure 2 shows:

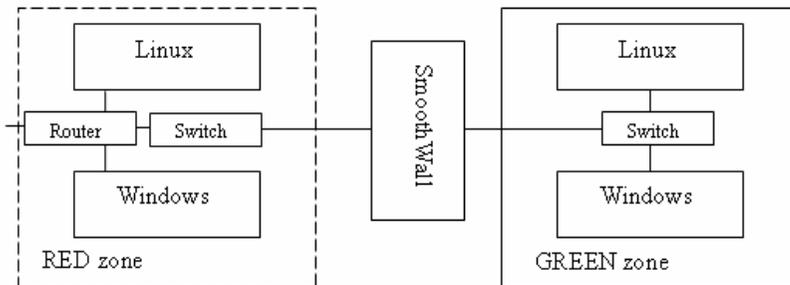


Figure 2 Test bed structure

The network divide into two zones, GREEN and RED. Each zone contains 1 Linux computer and 1 Windows computer, each machine has installed HTTP server and FTP server in order to test firewall functions. All computers setting as below:

GREEN: Subnet mask: 255.255.255.0, default gate way: 192.168.3.1
SmoothWall Interface: 192.168.3.1
Linux: DHCP/192.168.3.199
Windows: DHCP/192.168.3.200

RED: Subnet mask: 255.255.255.0, default gate way: 10.0.0.2
SmoothWall Interface: 10.0.0.10
Linux: DHCP/10.0.0.3
Windows: DHCP/10.0.0.4

All RED zone can be consider as Internet environment, we need to simulate port forwarding test and attack test from the RED zone. We also need GREEN to Internet service, so we have set all 4 computers as servers.

We are going to give very brief test results report, the purpose of test is used for examination of firewall functions rather than demonstration. =

DHCP Server: Linux and Windows are both assigned IPs information by SmoothWall Server, DHCP => PASS

NAT server: Linux and Windows are both can access internet, NAT server=> PASS

Proxy Server: Linux and Windows setting proxy server address: 192.168.3.1, port: 800, then can access internet, proxy Server => PASS

Port forwarding: add rule in SmoothWall, forwarding 192.168.3.200:80 port to 192.168.3.1. RED zone access 10.0.0.10, and opened 192.168.3.200's web page. Port forwarding => PASS

IP block server: add rule in SmoothWall, block RED Linux IP 10.0.0.3, and using 10.0.0.3 to access 10.0.0.10, access denied, and remove the rule from SmoothWall, 10.0.0.3 can access 10.0.0.10:80 (192.168.3.200) port. IP block server => PASS

Logs service: Access SmoothWall web administrator interface, click logs, you will find comprehensive logs recorded by SmoothWall, it can be used for security and networking analysis. Logs service => Pass

6. Port forwarding Vs DMZ

DMZ exposes all of ports of server, it means DMZ has no firewall protection at all. And only single server available in the ORANGE zone, you need to build all of service into one server, it reduces network structure flexibility. Or you need to set up “dmz pinholes” for forwarding GREEN's ports to ORANGE zone, but similar complex as “port forwarding” setting. Using port forwarding you can only forward wanted services port to the internet, and you do not need to build a ORANGE zone, the services port from GREEN zone.

The choice between DMZ and port forwarding depends on your network requirement and existing network structure. You need to consider your broadband upstream speed, for ADSL connection, downlink and uplink have different data rate, ISPs normally provide 256kbits/s speed for upstream even you may have 8mbits/s download speed, you can not expect using ADSL to build a server for public, but these two techniques help you to easily build a mobile office in your home.

7. Proxy firewall and packet-filter firewall

We have introduced application level proxy and packet-filter firewall: they both have advantage and disadvantage. Norbert etc compared two solutions: Packet-filter can provide higher flexibility, faster process speed, full application supporting. Application proxy can provide higher secure, better management ability, require identification and authentication before transporting. Xinli website described, Proxy server will dramatically increase server usage when high inbound and outbound more than 75MB/s, nowadays, ISPs can provide highest internet connection is 8MB/s, you do not worry about 75MB/s bottle neck, but faster and more RAM is recommend for using proxy service.

8. The weakness of firewall systems

Oppliger (1998) described, “firewall can not protect against insider attack”, firewall only checking when the packet pass through the firewall, for insider communication packets will not need to go through the firewall, no device will stop inside attack. To resolve this problem, we can add extra firewall for higher security department, like financial department. Central firewall + ICF for each computer is a good practice for organisation.

Firewall can hardly detect virus pass through it, especially when the data be compressed, it is impossible to build a scan to check virus signatures, compression with different algorithm can produce any number of a packet format. The efficient way to detect virus is installation of an anti-virus program, the software can check full file rather than packets.

Firewall can not improve the rules by themselves, firewall will do exactly things as you set, a poor setting will results higher vulnerability and firewall can not fix the poor setting. You need to maintain the firewall regularly, you need to view and check firewall logs information, you also need to understand the users expectation, under company policy to improve security policy.

Firewall can be bypassed, in the GREEN zone, the end user may have different internet connection as you provided, a modem connection or a remotely VPN with dual sub-net, other people will have chance bypass you firewall server, firewall can not provide protection that not pass through it, and you whole GREEN may will have a directly link to the internet. You need to pay attention for the security policy again, security policy needs to be discussed with users or you need to force users following the rules.

9. Conclusion

The paper introduced an overview of SME security aspect, briefly discussed SME currently faced problems, and three methods need to be adopted to resolve the problem. Three protection methods must be used at same time. The project did questionnaire and interview for collecting customer data, and the project chose SmoothWall as firewall system for SME, some functions and security test has conducted by the project. The paper discussed some functions may confused SME, and pointed out the weakness of firewall system, also gave the solution to overcome those drawback.

10. Reference

Firewallguide (2006), “Home PC Firewall Guide”, www.firewallguide.com, (accessed 31 July 2006)

Gartner Group (2000), NetworkWorld, “Half of small, midsize enterprises will suffer Internet attack”, www.networkworld.com/news/2000/1011attack50.html, (Accessed 31 July 2006)

Oppliger, R. (1998), Internet and Intranet security, Artech House, Inc.

Penn, Scheon and Berland Associates (2005), SOHOWare Inc., “BroadScan TM solutions overview”, www.sohoware.com/support/pdf/BroadScan_Solutions_Overview.pdf, (Accessed 31 July 2006)

Pohlmann, N. and Crothers, T. (2002), *Firewall Architecture for the Enterprise*, Wiley Publishing

Proctor, P.E. and Byrnes, F.C. (2002), *The Secured Enterprise Protecting Your Information Assets*, Prentice Hall PTR

XinLi (2004), “Firewall five main functions”, www.xinli.com.cn/showPage.phtml?cID=67&oID=322 (accessed 31 July 2006)

Zwicky, E.D., Cooper, S. and Chapman, D.B. (2000), *Building Internet Firewalls*, 2nd Edition, O’ Reilly & associates

Section 2

Information Systems Security & Web Technologies and Security

Information Security Awareness & Training

H.AI-Ghatam and P.S.Dowland

Network Research Group, University of Plymouth, United Kingdom

e-mail: info@network-research-group.org

Abstract

This paper identifies users' weaknesses regarding information security awareness, based on a survey which was conducted to assess security awareness. The survey showed that it is often the actions of users which makes them vulnerable, and that many are aware of the problems but continue to compromise themselves. The survey result shows that training and education do benefit users, and there is a clear difference between trained users and the rest – demonstrating that education, and training can work. The survey results have been used to help develop a training tool, to provide resources and interactive learning tools for users. These tools have been assessed and demonstrated to help promote information security awareness.

Keywords

Security awareness, training, tutorial.

1. Introduction

Computer Security is something that all organisations want to achieve, and a problem which has been rising more and more, without a real solution being developed. A problem that all have, since the moment many got involved with the computer technology, and being part of the world global network, which is the internet. Organisations, and users started to complain about applications they use, and IT professionals for not providing much protection. Most of them do not know that the actual protection is in their hands. By having a level of security awareness themselves, security can be more realistically achieved. It is important to have a security awareness, and by using the right way to promote security awareness, the real security can be achieved.

A survey which been conducted to support this research, and which provided evidence that users which had previously training related to information security in the past, do have a better security awareness. After proving evidence of the problem existence, it became important to find a way to help to solve it. Training seems to be the best way to do that, that's why a security assessment and training tool been designed, and implemented. The tool provides users with the information and resources needed. In addition, the online assessment tool provides, allows users to assess themselves and provide them with a feedback, and supportive information. This paper has all the details, and information regarding the security awareness tool, the survey and other information related to the research.

2. Contexts

Having security awareness is very important to have for computer users, and employees. It is as important as having the latest firewalls, anti-virus, and security hardware. Aware computer users are the gate which attackers getting throw at the moment. Attackers trying to get the most from user's awareness, to try trick them, and get as much sensitive and confidential information as they can. To avoid that, computer users should be more aware about security aspects, and they should know that they are in the danger zone if they do not. Security is like a chain, and computer users are the largest link on this chain. Being the biggest link do not always means that it's the strongest one, actually computer users appear to be the weakest link on that chain. The reason of that is that users do not think that it have anything to do with their action and use of systems. 435 senior college and university administrators participated in a research for EDUCAUSA (Kvavik, 2004), which shows that user awareness and other cultural factors are the second major problem in US institutions, and it presents 42%, which shows how much awareness needed to prevent any security failures in our organisation.

3. Aim & Objectives

The project aims to prove the need of security awareness, and to help to produce a tool which provide normal computer users at home and at work with the basic knowledge they need to know in order to increase the security awareness level, so they can be more robust against computer security threats which targets computer users. The idea is to have a tool that have the ability to assess users, and provide employers to have an idea about the knowledge that their employee have. The tool will makes it easier for users to find their weaknesses, and will advice, and provide users with the appropriate materials, and tutorials that they have to look at depending on their assessment. The tool will help computer users to enhance their security knowledge, and gives them an idea about how they can avoid being trapped by their own mistakes, especially in a time when everybody connected to internet, the place where users get very vulnerable to security thetas without their knowledge.

4. The User Awareness Problem

A survey about users attitude and awareness, which been carried out in UK (Dowland *et al* 1999), about public attitudes and awareness. Aimed to assess the attitude and awareness of general public about the computer crime, and abuse. The research shows that most people have the wrong idea about where the danger is coming from, and the people who is responsible. The research showed that the media have a very big effect on computer users, where media always tend to put the blame on computer hackers on any breach of computer security, and the way that media defining hackers as a very skilled people in IT which nothing can stop them from backing into your system. The research shows that 30% of respondents think hackers are lonely, young, male and lacking social skills, which do not really refract the real image of hackers. This is just an example of how people are really unaware about what's really happening and what is the danger. Lack of security education, and the media way of presenting the problem is driving computer users to the wrong way.

But at the other hand, Media doing a great job by getting people to understand that there is a big danger of been attacked, and showing them the type of crimes that might happen using computers. The research shows that 80% of respondents felt that computer crime and abuse was a problem, and most people highly consider computer crime as a serious concern.

5. Promoting Security Awareness

Most organisations do understand that are security problem, but they are not fully understanding what to do about it. Most the time they end with doing the wrong things to solve the problem, or they try to do the right thing, and some do the first right step, but they stop there (Furnell *et al*, 2002). Like obtains guidelines like ISO 17799, and British standard 7799, but they do not really do what these guidelines says. Organisations have to understand that these guidelines are all about insuring security and responsibilities with the organisation to be highlighted and reinforced, by doing number of steps, like having a security officer within the organisation to handle security issues, and to determines what need to be done to reinforce the organisation security at the right time and by using the right tools. Promoting security awareness among organisation members, during day-to-day activities, and not forgetting the most important thing, which is ensuring that organisation members to take training courses.

Training is an essential part to ensure the security of your organisation, by teaching all members the right way to use they systems securely. Training is not meant to be just for key staff members, or just IT staff, but it should include all type of members. Different type should be placed for all types of members.

6. User Security Awareness Survey

It's important to know the level of security awareness users have, before taking any actions. In order to determine that, a survey has been conducted to help find some of the common security mistakes that users do, and how user normally reacts to some common can faces in their daily use of internet and computer, along with some other electronic devices like mobile phones. Users normally unaware of the danger that surrounding them while they are using internet, or even a computer connected to a privet network. Users tend to get them self in trouble by the action they do while using computer, and internet. Most users are unaware that the actions they are taking, makes them so vulnerable to security threats. One of the important question that this survey tries to answer is whether the level of awareness that users have does reflects their actions while using internet and computer, and if an computer security awareness campaign needs to be taken in mind, and if that can help increase the security level among all computer users.

The survey distributed online, and it was linked to the University of Plymouth's Network Research Group site, and to the British Computer Society's south west branch site which allowed visitors from both sites to participate and take part of the survey, which titled User Security Awareness. To get the best results from the survey, and not just from computer professionals, people from different

backgrounds, ages, and education levels been invited to participate as well. and The survey has been conducted for a period of 2 months, between March and May 2006, with 135 participants taken part in the survey.

When respondents quizzed about their actions regarding internet pop-up, by presenting them with an internet pop-up messages. The right action to take is to ignore the pop-up, and that to avoid the possibility of clicking on a link which can lead to virus infection or security attack on your computer. 77.78% of respondents said that they will ignore the message if it come up, while the remaining 22.22% of respondents said that they will response to the pop-up message.

It may sound positive that most of respondents had chosen to ignore the pop-up message, but 22.22% still a big number, especially taken in mind that 80% of the respondents who chosen to respond to the pop-up message are using the internet more than once every day. You can imagine the security threat that can face those who responds to those kind of pop-up, especially with the increase of malware threats, and pop-up on the internet. The survey shows as well that most of the respondents who responded to the pop-up message are at the age of 17 to 29 with 70%, with 30% for respondents at the age of 30 to 59. That give you an idea about the importance about educating young internet users about the safe use of computer and internet in particular, knowing that 88.8% of respondents whose aged 17 to 29 do use internet more than once every day, which shows how important is to consider training and educating users about the safe use of computer and internet, not only for employee, but also for young users as they make the majority of users at the present time.

Passwords is very important to keep personal information safe from others, and for passwords to be effective, they need to be used correctly. First of all, participants asked whether they use the same password for more than one application and e-mail. 55.5% said that they use the same password for more than one application and e-mail, where the 44.4% claim that they do not. Then participants asked about what do they do to remember their passwords. 73.3% of respondents stated that they do not have any problem memorising passwords. 6.6% said they do write their passwords down and keep them in a safe place, while 6.6% of respondents said they will do the same and write it down, but they will keep it somewhere close to the computer, or stick it on the computer monitor. 4.4% of respondents do save their passwords in their computers, by writing it in a text document, which is saved in the computer. Other 4.4% of respondents said they will use a password management software to keep their password in a safe place in case they need it to have a look to them, in case they forget them. 2.2% of respondents said that they save their passwords in their mobile phones. The remaining 2.2% of respondents said that they have another way to remember their passwords, but without stating what is that way they using.

The respondents, who said that they do not have any problem in memorising their passwords with 73.3% of respondents. It appears that 81.8% of them are at the age of 17 to 29 years old, which can make since because normally old people have less ability to memories things than young people. In fact, the reason why 73.3% of respondents do not have problem memorising their passwords, isn't because their age and ability to remember, but because that 51.5% of those who do not have problem

memorising (73.3%), are actually use the same password for more than one e-mail and application, which explain a lot why those do not have problem memorising, or they think they do not have problem. Memorising password always is the best option, but using the same password for more than one application will take the advantage from memorising the password, and users will have the disadvantage of making themselves very vulnerable, once your password get discovered by someone. In fact, the amount of lose can be more greater than, if user using different passwords for different applications, because who have the password will be able to access any system that user use, specially that the user use the same password for the different applications.

When participants asked about if they do open e-mails from unknown senders, 44.4% said they do not. 20% of participants said it depends on the e-mail subject whether they open it or not. 17.7% said they do sometimes open e-mails from unknown senders, with 8.8% not sure. Some minority of respondents with 6.6% say they do open e-mails from unknown senders. One of the respondents gave an interesting answer, by declaring that he use the university computers to open any suspicious e-mails messages he receives.

What's more, is that 66.6% of respondents who do open e-mails from unknown senders, do not check file extension, when they download files from the internet. In contrast, 90% of respondents who do not open e-mails from unknown senders, do check file extension before they download any file from the internet. Also, 66.6% of respondents who do open e-mails from unknown senders, never do back-up their computer files. To make matter worse for those who open e-mails from un-known senders, 33.3% of them said they would use the computer and internet normally, and will not care about if there is a virus attack going to struck soon, and they even will not take any action to prevent their computers get attacked. In fact, most of those respondents (84.2%) do use the internet more than once every day.

To see how training is effective, if it's been given to users, participants been asked whether they had IT security training or not. 66.6% of respondents said that they didn't have any training, while the other 33.3% said they did have IT security training. As well as having training, 53.3% of respondents said they get reminded time to time about computer and internet security. The positive thing, that 73.3% of the respondents who had training, do have a strong passwords, which is a combination of characters, words, and numbers. In addition, 60% of them, do change their passwords every one to three months. On the other hand, 53.3% of respondents who had training, do open e-mails from unknown senders.

Having a computer and internet security training is very important, but not all kind of training can be beneficial to computer users. But it seems that many respondents who had training in the past, and benefited from it, would recommend having training to other computer users. In fact, 41.9% of respondents who had training in the past, thinks that computer users should have training, related to computer and internet security.

7. Why Do We Need A Training Tool

Having a training tool at your organisation can get you a great benefit, and it can be a starting point for all members to know about IT security. Each organisation has to ensure that their members have got at least the basic knowledge of security, and know what to do in some circumstances. And training tools can be used as a security assessment tool at the same time, where organisations can assess their member of staff, and train them at the same time. The training tool make it easy for organisation to ensure all members have the minimum amount of security information required, which will make their future work more easier, as it going to be about keeping members updated at all time, and reminding them about what already know.

8. Security Awareness Assessment & Training Tool

A training tool has been designed in a way which makes it accessible by everybody from everywhere, that's why it been designed as an online application, with an easy navigation, and an easy way to use. The main part of the tool been designed using flash which can provide a nice, and easy interactive interface, with multimedia. The tool offers assessment, and tutorials for users to take.

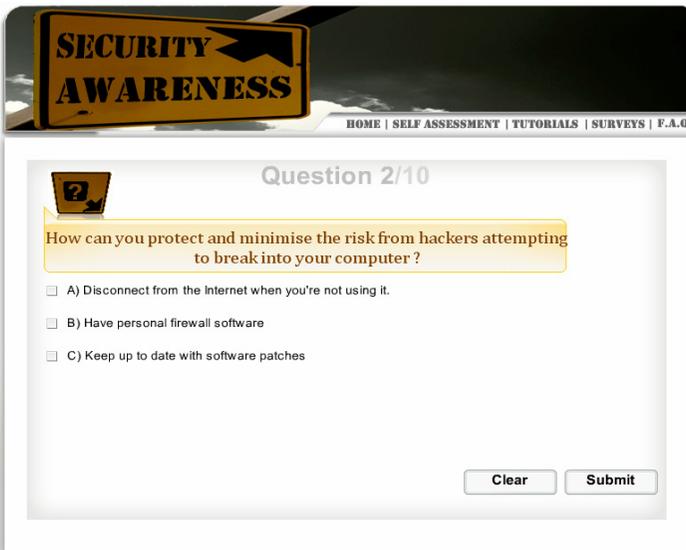


Figure 1: The assessment tool

When users access the assessment tool, they will be presented with number of questions, and they going to be assessed upon their answer to the questions (See figure 1). The users will be able to hear the question as well as reading it from the screen, which will provide a better level of interaction between user, and the application. If the user gives the right answer then will be presented with the next question or case. If the answer was wrong then he will be shown the right answer and why his question is wrong, before taking the user to the next question. That will help

providing users with just the information they do not know, while answering the right question will prevent giving them information they already know.

The online assessment tool will provide users with their assessment's result, by the end of the assessment. After users finish their assessment, and by the end of it, users will see the score they have got, and a short description what that score mean. In addition, users going to be provided with a list of tutorials which they can take to improve their security knowledge. In addition, the online tool will provide users with number of tutorials, which they can access and use any time. These tutorials can be found on the tutorials section, which can be accessed easily from anywhere in the tool. When users click on the tutorial they want to take, the tool will load the tutorial they are looking to take. Tutorials will display a question and an answer, to allow users to understand what the given information solves, and how they can they benefit from it.

9. Conclusion

The initial idea of this project is to develop a user awareness assessment and training tool, which been implemented and designed upon the survey result. The survey which been conducted for a period of two months, did provide some valuable information. The survey did prove that most computer users do have unsafe computer practices, which make users, and the computers they work on very vulnerable. In fact, it's been proven by the survey, that users who had previously training, have a better, and safer computer practices. They do have stronger passwords, do back-up regularly, and have a better idea how to protect their computers. In addition, educated users are more difficult to be tricked my malware writers.

10. References

Dowland P.S, Furnell S.M, Illingworth H.M and Reynolds P.L (1999). "Computer crime and abuse: A survey of public Attitude and awareness", *Computers & Security*, vol. 18, no. 8, Pages 715-726, 1999

Furnell S.M, Gennatou M, Dowland P.S (2002)"prototype tool for information security awareness and training", *International Journal of Logistics Information Management*, vol. 15, no. 5, Pages 352-357, 2002

Kvavik B.R (2004). "Information Technology Security: Governance, Strategy, and Practice in Higher Education", EDUCAUSE Centre for applied research at University of Minnesota, Available: http://www.educause.edu/ir/library/pdf/ecar_so/ers/ERS0305/ECM0305.pdf, [Accessed January 2006]

Security Technologies: Why are they not used correctly?

M.Al-Tawqi and S.M.Furnell

Network Research Group, University of Plymouth, Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

It is a fact that security technologies suffer usability difficulties, with prior studies revealing that end-users are not able to correctly use security technologies as a result of difficulties arising from complexity of design and mis-presentation of security features. In this aspect, this research presents the findings of an interview study conducted amongst over 75 end-users to explore the reasons behind such behavior, where known applications (Windows XP, Word, Internet Explore and e-mails & passwords) were used as examples. The findings revealed that participants were keen to show their wide awareness of security usefulness, with 87% of them using antivirus software, but with only 34% updating them either daily or weekly. It was also revealed that more than 57% of the participants disagreed that it is very easy to use software security features, while more than 60% disagreed that understanding the security features in software is easy. The roots of the problem were attributed to the complexity and unfriendly nature of the software, which require urgent moves from designers to simplify their security products.

Keywords

Security, Usability, Human computer Interaction, Windows XP, Word, Internet Explorer

1. Introduction

Information Security experts are now aware that in many cases breaches occur as a result of software not being used appropriately. Although suitable technology solutions are available to prevent computer users from such incidents, end-users are frequently unable to use them in the correct manner. This can take the form of end-user not knowing that those features exist, not caring enough to install them, or even willingly neglecting security, as in the case when people traded their passwords for candy bars (Saita, 2004). The other reason for such problems is related to the complexity of the software provided, which often scares away end-users instead of encouraging them to use such technologies.

2. Factors and contributors to the usability issue

One of the most common problems encountered within software security features is that it is *difficult to find and utilize security options*, as most of the times they are hidden instead of being placed in the forefront of the user interface. As a result, a user would only be able to access security features through routes such as the Tools – Option menu; which in many cases does not happen because of lack of knowledge that they exist (Furnell, 2005). Unfortunately, even when features are located, another immediate problem may be *the ability to understand and use the features*.

Indeed, the understandability problem was recognized and hence considered by the CRA (2003) as one of the four grand challenges in Information Security. The reason behind this can be attributed to either the lack of awareness amongst end-users, or the difficulty in finding meaning in the provided interface. The latter again relates to the unfriendly nature and complexity of the software itself, often as a result of designers and developers being more concerned with the software itself rather than taking the needs of regular end-users in consideration. The two problems were then extended by Furnell et al. (2006) to cover many other problems, as listed below:

- *Forcing Uninformed Decisions.* Difficulties in using security features can encourage poor security decisions (Zurko, 2005). The other problem faced by end-users is the detection of intrusion attempts accurately, whether before hands or afterwards (McHugh, 2001).
- *Lack of Integration.* If different elements of security do not work together, users might end-up getting a deceived message asking them to do something or install a software such as an anti-virus protection one when the user has already installed one. Good et al. (2005) agreed that this problem is in fact is doing more harm than helping at all.
- *Lack of Visible Status and Informative Feedback.* In addition to having difficulties finding the security options, users will also not have feedback from the system to inform them about the new state of security configurations.

Others have found that aspects of the Human Computer Interaction (HCI) were ignored by developers of security related products, who then tried to prove that use of security technologies can be improved if HCI techniques are employed. In this regard, Zurko and Simon (1996) came with solutions to help in providing user-centered security; these are:

- Applying HCI design and testing techniques to secure systems.
- Providing security mechanisms and models for human collaboration software.
- Designing security features directly desired by users for their immediate and obvious assurances (for example, signatures).

3. Investigative methodology

Much previous research has been conducted in a survey manner, where there was no interaction between participants and the questionnaire initiator; and with all of the factors mentioned above in mind, the investigation for this study was conducted in a form of a structured interview during the period 10th June- 15th July 2006. This method was considered to provide the benefits of having a questionnaire and a face-to-face interview at the same time, so that accurate observations are easily obtained and recorded from the mouth of the interviewees. The questionnaire was titled *Software Security Usability Survey*, while 71 participants' replies out of 76 replies were taken in consideration as there was evidence that the remaining 5 replies were either answered randomly or returned uncompleted. The questions of the research touched on the following areas: background, awareness, utilization and importance

of certain security features, E-mails and their passwords, use of Windows XP and some of its applications. The analysis and interpretation of the questionnaire has been guided (in most cases) by Nielsen's usability heuristics to check on the usability of the operating system and applications discussed (Nielsen, 1994). In addition, previous studies of the same subject were used as means of guidance in terms of comparisons of findings.

3.1 Participants demographics and background

Two thirds of the participants were male, and the majority (64%) was between 18-34 years of age, which suggests that most were computer users from a generation who have grown up with such technology. Findings also indicated that majority of participants were degree holders (79%), with most of them holding a Bachelor degree as a minimum. Although 56% of participants viewed themselves as being intermediate users, it turned up not to be quite right after having heard their answers to some of simple questions during the interview. Those participants can be excused as they have no other choice but to select that option since they neither consider themselves as novice users nor advanced. It is worth noticing that such result is applicable to many other studies in regard to assessing computing experience. On the other hand, computers usage appeared to be widely spreading, with 80% using it continuously at work and home for various reasons. It is also worth mentioning that among those who use computers at work only, are IT-related employees who try to escape the work environment by trying to live a computer life free when they are at home.

3.2 Security Features Awareness, Utilization and Importance

When participants were asked about the knowledge of security features within MS Windows and some of its applications, as well as the use of antivirus software for their home computers, they were keen to show that they have a wide awareness of those features. It was noted that 87% used antivirus software, which is a promising result in a way. Although Word, Outlook Express and Internet Explorer have received the least percentage of awareness, this was justified since these results were utilized from the number of participants who only said they used them, which meant that more than 80% of those who claimed to be using the 3 applications were aware of there features. When compared to findings from other research, these results show an improvement in awareness amongst participants; with previous research having reported percentages of 68% for Internet Explorer, 56% for Word, and 32% only for Outlook Express (Furnell et al. 2005).

The results were less promising when participants were asked whether features shown in Figure 1 were actually utilized, as they revealed that majority have either never used the feature or used it seldom, even though they were concerned about security breaches. Reasons stated by participants for not using the security features have agreed to a great extent with some of other researchers findings, such as: *visibility*, *unfriendliness*, *difficulty* and *performance penalty* (Nielsen, 1994; Johnston et al. 2003). Participants also listed other barriers such as: their feeling that whatever type of protection implemented is *breakable any way*; *time consuming* to set up to desired functions; *insufficient knowledge & information*; and *carelessness &*

laziness. All of these factors make it difficult to convince end-users to take the initiative to secure their systems to the required level.

Participants' answers were in line with the findings of security feature awareness and utilization sections, which proved that participants have the sense for importance of security and that they are aware about the obstacles preventing them from fully utilizing security technologies. This shows that although people are aware of the implications of security breaches, they do not do much to prevent them from occurring, either because they do not bother, think that the problem will solve itself, or it will not occur to them for them being 'ordinary' users with 'ordinary' data. This was not necessarily correct when findings were compared to those of the IT-related employee findings, who proved to be better in applying necessary measures to protect their systems.

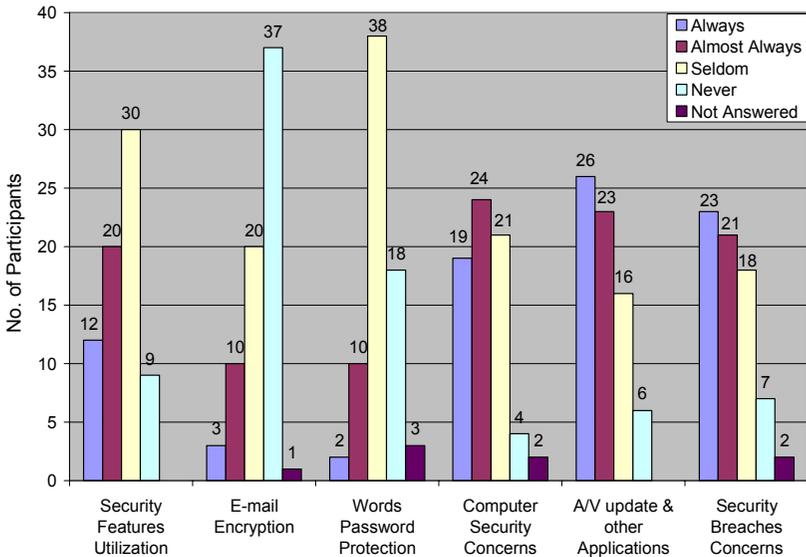


Figure 1: Utilisation of security

3.3 Windows XP, Anti-virus & Firewall

Despite the publicity of the inclusion of security features when Windows XP Service Pack 2 (SP2) was introduced, findings revealed that only 58% of participants heard of it, with less than 41% of total participants being able to list some (not all) of the features. This is another indication that end-users are not benefiting from an available facility that is been well advertised. Again, reasons can be attributed to carelessness and laziness from those who knew about it, or to the lack of knowledge for those who did not know about its existence.

When considering the aspects of protection that the Windows Security Center can control, the findings revealed that the majority of participants installed antivirus software, and were able to explain its role correctly (68%). However, this was not true for firewalls, as 59% of them did not know how firewalls are set to protect computers, while 63% did not know how exceptions are selected for allowing some

programs not to be blocked by firewalls. Lack of knowledge about firewalls was further confirmed, as 58% did not know about the difference between an antivirus and a firewall. These findings confirm earlier research, which found that almost 45% of participants did not know about the difference between anti-virus and firewall protection (AOL & NCSA 2005). However, when such answers were compared to those of the IT-related employees, it revealed that the second group is well-informed and more educated about such differences, with more than 76% being able to give the right answer.

3.4 MS Word

This section checked on the knowledge of protection for Word files, and on the knowledge of what Macros are. Results revealed that majority of participants are aware that password protection for Word exists (85%), with 67% utilizing it while 83% gave the correct answer for the difference between modify and read only files (this case was true for IT-related employees participants, with 95%, 76% and 86% results respectively). This overwhelming knowledge and utilization of such facilities is understood and expected for Word being popular with almost every single user of computers. As for the knowledge of Macros, results revealed that 62% of participants are aware that such facility exists, with 85% of these aware participants utilizing it, while only 48% saying that they know how to set security level for Macros. This case was almost true for IT-related employees, with 71%, 83% and 67% (slight difference) results respectively, which can be related to the familiarity of Word with most computer users.

3.5 E-mail

Answers of participants revealed that only 39% knew about SSL certificates, with only 64% of these participants knowing about the way of obtaining them. This demonstrates that if a feature's existence is not known, then there is no way it will be widely recognized and hence used. Findings also revealed that: 82% did not agree that *internet email is very secure* with more than 80% agreeing that *sent e-mails can be read, intercepted or modified by others* which indicated that participants are aware that they might be spied upon by others, as there is no security once they are online, and that (unless specifically protected) every single message sent by them can be intercepted by others. However, findings revealed that this knowledge of such problem did not prevent them from being victims of message interception, as 87% admitted that they have no means of knowing if messages received or sent by them were modified. Unfortunately, this is not the only area where end-users become victims to security breaches, but in fact it is them who make it easier for others to hack on them as a result of their reactions to security matters, when findings indicated that 67% of them will open an e-mail from a stranger with 57% saying that they will even open attachments before saving them first.

3.6 Passwords and other authentication methods

Participants were asked about their awareness of password, token, and biometric methods. The first significant finding was that the majority (69%) did not know about all three categories; which explains why only 25% have considered using the

other 2 methods instead of using a password only. Thus, it is understood why most end-users (85%) use passwords the most. However, although password authentication was still preferred over other methods, the percentage here was less than the current level of usage, with a noticeable portion of participants claiming that they would prefer other methods (61% *passwords*, 24% *combination* and 5% for *biometrics* and *tokens*). The preference to use passwords is understood, when it is known that most users have not encountered the other two methods, which means that they are forced to choose passwords as the only method they are familiar with. On the other hand, biometric authentication was preferred because fingerprints – for example - are unique and therefore cannot be imitated. However, this same justification was the reason why others refrained from preferring such method (as they would rather have their token or password stolen than their thumb), while others felt it would not be cost effective. Figure 2 gives a clear explanation of participants' awareness of password importance, where variation in wrong doing revealed the following:

- *Participants are having difficulties keeping up with more than one password*, which makes it difficult for them to remember them all or come up with new ones, and hence being forced to either use a changed password more than once or/and use it for more than one system.
- *Participants are less willing to share passwords with friends or write them down*, though they sometimes do that but to a lower extent. This highlights that participants are aware of risks of losing passwords, but they will only compromise when there is no other way of getting around it.

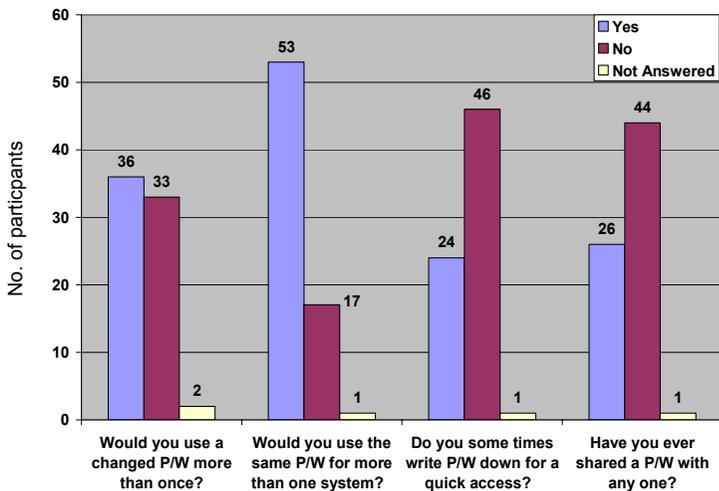


Figure 2: Participants' responses about awareness of password importance

Opinions provided by participants for guidelines or steps that should be taken or imposed for better selection of passwords were as follows:

- *Never use personal data.*
- *Use different combinations of characters, numbers and symbols.*

- *Change passwords every 45 days and never use old passwords again.*
- *Promote better security awareness and education.*

3.7 Internet Explorer

Users were asked about their knowledge of using trusted and restricted sites, and restriction of cookies. The results revealed that 61%, 62% and 55% of participants are aware of the their existence (81%, 81% and 76% of IT-related employees). However, this variation in percentages of the two compared groups did not prove that the second group is fully utilizing such available facilities, as results revealed that it barely reached 29% for the first group and 38% for the second group as the highest result reached for the utilization of any single feature.

Participants were also asked to consider the manner in which security features are presented and explained. When they were asked about Figures 3 and 4, their most common comments were that: *there should be an explanation of the actions taken with examples, description is a bit vague and may scare novice users away, such screens should be automatic, there should be better explanation for beginners, and the figure should properly explain the feature, not the words.*

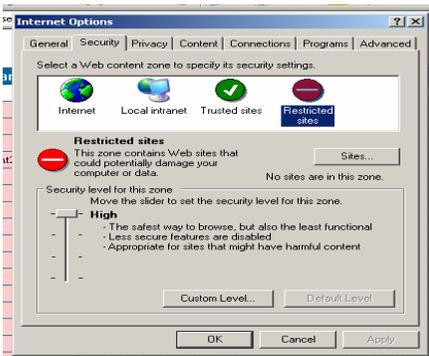


Figure 3: Explanation of what restricted sites are

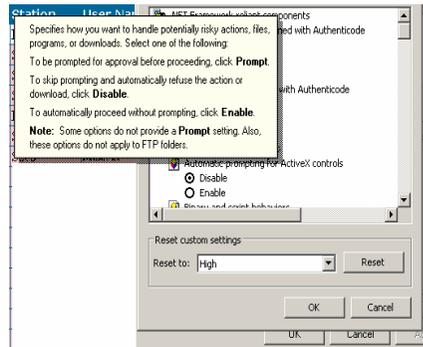


Figure 4: Explanation provided when clicking to enquire about settings

3.6 General

Evidence showed that end-users believe that security issues should be the concern of all of those dealing with or effected by them, as more than two thirds (68%) of participants said that security issues should be handled by both end-users and designers.

Other relevant findings were that:

- 80% agreed that security alerts provide a good indication of current status;
- 73% asked for a centralized security option for all MS-related software;
- 82% agreed to introducing different warning levels for moving from one site to another;

- 86% agreed that security tips and hints should be provided at the start of sessions.

Findings have also agreed and confirmed to those established by previous studies on the same field for e-mails being the main cause for infections with 64%, where other studies have revealed higher percentage (80%) for e-mails being the main source of virus infection, (Panda Software, 2006).

4 Conclusion

Unfortunately, some end-users do not take the initiative of updating themselves in regard to information security issues; mistakenly assuming that it is solely the duty of their organization to take care of training and educating them on how to fight security threats. In fact, it is end-users' responsibility to make sure that they keep up with the pace of information security development, as it is them who will eventually be directly effected by any breaches. On the other hand, designers must understand that they are developing such software so that they are useable, and that not every user has their level of knowledge, which means that they have to picture regular end-users and put themselves in their shoes so they can come up with tools that are widely accepted by majority of computer users. Therefore, the research highlights the need for action by designers and employers in order to assist end-users, where employers must foster knowledge and understanding of security features to their employees in order for those features to be appropriately utilized. On the other hand, designers must provide features which are visible and friendly with very clear explanation.

5 References

- AOL and NCSA, 2005. *AOL/NCSA Online Safety Study*, http://www.staysafeonline.info/pdf/safety_study_v04.pdf#search=%22AOL%2FNCSA%20Online%20Safety%20Study%20%22 (06 August 2006)
- CRA Web Site, 2005. *Challenges in Information Security & Assurance*, <http://www.cra.org/Activities/grand.challenges/security/home.html> (17 August 2006)
- Furnell, S.M. 2005. "Why users cannot use security", *Computer & Security*, vol.24, pp274-279.
- Furnell, S.M., Jusoh, A. and Katsabas, D. 2005. "The challenges of understanding and using security: A survey of end-users", *Computers & Security*, vol. 25, no.1. pp27-35.
- Furnell, S.M., Jusoh, A., Katsabas, D. and Dowland, P.S. 2006. "Considering the Usability of End-User Security Software", *Proceedings of 21st IFIP International Information Security Conference (IFIP SEC 2006)*, Karlstad, Sweden, 22-24 May 2006, pp307-316.
- Good, N., Dhamija, R., Grossklags, J., Thaw, D., Aronowitz, S., Mulligan, D. and Konstan, J. 2005. "Stopping Spyware at the Gate: a User Study of Privacy, Notice and Spyware", *Proceedings of the 2005 Symposium On Usable Privacy and Security*, Pittsburgh, Pennsylvania, pp 43 – 52.

Johnston, J., Eloff, J.H.P. and Labuschagne, L. 2003. "Security and human computer interfaces", *Computers & Security*, vol. 22, no. 8, pp 675-684.

McHugh, J. 2001. "*Intrusion and Intrusion Detection*", Integrated Justice Information Sharing IIJS, 2001, vol. 1, pp 14 – 35.

Nielsen, J. 1994. "Ten Usability Heuristics", http://www.useit.com/papers/heuristic/heuristic_list.html. (25 January 2006)

Panda Software, 2006, Virus Entry Points, http://www.pandasoftware.com/virus_info/about_virus/information1.htm (18 September 2006)

Saita, A. 2004. "Password protection no match for Easter egg lovers", http://searchsecurity.techtarget.com/originalContent/0,289142,sid14_gci960468,00.html (17 August 2006)

Zurko, M. E. 2005. "*User-Centered Security: Stepping Up to the Grand Challenge*", 21st Annual Computer Security Applications Conference (ACSAC'05) pp. 187-202

Zurko, M.E and Simon, R.T. 1996. "User-Centered Security", *Proceedings of the 1996 Workshop on New Security Paradigms*, Lake Arrowhead, California, United States, pp. 27 – 33.

Web-based Plankton Data Visualisation

T.Y.Aung and P.S.Dowland

Network Research Group, University of Plymouth, Plymouth, UK
e-mail: info@network-research-group.org

Abstract

Data visualisation over the Internet is a challenging task for web developers. North Sea plankton database is available to the scientists, researchers and students all over the world by using the WinCPR software. However, users need to download the WinCPR software to use the plankton database. The web-based plankton data visualisation research paper provides the various techniques to visualise the data from the North Sea plankton database without downloading WinCPR software. It also suggests using the different database system for online high data traffic. This paper proposes specific drawing methods for each chart type to visualise the data based on the Web-based CPR project which is developed using ASP.NET 2.0 and VB scripts. It also discusses the result of the different drawing methods and using different image formats. At last, it suggests the improvements for the current project and possible features for the future works.

Keywords

Data Visualisation, Image generation dynamically

1. Introduction

The web-based plankton data visualisation is based on the WinCPR software. The WinCPR software is a gridded database browser of North Sea Plankton database which is collected and maintained by the Sir Alister Hardy Foundation for Ocean Science (SAHFOS), Plymouth, UK. This database has kept recording monthly plankton abundance from the Continuous Plankton Recorder (CPR) survey. However, WinCPR software has the limitation of the database and it is the Windows-based software. Therefore, users need to download to use the software. The web-based CPR project is able to provide main functionalities of WinCPR software without downloading any software.

There are a number of processes to accomplish to produce an image dynamically on the Internet. Users have to make a query to get a result for their requirement. So, data collection from users is one part of data visualisation. Once users have filled the query, a connection to database and SQL command has to be done programmatically. After fetching the data from the database, the result data will be used to draw the appropriate graph or chart. The last step is to send the result graph or chart to the client browser. Entire task has to be accomplished dynamically. There are a lot of issues on each step. The foundation of this research paper is the web-based CPR project which is based on the WinCPR software. Three main parts can be classified to visualise the data on the fly in the project. They are Data Collection from the users, Database Connection, Image generation and Visualisation to the users.

2. Data Collection from the users

Each user has different level of understanding and familiarity to the database that they are trying to use. In the most case scenarios, users are assumed to make mistakes while they are completing the query. Thus, the developers have to make sure that the data filled by the users are accurate. In this Plankton Data Visualisation, users have no need to type to fill the query. Users have to choose the required data from the wizard on each step. Appropriate data will be displayed by using server controls on each step of wizard. It makes sure to be accurate the data by using the wizard server control. This Wizard server control <asp:Wizard> is supported by the ASP.NET 2.0. It builds a series of steps to display to the users. It has several properties to customize the appearance and to meet the requirements of developers. In visual studio 2005, the wizard control can be configured visually. Wizard steps can be added or removed visually in the developing time and programmatically at the run time. In the project, the steps of wizard display the data from the database. Database connection has been made depending on the choice of user on each step of wizard. After user has completed the wizard, the SQL command will be created.

3. Database Connection

In this project, although Microsoft SQL Server was intended to use in the first place, Microsoft Access database is used as the backend database. It is because the size of the database used in the project is small and it was not intended to use online as a professional web application. For online used as a professional web application, other relational database system such as Microsoft SQL server, Oracle and MySQL will be appropriate to use as a backend database because Microsoft Access is not designed to use high data traffic. In Microsoft article Q300216, it describes that when a file-sharing database is used in a multi-user environment, multiple client processes are using file read, write and locking operation on the same shared file across a network. If, for any reason, a process cannot be completed, the file can be left in an incomplete or a corrupted state. The other reason to use Microsoft Access in the project is that MS Access is portable and easy to use.

Entire project has used straight forward database connections. All of the SQL commands are created on the fly programmatically by using the user inputs. Moreover all of the command types are “Select” command type. There is no “Update” or “Delete” command to the original Continuous Plankton Recorder (CPR) database through out entire project. Therefore, underlying database has never been modified. If this project goes online and uses Microsoft SQL server or other database system, there is no need to implement to modify the database online. It will reduce several of security threats to the database system.

3.1 Microsoft Access Database Connection in .NET framework

The .NET framework provides the “System.Data.OleDb” name space to access an OLE DB data source. This name space has a number of classes to use to connect to the MS Access database. In the project, only three classes are used to connect to the database. They are “OleDbConnection” class, “OleDbCommand” class and

“OleDbReader” class. When user has finished the wizard and click to view the graph or map, all the data user has chosen will be sent to the image generate page via the URL. All the data will be retrieved from the URL on the image generate page by using “Request.Params ()” method and put them into local variables. Then a database connection object will be created by using “OleDbConnection” constructor which takes a connection string as a parameter. The connection to the database will be opened by using “Open ()” method of “OleDbConnection” object. The variables which are generated from the URL will be used to create a SQL command. Then, an “OleDbCommand” object will be created using the SQL command and the “OleDbConnection” object. After that, “OleDbReader” object will be created by using the “ExecuteReader ()” method of the “OleDbCommand” object. That method will return the rows of the result of the SQL command. The values of rows will be put into the appropriate variables and it will be used to draw the image.

3.2 SQL Server Database Connection in .NET framework

To use SQL server in the future works, “System.Data.SqlClient” and “System.Data” name space are required to import. The “System.Data.SqlClient” name space provides access to SQL server database and the “System.Data” name space consists of most of the classes that constitute ADO.NET architecture (source: MSDN). “SqlDataAdapter” class will connect to the database and fill the dataset with data from the specific SQL command. “SelectCommand ()” method of “SqlDataAdapter” object will send a query to the database along with a connection object. In this case, the connection object will be created using “SqlConnection” object. “Dataset” object will then be filled by using “Fill ()” method of “SqlDataAdapter” object. “DataSet” object will contain “DataTable” which consists of “DataColumnCollection”. The required data will be cached into the dataset and the necessary processes to draw images can be accomplished by using it.

4. Image Generation and Visualisation to users

The main subject of this research paper is generating the image on the fly and sending it to the users. It can be divided into two parts. The first one is drawing the image and second one is sending the image to the users. Both parts have different techniques and approaches.

The first part is to generate the image from the result of the query that user made. In ASP.NET 2.0, Graphics Device Interface (GDI+) supports the powerful graphics interfaces which allow developers to generate graphics and image handling code. The GDI+ interacts with graphical display devices such as a screen or printer instead of developers. Therefore, the developers have no concerned to the particular devices. “System.Drawing” namespace provides basic drawing functionality and “System.Drawing.Imaging” namespace and “System.Drawing.Drawing2D” namespace support advanced drawing and imaging functionality. The “System.Drawing” namespace is used in this project to draw the image using its classes such as Bitmap, Color, Pen, Font, Graphics and Image classes and “System.Drawing.Imaging” namespace is used to set the image file types.

A Bitmap object is used as a canvas to draw the images and graphs. It can be created with specific width and height and also from a bmp image. When the Bitmap object has been created, a Graphics object is created by using the Bitmap object. After the Graphics object has been created, all the drawing can be started on the Bitmap by using Graphics object. The Graphics object supports all the functions to draw rectangles, circles, lines, strings and images. After drawing required elements on the Bitmap object, it has the function to save all the drawings. The save () method of Bitmap object takes 2 parameters. The first one is the location to save and the second one is the image type. The location can be the file system as well as memory stream buffer. The image can be saved in the file system by providing the specific file path and name and the image can be saved in output stream buffer to send it to the client browser. The image format can be set using ImageFormat class. This class provides several image formats including GIF, JPEG, BMP and PNG format.

To send the result image to the client browser, the content of the page must be set to image. The “Response.Content” has to be set to “image/gif” or “image/Jpeg” to let the client browser know that the sending page is an image file. “Response.BufferOutput” is set to “True” to process all the functions in the page before the image is sent to the client browser. Then, “Response.Flush” method will send the image stored in output stream buffer. The general idea and codes of creating drawing place and sending it to the client browser are described in listing (1).

```
Dim objBitmap As New Bitmap(800,1000)
Dim objGraphic As Graphics = Graphics.FromImage(objBitmap)

"Drawing Process Here"

Response.Clear()
Response.BufferOutput = True
Response.ContentType = "image/jpeg"
objBitmap.Save(Response.OutputStream, ImageFormat.Jpeg)

objGraphic.Dispose()
objBitmap.Dispose()
Response.Flush()
```

Listing (1)

Even though original WinCPR software provides 8 different chart types, there are three different chart types to visualise the data in the project. According to the survey which is carried out for this project, the selected three chart types are the most useful chart types for the users. Those chart types are Seasonality Graph, Annual Plankton Abundance and Monthly Plankton Abundance. The survey was carried out in summer 2006. 12 Marine Biologists, 4 postgraduate students, 2 undergraduate students, 1 teacher and 5 people from other area are participated in the survey. The survey was organised with 15 questions. There is a question in the survey to rate the usage of the chart types in WinCPR. Half of the participants rates very useful for Monthly Plankton Abundance chart type and Annual Plankton Abundance. Seasonality Graph has been voted as very useful by (37.5%) of participants. According to those results, those 3 chart types are included in the project. Moreover, those three chart types have different drawing approaches.

Seasonality Graph is a graph presenting the selected species data of specific location in the North Sea of a selected Year. The user has to select the species name, a year range between 1948 and 1997 and a pixel location. The pixel locations on the North Sea are shown in figure 1. Based on the user's input, the required information is retrieved from the database to draw the graph. 12 decimal values are retrieved from the database to draw the seasonality graph. Figure 2 displays a sample of seasonality graph produced by web-based CPR. First, create a Bitmap object with specific width and height and create a Graphics object using that Bitmap object. Then the bars will be drawn with same width and different height according to the result values. And the legend will be drawn and save the image to the Output stream buffer. That buffer then will be sent to the client browser.

Monthly Plankton Abundance is a chart presenting the density of specific species on the North Sea of a selected year. The user has to choose species name and a year. 2 approaches are provided to draw this chart type. First approach is to draw the image of each month separately and put it together in a table and show it to the client browser. The Bitmap object will be created with North Sea map bitmap file. Then the pixel location boxes on the map will be filled with specific colours according to the value of that location. The image will be saved into the memory output stream and send it to the browser. This process will repeat for each image for 12 months.

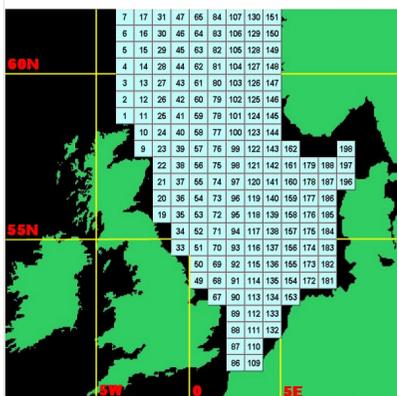


Figure 12: Pixel Locations on the North Sea Map

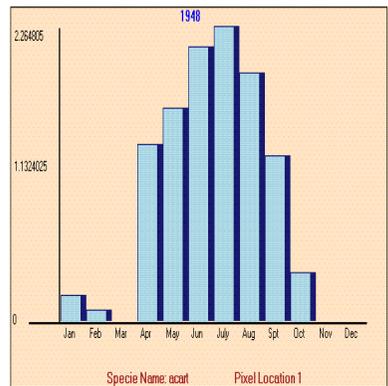


Figure 13: A sample Seasonality Graph

The second approach is to draw each month on a single image and send it to the client browser. This chart type uses the North Sea map as the background. Figures 3 and 4 show the sample of Monthly Plankton Abundance chart type produced by both methods. In this approach, a Bitmap object array which holds 12 elements of the array. The image for each month of the year is drawn and put into each elements of the array. Then, a big Bitmap object and a Graphics object are created. The DrawImage () method of the Graphics object draws all the images in the array into the big Bitmap object. The big Bitmap object is saved in the output stream buffer as an image and sends it to the client browser. The listing (2) shows the brief demonstration of all the process.

```

Dim objBitmap(12) As Bitmap
Put all the images for each month in the array
Dim bigBitmap As New Bitmap(1200, 1600)
Dim bigGraphic As Graphics = Graphics.FromImage(bigBitmap)

For a = 0 To 11
    bigGraphic.DrawImage(objBitmap(a), xpos1, ypos1, 350, 350)
    xpos1 += 400
    If a = 2 Or a = 5 Or a = 8 Then
        ypos1 += 400
        xpos1 = 0
    End If
Next

Response.Clear()
Response.BufferOutput = True
Response.ContentType = "image/gif"
bigBitmap.Save(Response.OutputStream, ImageFormat.Gif)
bigGraphic.Dispose()
bigBitmap.Dispose()
Response.Flush()

```

Listing (2)

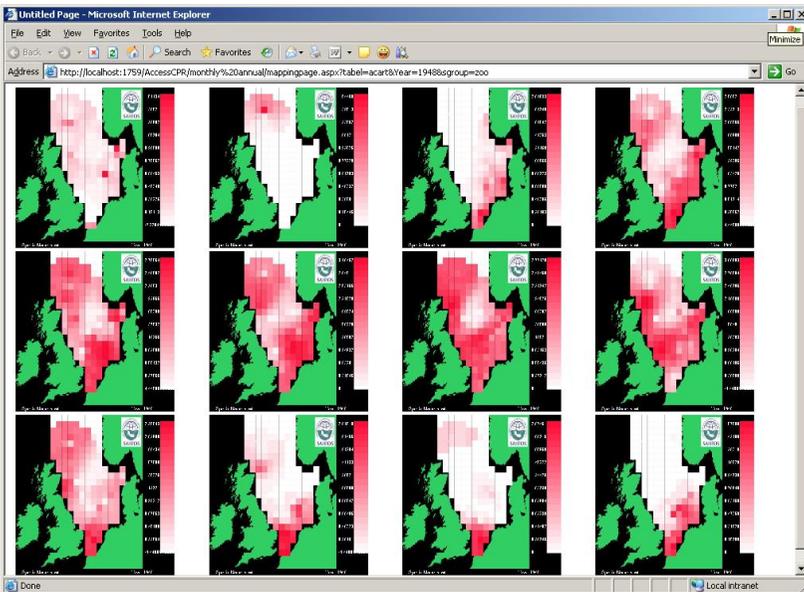


Figure 14: Monthly Plankton Abundance produced by first method

Annual Plankton Abundance is a chart presenting the density of selected species on the North Sea of a selected year. The user has to choose species name and a year to visualise the data. This chart type also uses the North Sea map as the background and fills the colour boxes of each location on the map. Figure 5 shows a sample of Annual Plankton Abundance chart type. In this chart type, a Bitmap object is created with North Sea map as the background and a Graphics object is created by using it. The pixel boxes are filled with colours according to the density of species that user selected. Then the image will be saved into the output buffer and send it to the client.

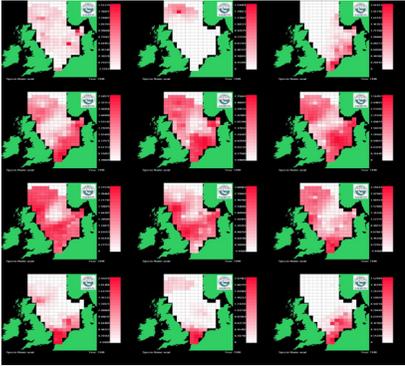


Figure 15: Monthly Plankton Abundance produced by second method

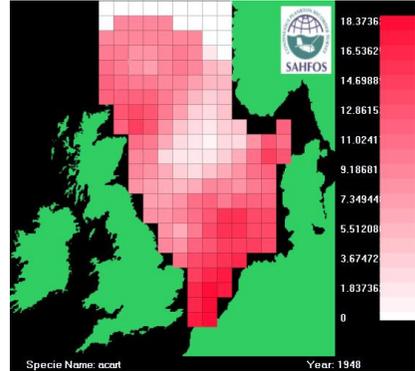


Figure 16: A sample of Annual Plankton Abundance chart type

5. Discussion

5.1 Image Quality and Image Size

The size of the image is critical in the data visualisation techniques. The image transfer time over the Internet depends on the size of the image. The smaller image size increases the speed of the transfer time. The image size produced by the project is between 10 kb and 300kb. The Seasonality graph has the size between 10kb and 30kb, Annual Plankton Abundance chart type has the size between 35kb and 40kb. Monthly Plankton Abundance chart type has the image size between 250kb and 300kb with second method. First method will send 12 images to the browser at the same time and each image has the size between 35kb and 40kb. Therefore, the first method increases the amount the data sending to the client browser. Two image formats are available to produce in the project. JPEG and GIF image formats are widely used over the Internet. JPEG format has larger size than GIF format but it has better quality.

5.2 Image Producing Techniques

The image can be produced in two ways. The first one is to produce the image on the server side and the second one is to produce on the client side. The server-side scripts are used to produce the image on the server and this technique is used in the project. Producing the image on the server increases the speed of image generation process rather than producing the image on the client side. The result image only is sent to the client browser. In this technique, all the imaging and drawing functions are done on the server. Therefore, all the requirements to draw the image are supported by the server. However, if the multiple users are generating the image on the server at the same, the performance of server will be decreased.

The second approach is to produce the image on the client side using client-side scripts. In this approach, server will send the client-side scripts to the client browser along with necessary data to draw the image. This approach will reduce the numbers of processes to generate the image on the server. Therefore, it will increase the

performance of server. However, the client machine has to permit to run the scripts on it. Most of the client-side scripts such as Java Applet need permission from the browser to run. The developers need to be aware of the target audience environment to use the specific client-script. Office and educational environment are not permitted to run the scripts on their system for security reasons. One of the security threats for client-side scripts running on the machine is that the client-side scripts can access to the resources of the machine. Moreover, the target audience for Web-based CPR is office environment and educational environment. Under those circumstances, the server-side scripts are more appropriate to use in the project than the client-side scripts.

5.3 Image Saving Methods

According to the survey result, (54.1 %) of the participants wanted to save the images on the server. However, there is no save function in the project. There are three possible approaches have been suggested to save the images on the server. Each approach has advantages and disadvantages.

First approach is to save the image as an image file on the server. If user chooses to save the image, the image name and location will be specified programmatically. When user retrieves the image next time, the image will be fetched from the server's file system and display it to the user. In this approach, server will need to provide the significant amount of space to store the images. However, when user retrieves the image, there will be no process to produce the image again. The server has to send the image directly to the client browser.

The second approach is to save the Bitmap object into a file. The Bitmap object will have all the drawings in it before it is sent to the file. When the user retrieves the image next time, user will have the chance to choose the image file type. This approach will reduce the size of file to save on the server. Nevertheless, it will have a small process to change it to the image file and send it to the browser.

The third approach is to save the result of user's query into the file along with a flag which can describe the chart type of the image. When the user retrieves the image, the flag will be checked to choose the chart type and the other data will be used to generate the image. This approach will have the smallest amount of data to save in the file system but it will also have the whole process to draw the image.

6. Conclusion

This study has accomplished to visualise the plankton data over the Internet without downloading any software. Moreover, this study has used only server-side scripts to produce the images on the fly. The advantages and disadvantages of using server-side and client-side scripts are highlighted. However, client-side scripts can achieve some of the features which server-side scripts are not able to achieve. For instance, the client-side scripts can save the images on the client machine by the web application itself. The density of plankton data on North Sea is shown by the colours in the project. However, producing the images with the client-side scripts can also let

user to change the colours and image file types on real time. This research paper also advises three approaches to save the images on the server. Nevertheless, this study is the basic step to visualise the plankton data over the Internet and it gives a range of ideas to produce the images dynamically.

7. Reference:

Cogan, A., (2004), “Moving Your Access 2002 Database to SQL Server” Microsoft Regional Director, Australia, December, white paper

Esposito, D., (2004a), “Cutting Edge Image generation Service for asp.net 1.1” MSDN magazine, April.

Esposito, D., (2004b), “Image generation service for ASP.NET 1.1” <http://msdn.microsoft.com/msdnmag/issues/04/04/CuttingEdge/> (accessed 15 August 2006)

Latarre, U., (1998), “Graphic file formats”, <http://www.why-not.com/articles/formats.htm#INDEX> (accessed 20 August 2006)

Microsoft Developer Network, “System.Data name space”, <http://msdn.microsoft.com/library/default.asp?url=/library/en-us/cpref/html/frlrfSystemDataOleDbConnectionClassTopic.asp> (accessed 30 august 2006)

Microsoft Developer Network, “System.Data.SqlClient name space”, <http://msdn.microsoft.com/library/default.asp?url=/library/en-us/cpref/html/frlrfSystemDataOleDb.asp> (accessed 30 august 2006)

Microsoft Developer Network, “System.Data.OleDb name space”, <http://msdn.microsoft.com/library/default.asp?url=/library/en-us/cpref/html/frlrfSystemDataOleDbCommandClassTopic.asp> (accessed 30 august 2006)

Improving protection and security awareness amongst home users

P.Bryant, S.M.Furnell and A.D.Phippen

Network Research Group, University of Plymouth, Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

With increased protection making businesses a much harder target, home users are now becoming targeted more significantly. In the past home users have been somewhat neglected and only in recent years have a few surveys emerged to provide some insight into their level of protection.

This paper reports findings from a survey of 415 home users, examining their knowledge of security issues and the usage and protection of their computer systems. It revealed that while users are generally confident about their security and feel they have heightened awareness about the main threats, protection measures such as firewalls and anti virus are not updated regularly enough to provide sufficient protection. There is also a lack of awareness about phishing threats, and insufficient protection against more recent threats such as spyware. This may have worrying implications for the future, with rising uptake and consequently threats from Instant Messenger and Voice Over IP applications and technologies.

Keywords

Security, Home user, Awareness, Perceptions, Survey.

1. Introduction

Many online security threats can pose a serious risk to both home users and organisations. As Internet and especially broadband connectivity continues to increase, home users are becoming even more vulnerable, especially with the wide use of computers for tasks including communication (e.g. email and instant messenger services, surfing, gaming, file sharing, and storing sensitive and personal information). While their awareness of the threats has increased over the years, what they actually do to protect themselves is generally not enough, leading to a false sense of security. Alongside this, users' machines are holding more sensitive information and online services are inviting them to part with vast amounts of personal and financial information, with shopping, banking, gambling and auctions to name a few.

While organisations are tightening their defences' users are becoming more and more attractive targets, especially with the rise of botnets and related denial of service attacks. "In the first six months of 2005, Symantec identified an average of 10,352 bots per day, up from less than 5,000 per day in December 2004" (Symantec, 2005) and denial of service attacks have increased by 51% from an average of 927 attacks per day the first half of 2005 to an average of 1,402 attacks by December 2005 (Symantec, 2006). The home user's machine is the ideal candidate for zombies

from which the resources can be utilised for denial of service attacks or the widespread distribution of spam and phishing emails (MessageLabs, 2005).

In order to gauge the extent of current vulnerability, a survey was conducted to assess users' awareness and understanding of threats and countermeasures as well as the protection practices they currently follow. The intention of the study was to gain an insight into the vulnerability of users and what might be done to help them.

2. A survey of home user security perceptions

In order to understand the current practices and knowledge of users, an extensive questionnaire-based survey was carried out. The survey included mainly closed option (tick box) questions to provide a high amount of statistical data, but it also included open-ended questions where users could add their own thoughts and input (Fink, 1995). Another essential element was to capture demographic details, such as age and education, to assist in the analysis and conclusions. A maximum of 28 questions (excluding the demographics) were included so as not to overload the participants, while also ensuring that enough information was gathered.

To increase the credibility and distribution of the survey it was placed in its own domain on the Internet at www.securityperceptions.net. This was then promoted via email and word of mouth, but also through a variety of web forums such as ultimatereef.net, ukip.co.uk, allotment.org.uk, webuser.co.uk and the studentroom.co.uk, which gave particular access to home users with differing interests and lifestyles who regularly use the Internet. The research was also supported by the Trustguide initiative (a DTI-funded joint project between British Telecom and Hewlett Packard). In order to have a consistent basis for analysis, the respondents were restricted to UK users, and this requirement was constantly pointed out during the promotion of the survey.

2.1 Respondent demographics and background

The survey received 415 responses, a large majority of which classed themselves as intermediate or advanced users, with 58% stating that their highest level of education was at degree level or higher. Therefore the results could be considered as somewhat unrepresentative of the national population. What this does though mean is that the users who completed the survey could be considered as technically and academically more advanced than the general population. This is supported by the fact that only 8% stated they were novice users, while 43% stated they thought they were advanced users. With this in mind it should be considered that the results would be more positive than normal and show more bias that the home users are more secure than they really are. One other point that should be considered is that a number of respondents were from web forums, and so could be considered to be more experienced (particularly using the Internet) than the average user.

2.2. Vulnerability of users

In terms of assessing their vulnerability the results showed that 87% of the respondents had a broadband connection, much greater than the number of dial-up connections. While the advantage in security terms is that it allows updates to be downloaded fairly efficiently, the systems using broadband are at a severe risk if they do not have the appropriate protection and the users do carry out the required procedures to ensure they are protected. The main reasons for this have been highlighted by Furnell (2005), because of the speed that packets can be transferred over the network and unlike dial up broadband is always on:

- The bandwidth can be harnessed for a denial-of-service attack;
- A mass mailing virus or worm program could be deployed;
- The broadband connection can be used as a spam generator.

Out of those with an Internet connection 38% of respondents have a wireless connection, which if unsecure (e.g. without encryption mechanisms so that only the user's machine can connect to it) can provide an additional level of vulnerability. With an insecure connection it can allow another with wireless connectivity within range to connect to the Internet through the network and use the Internet and resources that may become a problem where the victim has a limited download, and the fact that the criminal is getting free Internet access. It is also possible for a hacker to gain access to the user's machine itself, and so use its resources, plus steal or alter files.

2.3 Home Users Security knowledge and practices

The first element to note is that users have a high level of confidence about their security, with 51% indicating that they were 'satisfied' with their security while a further 20% were very confident and additionally 41% felt they understood all the issues and devote time to security. Therefore it is essential to understand how secure these users are to see their level of vulnerability to these threats. The first indication of the level of knowledge possessed by the users was to look at their understanding of the main terminology that is used to explain and refer to the different issues with in security and IT. With this in mind, Table 1 shows a relatively high level of understanding amongst the respondents for what could be considered as the three main issues of Virus, Hacker and Firewall.

	% of Respondents
Virus	99%
Hacker	98%
Firewall	96%
Spyware	89%
Phishing	68%
Identity Theft	92%
Worm	85%
Trojan Horse	83%

Table 1: Respondents' understanding of security terminology

The main area that stands out is that 32% of respondents do not understand what the term ‘phishing’ means. It is therefore likely that they are unaware of the threat, and so will be unlikely to act in an appropriate manner to ensure they do not become victim of a phishing incident. It can be considered a high proportion when the seriousness of the threat is considered with the possibility that users can be fooled into divulging their personal or financial information, such as bank account details, and the ability for criminals to send vast amounts of phishing emails using botnets and spam. Despite this, however, 92% do understand identity theft so they may be aware in some cases the need to be careful with their information, but again they are not likely to be prepared for a well-designed phishing email that appears legitimate.

Worms and Trojan horses are also a cause for concern where 15% of users still do not understand what the terms mean. This is important when considering that these threats have been in the public eye for at least six years or more, particularly with worms such as the Love Bug in 2000 (BBC News, 2000) and the Slammer worm (Clyde, 2003) that both caused significant damage and were publicised across the world.

There is a significant interest in relating this to the use of firewalls because where only 4% did not know what the term firewall meant, 13% do not have a firewall on their machine. In addition 98% stated that they also knew what the term hacker meant. Therefore the survey tells us that while a number of users understand that there are threats and countermeasures they do not consider the risks sizeable enough to warrant protecting themselves against, and so do not have protection such as a firewall installed. This can also be applied to spyware, where 89% understand what spyware means but only 77% actually have anti-spyware installed on their system, leaving 22% unprotected against spyware threats. Naturally it could be that their understanding of the terms is incorrect or they do not appreciate the threat or alternatively have no knowledge of how to protect themselves against it. Furthermore the fact that only 60% of users have anti-spam installed is not only another indication of the failure to protect themselves, but is closely linked to their lack of understanding of the need for anti spam and the threats. This is highlighted by the fact that only 68% understood the term ‘phishing’, which is especially linked and distributed principally through spam. A reason for the lack of understanding and protection against spyware, phishing, spam and even to some extent Trojans and worms, is most likely that there is a failure amongst users to keep themselves up to date with the latest threats and countermeasures and integrate this protection into their machines. This suggests a problem for the future if users carry on with this and do not become protected against future threats, especially where the introduction of new technologies provide new opportunities for criminals that users may not be aware of. Instant Messenger (IM) and Voice over Internet Protocol (VOIP) technologies are significant examples of this, where currently 51% of all respondents use IM and 23% already use VOIP. While the current threat posed is not very significant it is predicted that the threats will greatly increase in the future in terms of number and severity, as pointed out in the following quote. “IM-worms are at the initial stage of evolution. And the fact that the vast majority of the worms are written in Visual Basic demonstrates that most of the authors are fairly new to the virus writing scene and are relatively inexperienced programmers” (Gostev, 2005).

Consequently users need to be prepared for the change in the threats which can be done by ensuring that the importance of keeping their knowledge up to date is passed onto users, as there is no point only knowing about the threats in the past. It therefore needs to be emphasised that the threats will change all the time, particularly with the uptake of new technologies, while the means of protecting themselves and their machines will also evolve.

Related to this has to be the fact that not all the respondents fully understood the role of the technologies installed on their machines, at 27%. While 41% had never attempted to configure their firewall thus leaving the possibility that they are left unprotected, with the default settings and this could also cause the firewall to block a significant amount of legitimate traffic, causing frustration for the user. 12% stated that security impedes their use of the computer. A frequent occurrence with firewalls is that it prompts whether the user wants the traffic blocked or not, there is a possibility therefore that users may permit dangerous connections as they may not truly understand what it means. Alternatively users could just turn off the firewall if it becomes too much of a nuisance.

While having the appropriate technologies installed is one thing, they can be considered useless if they are not regularly kept up-to-date with the latest bug fixes, signature lists or patches to fix holes within the code. With regard to the key security applications displayed in Figure 1, a worrying result is that even though the majority of antivirus, firewall and anti-spyware and anti-spam updates are carried out weekly, these figures are still relatively low, with only antivirus getting a positive response from more than half of the respondents (at 63%), while only a further 20% update the antivirus software monthly. This therefore means that a further 16% are very inefficient in updating their antivirus or never do it, thus leaving themselves very vulnerable as it is particularly important to update antivirus frequently to ensure sufficient protection. Even updating the software on a monthly basis can pose a significant risk to the user, as they will be unprotected against any new threats that appear within that period. During this time they may also become victim of a botnet or backdoor program, which can allow a hacker to control their machine.

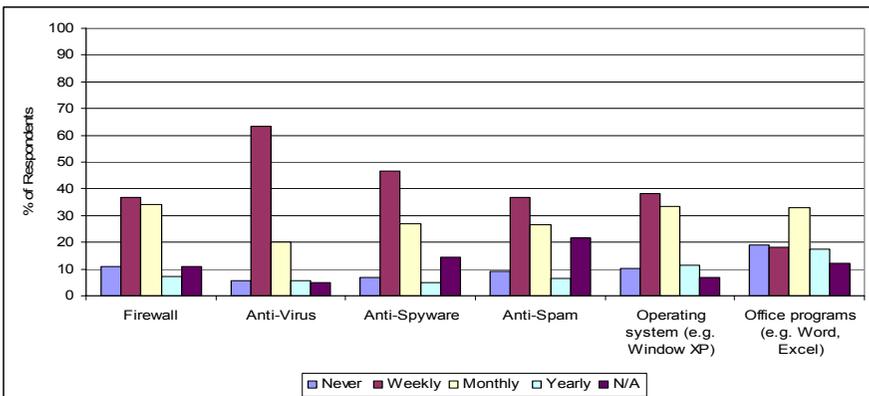


Figure 1: How often do users update applications?

Anti-spyware works in a similar fashion to anti virus software so it is as important to prevent criminals from spying on your behaviour or stealing your data, by ensuring that it is up to date as possible and so prevent new spyware infections. This seems to be another area where user's security is seriously lacking and they are exposing themselves to these dangers as only 47% update on the required weekly basis, while 27% update monthly. This leaves a considerable 26% of respondents who are almost completely unprotected against this threat, by stating that they only update yearly, never or it is not applicable. While it is not as important to update the firewall as regularly, a weekly or monthly basis should provide a basic level of protection for this part. Despite this 29% of users update their firewall either yearly, never or feel it does not apply to them. Thus, only 37% update weekly.

Similarly anti-spam programs do not require the malware list updates of anti-virus and anti-spyware, but the programs may have the requirement to be updated fairly regularly so they can operate at maximum efficiency. Despite this, only 64% of users update their anti-spam on a weekly or monthly basis, subjecting themselves to more spam than necessary which can include phishing emails, while also increasing the possibility that legitimate mail is blocked by the older software.

Additionally security holes within applications that operate on the machine can provide a backdoor past the security technologies, to infect the machine or allow an outsider to spy or even gain control of it and the information within. An essential element for ensuring security is to update these applications with the latest patches as regularly as possible. These exploitable holes occur because of the vast amount of code required to construct the applications, which make it virtually impossible to test the whole system, so it is released with these unknown flaws included. This is especially important with the operating system, which is the backbone for the whole machine and so if it can be compromised in some way then the whole system and everything on it possibly can be as well. Despite this, only 38% updated the operating systems on a weekly basis and a further 33% monthly again leaving 28% of users extremely vulnerable by not updating or only on an annual basis.

Office-type programs, including spreadsheets and word processors, may provide a similar weakness for the system. There is again a deficiency in the percentage of respondents who update their applications on a fairly regular basis at just over half (51%) doing it weekly or monthly. This seems to be an area where there is the least amount of knowledge, as it had the greatest percentage of users who stated that they never update (19%), while 12% put this as not applicable to them.

Another reason that could be used to explain user lack of security could well be the fact that 19% of respondents (second highest response for this question) stated that they felt security packages and services are too expensive. Therefore, it is likely that once an initial subscription with the security software vendor runs out (as its normally supplied when the machine is first purchased) that they do not update this subscription and so are not able to receive new updates.

All this shows that the awareness of updating non-security applications is particularly lacking, with the idea that security only resides with these applications, supported by the results about user awareness of security features. A significant part

of managing user's security is the ability to personalise and ensure that the security settings that exist within a number of programs can be used by the user. For one, most Internet browsers have security features where users are able to manage different aspects of their security and things that they are subject to when browsing the Internet, such as content filtering or more importantly cookie management. Without these they may be less secure than they think. Despite this, only 40% actually know that the security features exist within web browsers, while more importantly only 22% actually understand how they work. Given that 97% use the Internet for web browsing, this is a significantly low proportion, particularly in relation to the amount of users who understand the features.

Other types of programs show a similar picture, with only 39% knowing that security features in office programs and email clients exist, while 37% know about the operating systems features. Moreover 25% of all respondents actually understand the security of office programs, 24% the features in email clients and a lowly 21% who actually understand the security features of the operating system (OS).

Further issues to do with the usability of security come down to the fact that 11% of users stated they do not have time to deal with security issues, and 12% feel that security impedes the use of their computer. This may be considered a small proportion, but if this was taken to apply to the whole population, then 11% is a large number of users from the 62% of the UK population who use the Internet.

3. Discussion

The survey has been able to reveal interesting conclusions that while the majority of users are confident about their security and feel they have heightened awareness about the main threats, there are areas where their practices and knowledge to protect themselves is insufficient. Particularly, while firewalls and anti viruses are installed on their systems, they are generally not updated regularly enough to provide sufficient protection. There is a specific problem in considering malware that even disables these applications, leaving the users extremely vulnerable if they do not update their virus lists in time before becoming infected.

There is also a lack of awareness about phishing threats, and insufficient protection against other more recent threats such as spyware. This may have worrying implications for the future with rising uptake and consequent threats from Instant Messenger and Voice Over IP applications and technologies. More importantly, users are taking large risks by not updating the other applications on their machines on a regular basis. Key to this is the operating system and Office style programs, where coding errors and bugs have been exploited in the past by malware writers and hackers to spy on the user, corrupt or steal their data or control their machine. This is particularly important where a backdoor is created past the anti virus or firewall protection, rendering them useless, while the user has no knowledge and still feels safe while using sensitive information or data on their machine.

Therefore users are not doing enough to protect themselves, so there is a need to significantly improve home user awareness and protection. Additionally, this is

particularly important when considering the increase of the threat financially and especially to individuals: “Attackers appear to be moving away from threats that destroy or compromise data and toward the theft of confidential, financial and personal information for financial gain.” (Symantec, 2006).

4. Conclusion

The survey has provided an insight into home users’ security knowledge and practices of which there are areas that still need to be improved especially concerning knowledge about more recent threats and updating of their applications. It is worth again pointing out that the majority of respondents were advanced or experienced users, and so for the results to be applied to the population of home users the bias towards suggesting greater security awareness and protection needs to be taken into account.

The survey also assessed areas such as what users use their computers and the Internet for, where they might go for security advice and in particular whether users had any knowledge or experience of information sources that are designed to assist the home users, as well as their understanding about reporting security incidents. These elements are under further investigation and have not been covered by this paper.

Particular areas which could be further investigated include how users gain their security knowledge and whether they feel it is useful. An investigation could be applied to all sorts of information sources from the websites to retail stores and even from friends and relatives to understand if this information is appropriate and understandable for different types of home users needs. It is important to know how they learn, and to improve the advice that is given, in order to subsequently improve their protection.

5. References

BBC News (2000), “Papers fall for Love Bug”, Friday 5th May 2000, <http://news.bbc.co.uk/1/hi/uk/736570.stm>, (Accessed 3rd September 2006)

Clyde, R (2003), “Inside the Symantec Internet Security Threat Report”, Summer 2003, Issue 19, Symantec Corporation, <http://www.symantec.com/symadvantage/019/report.html#read> (Accessed 28th September 2006)

Fink, A, (1995), “The survey kit / Vol.2, How To Ask Survey Questions”, Sage Publications, London

Furnell, S (2004), “Hacking begins at home: are company networks at risk from home computers?”, *Computer Fraud & Security*, Volume 2004, Issue 1, January 2004, Pages 4-7

Gostev, A, (2005), “Malware Evolution: January - March 2005”, April 2005, <http://www.viruslist.com/en/analysis?pubid=162454316#mobile> (Accessed 1st September 2006)

MessageLabs, (2005), "Intelligence 2005 Annual Security Report: Cyber-criminals narrow their focus", http://www.messagelabs.com/publishedcontent/publish/threat_watch_dotcom_en/intelligence_reports/2005_annual_security_report/DA_123230.chp.html, (Accessed 31st August 2006)

Symantec Corporation (2006), "Symantec Internet Security Threat Report IX", March 2006, http://www4.symantec.com/Vrt/offer?a_id=22651, (Accessed 31st August 2006)

User security awareness of social engineering and phishing

A.Karakasiliotis, S.M.Furnell and M.Papadaki

Network Research Group, University of Plymouth, Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

Social engineering is a significant problem involving technical and non-technical ploys in order to extract information from unsuspected users. This paper presents an assessment of user resilience to such ploys in the form of email phishing attack. Our experiment used an online web survey which included a mix of legitimate and illegitimate emails and asked users to differentiate between them. A total of 179 participants were involved and the assessment shows that they correctly identified legitimate emails on average of 50%, whereas illegitimate emails were correctly identified on average of 60%. However, in many cases participants who correctly identified illegitimate emails could not reason their selection based on criteria that illustrate their security awareness.

Keywords

Social engineering, phishing, attack, security, awareness, criteria

1. Introduction

Social engineering is a significant and crucial threat that to information system security, both in its own right and as a technique within other threats such as phishing, vishing, and malicious attachments. Over the last few years various sources have highlighted some basic information about the kind of techniques that were used, the success rate of this attacking method and its relation to the user behavior (Leyden, 2004; BBC News, 2005; Silicon.com, 2006).

Harl (1997) defined social engineering as “the art and science of getting people to comply with your wishes. It is not a way of mind control, it will not allow you to get people to perform tasks widely outside of their normal behaviour and it is far from foolproof”. Many other authors, such as Allen (2006), have established this definition as the most common describing term of social engineering in information systems. Nonetheless, it is extremely difficult for someone to define social engineering within just a couple of sentences, because of the myriad ploys that may be involved. So, to add more detail, social engineering is the term that refers to a ‘hacking’ method in which the attacker exploits the user’s behaviour via a series of psychological or/and social ploys, and through technical or/and non-technical communication processes, in order to gain the user’s trust and achieve desired result (normally in terms of getting them to part with information).

The purpose of this paper is to assess user awareness of social engineering threat through phishing attacks.

2. Background

The techniques that attackers may use to extract information from users can be separated in two main categories, namely psychological and technological methods. In our experiments these two approaches were analyzed together in the context of phishing emails.

According to research on the topic of social influence (Petty and Cacioppo, 1986) the main human behaviours that are based on judgment can be separated in two categories that refer to central and peripheral route of persuasion based on the ELM (Elaboration Likelihood Model) theory. Moreover, Cialdiny (2000) mentions that there are six basic tendencies of human behaviour that are responsible for a form of behaviour change-compliance with a request. These influential routes are defined as authority, scarcity, liking, reciprocation, commitment (consistency) and social proof (validation). Furthermore other researchers from the field of information technology have referred to some other behavioural traits such as ‘conformity’ and the ‘desire to be helpful’ (Stevens, 2002), as well as factors of ‘inexperience’ and ‘curiosity’ (Jordan and Goudey, 2005). In phishing attacks, these influential methods can be implemented through the technique of semantic deception (Fette *et al.* 2006), which is achieved through the language used in the text body of email.

On the other hand the technical method that is used to leverage user trust can be performed in other ways. More specifically, in phishing attacks, technical ploys can be defined based on user visual deception (Dhamija *et al.* 2006) through multiple techniques. A phishing attack can contain two main steps; a phishing email and a bogus web site. Moreover it is up to the attacker if he will use further techniques, such as malicious attachments (Everett, 2004) in the email in order to exploit a vulnerable to user system or if he will include a hyperlink in the email body. In most common phishing attacks, the URL redirects the user to a bogus web site in order to collect sensitive personal information such as login credentials (username, password) and financial details (account/PIN number), or alternatively to download a malicious file (Forte, 2005).

Visual deception in phishing attacks can be achieved through many technical ploys, such as masking the fraudulent URL to make the email appear legitimate (Huseby, 2004), and stealing HTML code from a genuine web site (in order to create the bogus one by mirroring it) (Drake *et al.* 2004). Other techniques could involve the inclusion of banners, logos and trademarks to give the email and the web site a plausible appearance. Also, in the email part, spoofing the email address of sender and displaying a URL that contains https could be possible. On the other hand, the bogus web site may contain plausible security indicators, such as padlock icon (denoting SSL, Secure Socket Layer) (Dhamija and Tygar, 2005) and security certifications such as VeriSign.

The aim of the research at this stage was to assess users’ awareness of social engineering via a web survey that would investigate their knowledge of the above ploys and techniques. In common with other experiments (Robila and Ragucci, 2006;

Dhamija *et al.* 2006), our investigation focused on the email part of the phishing attack and specifically on the participants' ability to identify such techniques.

3. Methodology

The experiment used an online survey and included two main sections. The first collected personal information about participants, and included seven questions covering demographic details and technical background (e.g. Internet habits). The second section consisted of 20 questions, each of which presented the user with an email message and asked them to consider whether or not it was genuine or a phishing attempt. Each question had three options (illegitimate, legitimate and do not know) and an optional text box for participants to briefly explain their reasoning.

The demographic questions collected general details such as gender, nationality, and age in order to separate participants into different categories and make a comparison analysis between these categories. In the second part of the demographics we tried to investigate if the security awareness is related to the educational or employment background of the participants. The last part of the demographics asked participants about their internet habits (e.g. online shopping, e-banking, online purchases of bill payments, etc.) and the protection mechanisms (e.g. anti-phishing toolbar) that they use against phishing threat by giving three answering options (yes, no and do not know).

The design of the second part of the survey was more complicated, as it was based on a series of criteria that were related to methods of phishing. So the 20 email questions were separated in two main categories of illegitimate (11 email snapshots) and legitimate (9 email snapshots). As already mentioned above, each question had three possible answers, with the 'do not know' option set as the default..

The main concept of the survey was to benchmark participants' responses to different technical and influential ploys that attackers may include in their emails. The 20 messages used as the basis for the questions were selected from a combination of anti-phishing advisory sites, as well as from emails received by the investigators themselves. The emails included representation from a variety of online services that could be phishing targets (e.g. e-banking 8/20, online-shopping 4/20, online purchases of bill payments 5/20, etc.), as well as a range of different attacking techniques (such as collection of personal and financial information from bogus web site 7/11, downloading malicious software 1/11, opening malicious attachment 1/11, vishing 1/11 and PO BOX 1/11 technique). Figures 1 and 2 below illustrate examples of the illegitimate messages used in the study.

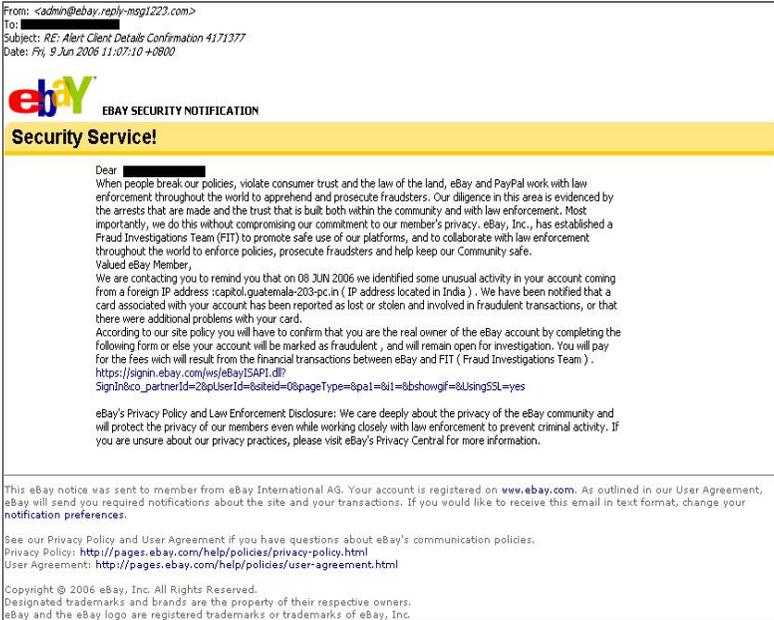


Figure 1: Illegitimate email from eBay

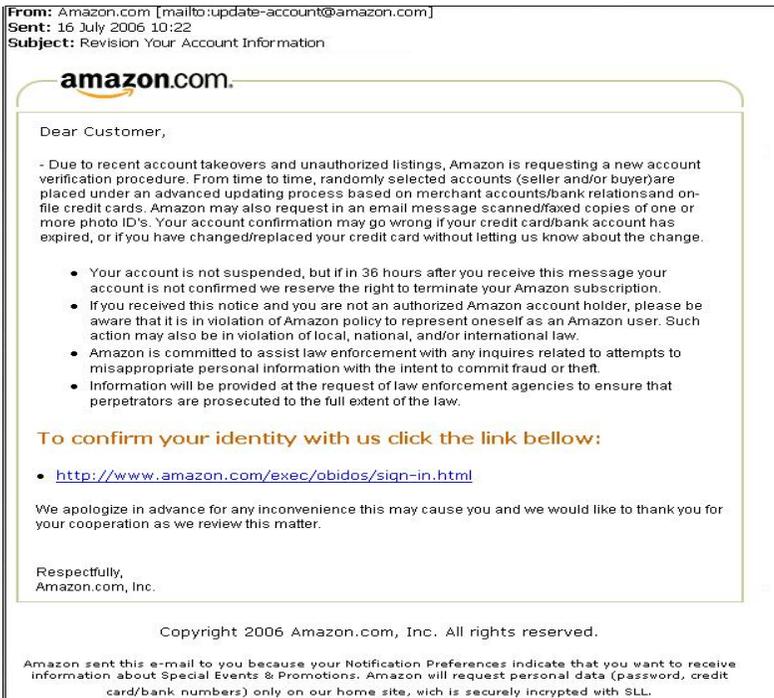


Figure 2: Illegitimate email from Amazon

Other design criteria were based on visual aspects in order to investigate the role that these can play in deceiving the participants. So, 14 emails were colored snapshots

with logos, banners, trademarks, etc. (six legitimate and eight illegitimate), whereas the other six were plain text messages (three legitimate and three illegitimate).

Other characteristics represented in the emails included typos, errors or even grammar mistakes (in 7 of the 11 illegitimate messages). Another language-related factor had to do with the influential techniques that the emails used, with a range of persuasive methods being represented, such as scarcity (2/11), authority (7/11), social proof (2/11) and desire to be helpful (1/11), with some emails including a combination of techniques.

4. Experimental results

A total of 179 participants (75% of whom were male) filled out the survey, which was available on the Internet for a period of 19 days. The requirements for someone to participate to our study were the understanding of the English language (as the emails were written in English) and the use of Internet. The total population of participants included representation from 22 different nationalities, with the majority (97%) having a higher education qualification. In terms of age groupings, 76% were aged 18-29 and 24% above 30.

According to the Figure 3 below that represents the total results of the participants for each question separately we can observe that in most cases, opinions are very divided and there is only a small number of cases where respondents clearly come down to one side (questions 3, 14, and 20) and in some cases they answered dramatically wrong (question 3).

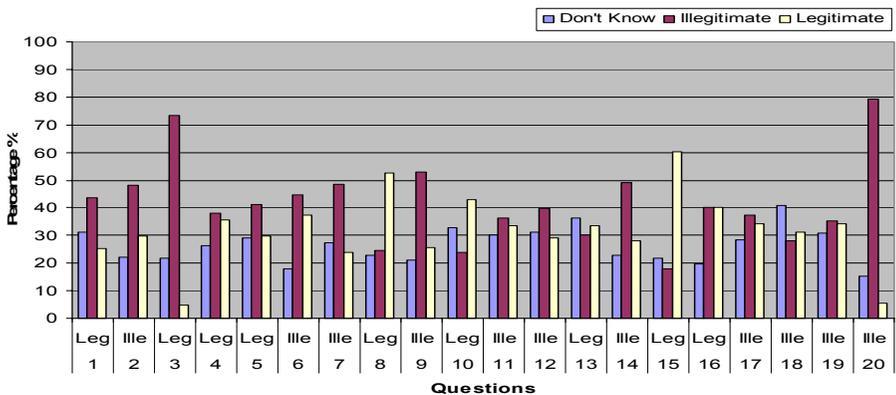


Figure 3: General results from survey

The overall success rate of participants in correctly identifying illegitimate emails was 42%; on the other hand 32% could correctly identify a legitimate email, and 26% selected the 'do not know' option (thus illustrating the confusion of the participants). From the analysis based on demographics we determined that there were no significant differences relating to age or gender, but there were notable changes based on the participants' work/study background. More specifically, 31% of the total population was related to the IT/Computing sector. From these

participants the success rate of identifying correctly emails was 25% for legitimate, 52% for illegitimate and 23% for those that didn't know. Also some changes compared to general results were observed when measuring subsets of participants based on their Internet habits. These changes showed a small increase of success in identifying illegitimate messages, but not as much change for identification of legitimate emails.

Feedback comments were left by 47% of participants, enabling a deeper analysis of their criteria for judgment in each case. We observed that 40 participants made judgments based on visual indicators, such as the presence of logos, banners, trademarks, footer, fonts and copyright symbol. From those participants, 55% selected the legitimate option. The results also showed that a plain text email was more likely to be judged as illegitimate compared to messages with color and images.

Considering the influence of technological factors, 52 participants made a judgment based on whether email contained a URL (70% select the illegitimate option). Furthermore, 26 participants mentioned the fact of http or https, and 39 made a comment about using a verification process (e.g. "have to check this by opening a browser window and typing the given URL into this"). Only 12 of them manually checked the correctness of the URLs and 40 participants made a selection based on the given email address. Considering judgments of personal information, 18 participants gave an answer based whether the email contained a recipient name or not, and 67 participants did so based on other personal information.

From the perspective of language, we understood that 19 participants focused on the language mistakes such as typos and grammar errors. Moreover 34 participants selected answers on the basis of emails that claimed to offer opportunities, while 26 did so based on emails that used forceful language. Also from an analysis of influential techniques it seems that the techniques of authority and desire to be helpful are the least correctly identified from the participants, compared to the techniques of social proof and scarcity.

5. Discussion

Our experiment revealed a significant failure by participants to correctly classify the emails (with average of do not know answers 26%). Comparing our findings to similar studies (Robila and Ragucci, 2006) we highlighted a slight difference in the overall results. More specifically, Robila and Ragucci (2006) mention that participants were able to correctly identify legitimate and illegitimate emails in 60% and 53% of cases respectively. Our findings showed that participants did the same on average of 50% and 60% respectively. However, this difference could possibly exist because of a series of reasons (e.g. different number of participants, different emails, and the addition of a 'do not know' option in our case).

Moreover from the investigation of visual attention, language attention and technological awareness we revealed interesting findings. More specifically participants made a selection based on incorrect criteria in many cases (e.g. based on

logos, copyright symbols and footers in the emails, and in other cases influenced from the type of language that the emails used). Moreover, the fact that many participants mentioned technological aspects in the email in order to support their thesis shows a level of security awareness. However, in many cases the participants mentioned these aspects but their reasoning led them to draw the wrong conclusion about legitimacy.

6. Conclusion

The practical study was a good idea to investigate the phenomena of social engineering through phishing attacks with emails. The need for security awareness on the topic is imperative, but the way to achieve such awareness could be a difficult process due to the technical unfamiliarity or the behavioral traits of each user.

Future work could address a deeper analysis of how individual factors (such as visual, technological and language characteristics) have an effect, rather than having messages that include combinations of several of them. Another possible direction for future work would be the use of emails with real links instead of email snapshots, as this would enable more interactive exploration by the users (thus more accurately mirroring the real-life scenarios in which phishing messages would be encountered).

7. References

- Allen, M. (2006), Social Engineering: A means to violate a computer system, SANS Institute, http://www.sans.org/reading_room/whitepapers/engineering/529.php, (accessed 10 August 2006)
- BBC News. (2005), How to sell your self for a song, BBC News, <http://news.bbc.co.uk/1/hi/technology/4378253.stm>, (accessed 06 August 2006)
- Cialdini, R. (2000), Influence: Science and practice, 3rd edn., New York: HarperCollins, ISBN: 0-3211-8895-0
- Dhamija, R. Tygar, J. D. (2005), Phish and HIPs: Human Interactive Proofs to Detect Phishing Attacks, in Proceedings of the Second International Workshop on Human Interactive Proofs, H.S. Baird and D.P. Lopresti (Eds.): HIP 2005, Springer-Verlag Berlin Heidelberg, pp127–141.
- Dhamija, R. Tygar, J. D. and Hearst, M. (2006), Why Phishing Works, to appear in the Proceedings of the Conference on Human Factors in Computing Systems (CHI2006), Montréal, Québec, Canada, pp: 1-10
- Drake, C. Oliver, J. J. and Koontz, E. J. (2004), Anatomy of a phishing email, in Proceedings of the First Conference on Email and Anti-Spam (CEAS), 2004, <http://www.ceas.cc/papers-2004/114.pdf>, (accessed 12 August 2006)
- Fette, I., Sadeh, N. and Cranor, L. (2006), Web Security Requirements: A Phishing Perspective, Carnegie Mellon University, http://www.w3.org/2005/Security/usability-ws/papers/13-cmu_requirements/#search=%22Web%20Security%20Requirements%3A%20A%20Phishing%20Perspective%20Fette%22, (accessed 30 August 2006)

Forte, D. (2005), Spyware: more than a costly annoyance, *Network Security*, Vol. 2005, No. 12, pp8-10.

Harl. (1997), People Hacking the Psychology of Social Engineering, Text of Harl's Talk at Access All Areas III, <http://www.noblit.com/docs/people-hacking.pdf>, (accessed 10 August 2006)

Huseby, S. H. (2004), *Innocent Code: A security wake-up call for web programmers*, John Wiley & Sons. Ltd, Sussex, U.K., 0-470-85744-7.

Jordan, M. and Gouday, H. (2005), The Signs, and Semiotics of the Successful Semantic Attack, 14th Annual EICAR Conference 2005, St.Juliens/Valletta, Malta, ISBN: 87-987271-7-6, pp: 344-364.

Leyden, J. (2004), Brits are crap at password security, *The register*, http://www.theregister.co.uk/2004/04/20/password_surveys/, (accessed 06 August 2006)

Petty, R. E., and Caciopo, J. T. (1986), *Communication and persuasion: Central and peripheral routes to attitude change*, New York: Springer-Verlag.

Robila, S. and Ragucci, J. (2006), Do not be a Phish: Steps in User Education, *ACM SIGCSE: Vol. 38, No. 3*.

Silicon.com. (2006), What the security 'stitch-up' should teach as, *Silicon.com*, <http://software.silicon.com/security/0,39024655,39156525,00.htm>, (accessed 06 August 2006)

Stevens, G. (2002), *Enhancing Defenses Against Social Engineering*, SANS Institute, GIAC, http://www.sans.org/infosecFAQ/social/defense_social.htm, (accessed 10 August 2006)

Evaluating the Perceptions of People towards Online Security

N.K.Jayakumar and A.D.Phippen

Network Research Group, University of Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

People started using online services in day to day life from the period; when they came to know about the use and ease of the technology. Services like online banking, online product purchase or sale, instant messaging, electronic mail, online voting, medical, paying tax, etc are mostly commonly used by the people. The growth of the online commerce is the reason for the significant increase in the use of online services and also increase in the variety of fraudulent web activities. Almost all the users are the victim for the internet threats, in some way. So the project aims in evaluating the perceptions of individual users about the online security and its issues. The perception of online issues differs from each and every individual like 'Normal User, Internet Users and Security Professional' because the normal user just operates/uses the system without having much knowledge about the causes and threats that are caused due to online services whereas the internet users are known to some available threats like virus, spyware, malware, etc but the security professionals are known to all kinds of threats that are available and its causes. Security professionals always keep themselves updated with the change in security issues and its threats because it's their part of the profession, they are known to latest virus, worms, malware, spyware, intrusion option, cookies, firewall, trojan horses and other security related options provided on the client/user systems etc. In this project, a survey is taken as a part of research from the people to reveal their views towards online security; which will increase the awareness about the threats that are caused due to the online transaction and services. The final survey result concluded that the users/people are known to some threats but not all, and there are some hindrances for the user from utilizing the full security, which have been discussed in the final part.

Keywords

Online security issues, Security threats, Grouped survey discussion

1. Background

In today's world almost everything is done with the help of online services 'Buying/Selling/Ordering/Contracting/Renting of products, Bank transaction, E-mail, Messaging, Online Vote, Tax payments etc', the use of these kind of services was increased recently due to the rapid growth of the internet during the past decade. Almost all the users of the system find that the use of technology was an ease but they are not fully aware about the causes and threats behind it. Own instance: - The People instead of going to shops to purchase the needed products; they started to use online to order things to their door steps. The online transaction made mans life easier in processing the information they need; and to do things faster and in an easier way, but it started to cause threats which is mostly unknown to the users because of the awareness towards the causes. Threats caused to the users / people in

the sense; during the online transaction to purchase a product from an online store, the user enters the personal information like credit card number, security code, card holder name, etc., which has been transmitted over the internet for placing the order but the user is not sure whether the given details might be hacked by someone for their personal use or whether it reaches the exact destination for placing the desired product order. The users of online services are facing most of the problems which are not known to them directly but the effects of the problem are caused indirectly by any means of source like 'Phishing attacks, Scam, Unsolicited adverts, Identity/Privacy theft, Virus, Spam' etc.

2. Importance of Security to prevent the online threats

Online services are used by the people in day to day life, which runs fully on the internet where the whole world can access the information stored on it and thus making it an insecure environment for someone who wants to be secure (Joris Evers, 2006). As the service started to grow day by day, the new threats have evolved creating a fear among the users, making them to be more concerned about 'what the particular threat is going to do...?' (Michael Bruck, 2006) and it appears to be like 'Internet/Online' has lots of threats when compared to other technologies that have been developed to ease the living of human being (Andy Sullivan, 2006) the main threats which affects the user system the most are 'Worms, Break-ins, Hackers, Crackers, Hi-jackers, Phrackers, Spoofing, Password sniffing, Denial of service' etc. There are still lots of users using the online service who are very vulnerable to attacks mainly due to the awareness of the nature of the attacks and they still believe that a good password is the only need to be concerned about and they are not worried about the behind threats.

3. Project Survey Procedure

The project work was started with the aim in finding the user's level of awareness in every security aspect and issues 'what they think, why they think like that, what are their fears in threats....'Etc. So after finding the awareness level of the user/people; it would be easy to eradicate the problems just by knowing the right steps to be taken 'whether the problem exist within the user due to unawareness or the problem exist within the developed software/application'(Robert Vamosi, 2006). For this finding, the project works by taking a survey from the user/people which enquires them in all aspects of security issues about the known and unknown threats. The survey is not only taken from students who are doing their degree but also from workman, professionals, normal person who does not know about the security issues, this kind of survey is taken to reach in-depth level of people awareness.

4. Discussion from grouped survey graph

The Figure 1 says that, users fully agree that they are known to 'Existence of viruses, Causes of viruses, Concerned about damage caused by viruses and Regular scanning of computer for virus' but their level of agree to 'Regular updating to virus definitions' is low when compared to other issues, it is because users/people normally forget that virus definitions is most needed part of the antivirus software, so

in total nearly 87% of users are agreeing that they are known to issues of virus threats and 13% of users are not known to the issues of virus threats.

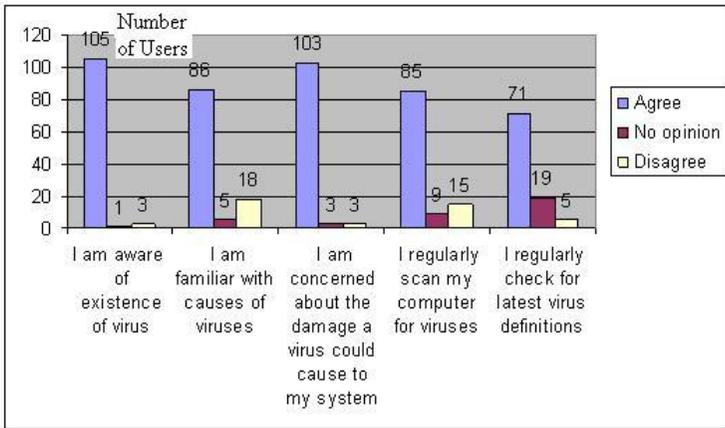


Figure 1: Grouped survey graph for virus threats

The Figure 2 says that, users fully agree that they are ‘Aware of spam attacks, Receiving unsolicited advert in email, Use filters to block spam, and Receiving spam daily’ but their level of agree to ‘Reporting spam to provider’ is very low when compared to other issues, it is because user/people do not want to waste time in doing it, so in total nearly 79.5% of users are agreeing that they are known to issues of spam threats and 20.5% of users are not known to the issues of spam threats.

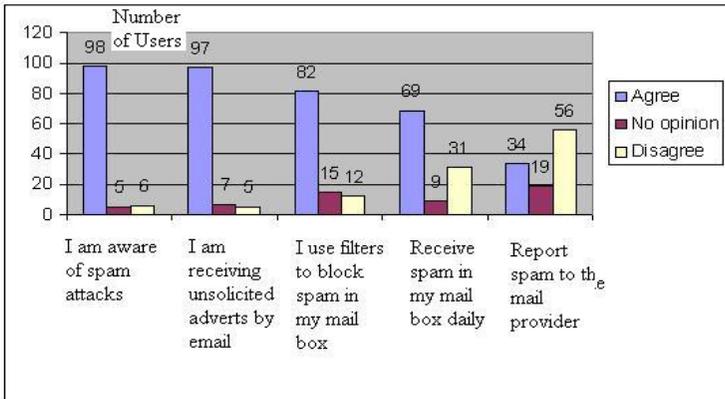


Figure 2: Grouped survey graph for spam threats

The Figure 3 says that, users fully agree that they ‘Always open email from family and friends, Always open email from known companies, Always open email from unknown senders, My responsibility to protect my system from threats and Scan all mails using mail scanner’, so in total nearly 79.5% of users are agreeing that they are known to issues of Email threats and 20.5% of users are not known to the issues of Email threats.

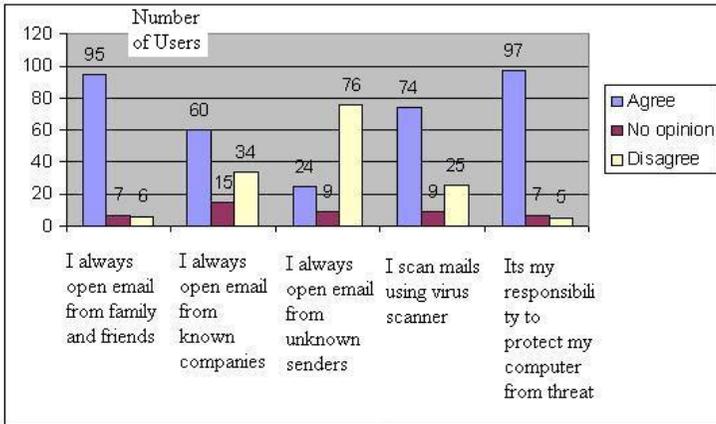


Figure 3: Grouped survey graph for email threats

The Figure 4 says that, users fully agree that they ‘Disclose bank/card details during online transaction, Always use online banking, Feel risk at online fraud, Have required knowledge to protect the computer’ and their level of disagree to ‘Disclose bank/card details via email’ is very low when compared to other issues, it is because user/people is fully aware of not to disclose anything in email messages, so in total nearly 52.5% of users are agreeing that they are known to issues of online threats, also using the online transaction in a safer way and 47.5% of users are not known to the issues of online threats, disclosing the bank/card details online will increase the chances of threat.

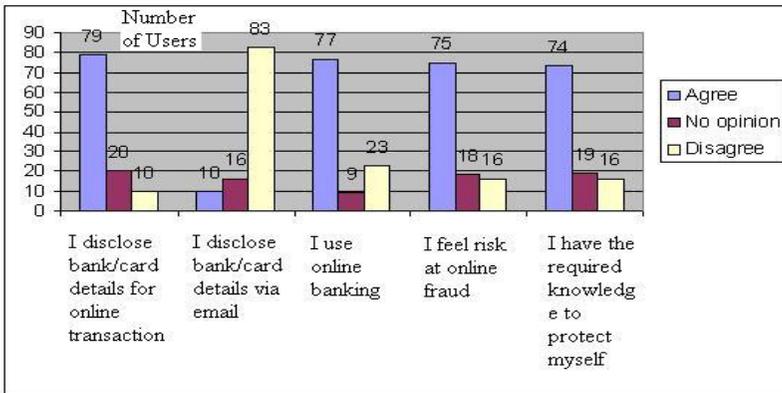


Figure 4: Grouped survey graph for online transaction threats

The Figure 5 says that, users agree that they are ‘Happy to fill tax return online, Happy to vote online, Happy to give medical details online, Worried about amount of data held about them in online’, it is because user/people are fully aware of what they are doing in online life, so in total nearly 50.5% of users are agreeing that they are known to issues of online life threats and 49.5% of users are not known to the issues of online life threats.

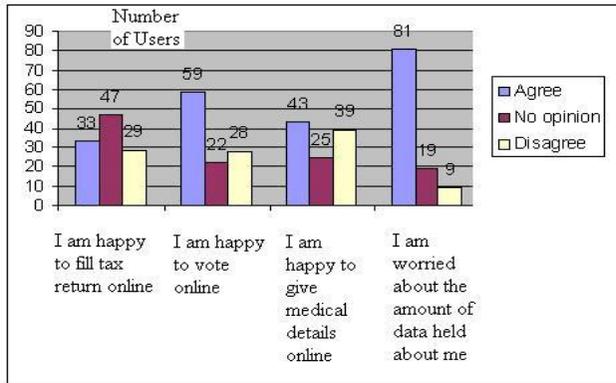


Figure 5: Grouped survey graph for online life threats

5. Limitations

The project research has analysed that there is no limitation in online service and security issues until or unless, the users of the system are getting aware of the upcoming/new online issues and attacks. If the user stays behind without updating their knowledge about the available threats; they might be one of the victims who will be attacked by the online threats, which will led them not to use the online services due to the fear. So its better all the users make themselves updated to latest threats, which will help them to get protected from the attacks and will be able to use the online service without any further interrupts.

6. Future work

Almost all the security companies like 'HP, Symantec, McAfee, eTrust, BT, Lavasoft...'etc are trying to implement a perfect secure system for the users to use in the future without any threat/attack in online services, for this purpose 'HP' has started taking survey to understand a depth knowledge of user 'In what areas of online service, the users faces the main problem..', (Peter Szor, 2006) their survey resulted in users awareness as a main conclusion, so the HP company has decided to give a clear view about the software they use or terms and condition they sign or company software licence. Likewise all the security software's features are to made short and clear, so that the user themselves can make a proper configuration and can prevent their system from threat attacks.

7. Conclusion

The project is concluded with the awareness level of the user related with online security and its issues, from the discussion of grouped graph it is clearly understood that the user is known to the 'Threats caused due to Viruses, Spam, and Email' because they are able to see and understand that they have been attacked by someone or something, but they are not known to the 'Threats caused due to Online transaction and Online life' because these threats cannot be identified by the user until they are known later that they have been attacked. We know a saying

‘Prevention is better than cure’ likewise it is better that the user being aware to all the existing/new threats, so that they can prevent themselves from being attack instead of noticing later that they have been attacked due to online threats. Not all the people will be able to browse the internet and come to know about the latest threats available because they might be very busy at work or did not find any time to search for new threats. So it is better that the security companies like ‘HP, Symantec, McAfee and others’ start to release threat notes monthly or quarterly year that should be easily readable and understandable even by a normal person, all the companies should take consideration from Normal user – Expert user in case of security issues and threats, so that it will easy for all level of users to be aware of online attacks.

8. References

Bruck, M. (2006), “Security threats from within people”,
<http://www.entrepreneur.com/technology/managingtechnology/article503414.htm>, (Accessed 26 June 2006)

Evers, J. (2006), “The security risk in web”,
http://news.com.com/The+security+risk+in+Web+2.0/2100-1002_3-60992128.html,
(Accessed 24 June 2006)

Sullivan, A. (2006), “Internet technologies emerge as new online threat”,
<http://www.forbes.com/markets/newswire/2006/06/26/rtr1335591.html>, (Accessed 26 June 2006)

Szor, P. (2006), “Security system extends relationship with HP”, <http://phx.corporate-ir.net/phoenix.zhtml?c=933082&p=irol-newsArticle&ID=309055&highlight=> ,(Accessed 29 June 2006)

Vamosi, R. (2006) , “AOL Active Security Monitor”,
http://reviews.cnet.com/AOL_Active_Security_Monitor/4505-3667_7-319294263.html,
(Accessed 28 June 2006)

Intrusion Detection System for mobile devices

D.S.Michalopoulos and N.L.Clarke

Network Research Group, University of Plymouth, Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

Mobile devices are getting very popular these days. Most of people use mobile phones and now, as the new generation of smartphones, the capabilities that are included in a portable palmtop device are amazing. However, single user authentication is not enough as their compact size makes them easy to be stolen. The variety of frauds on these devices has lead to the development of an intrusion detection system, that acts as a constant and transparent authentication mechanism. In this paper an algorithm for host based intrusion detection in mobile devices has is developed. Normal Intrusion Detection Systems use signature checking algorithms, something that can not be implemented easily on a mobile device. This system is based on statistical models that try to identify anomalies in users' attitude by utilizing a profiling procedure and then requires a second layer of authentication from the user, where at the same time network administrator is warned about possible abuse.

Keywords

Mobile, Security, Fraud, IDS, Wireless

1. Introduction

The main aim of this project is to develop an intrusion detection system (IDS) for mobile devices. Single authentication mechanisms are not enough and many frauds take place (Clarke&Furnell, 2005). As a result our thought is the development of a continuous and transparent authentication system capable of identifying any possible abnormal attitude. Before we analyze the designing plans of our system, some basic principles of the intrusion detection have to be understood. First of all the idea of intrusion detection is taken from desktop and laptop computers as they are developed before the mobile devices and there are many problems in terms of information security. The evolution of the mobile devices was very fast and started during the previous decade. The first of them were capable to perform phone calls through the GSM network, store names and numbers, provide calendar services like date reminding etc, storing notes etc (Mobitedia, 2006).

The last generations of these devices are far more developed. The definition of palmtop computers is precise enough. Having a processor, memory, operating system and many other computer infrastructure they can support many type of applications and of course at the same time to be normal cell phones. As an example for their evolution we can give Microsoft's windows mobile operating system, specially designed for mobile devices.

What is in our minds is to create a system, capable to identify frauds and misuses at the mobile device (Lundin&Jonsson, 2000). In order to achieve that, we install

multiple sensors on the device that monitor user's attitude and keep records of it. Then, as the device is in operation, the captured traffic is compared with what was kept and a threshold. In case there is enough aberration the IDS is activated, warning the user and the network about the potential danger.

In section 2 literature review is analyzed providing information about the sources and the references of this project. In addition the research method that has been followed is analyzed giving details on the aspects on where the author was focused on. In section 3 the main framework of the IDS is presented analytically providing all the necessary details. In section 4 some advantages and disadvantages of the framework are discussed

2. Literature review

Our aim is to develop a system that is capable to protect all kind of mobile devices. However, not all of them have the same capabilities. What is more, data connections for these devices are many, for example Bluetooth, GPRS, UMTS, infrared etc. Now, in order to develop this project, we need to research in depth existing vulnerabilities and further security aspects in this area.

As the project is more focused on the development of a host based anomaly detection IDS research is needed on these fields. In (Lundin&Jonsson, 2000) some effective strategies for intrusion detection are presented, like the mobile agents one. In addition in (Farshchi&Jamil, 2006) useful ideas about wireless intrusion detection are presented. What is more, in (McGraw, 2005) some interesting thoughts are presented about the necessity of effective security mechanisms on mobile devices and mainly for intrusion detection systems. Furthermore, in (Kemp, 2005) some interesting thoughts about effective tactic on intrusion detection are presented, some of them useful for our research. Besides, in (Elison, 2006) there is a very good discussion about potential problems in general for intrusion detection.

The system that it is designed, acts like a constant transparent authentication system. As a result, before the designing attempt of the IDS, a research for authentication methods is needed. What is more, a research for biometric authentication mechanisms is also suggested, as the method with the threshold, where the system decides whether the user is authenticated or not, can be used with a similar to the IDS identifying whether an activity is normal or not (Clarke&Furnell, 2005).

In a recent survey presented in (Clarke et al 2002) we can see that PIN is the only authentication procedure that is used. Indeed, sometimes it is not used properly and as a result frauds occur. Another interesting point is the concern that users express for security aspects. This is certainly an optimistic message to continue our work and achieve our goals.

2.1 Statistical analysis

Mobile devices are mobile phones as well and one of their basic functions is telephony. As mentioned above, by the increasing processing power and storage

capability of a device like these, statistical data about the calling activities and the user's roaming can be kept in the device and be analyzed. In case new data get over some specific limits, a warning signal may be sent. A very good algorithm is IDAMN, developed by (Samfat&Molva, 1997). Although it is a bit complicated, a very good point is the user classification they use: Domestic, Business, Corporate and Roamer according their usage. However, the complicity of the statistical data makes me think that it requires a lot of processing power where at the same time the percentage of false alarms may be high. On the other hand, a similar, simpler, algorithm can be used gathering and analyzing data, giving warnings when something unusual is happening. Finally, one more disadvantage may be the possible leak of personal data that may occur from gathering all these personal data of calls and roaming (Clarke et al 2002; Clarke and Furnell, 2005)

3. IDS framework for mobile devices

Now, all parts of the system need to be implemented at the mobile device.

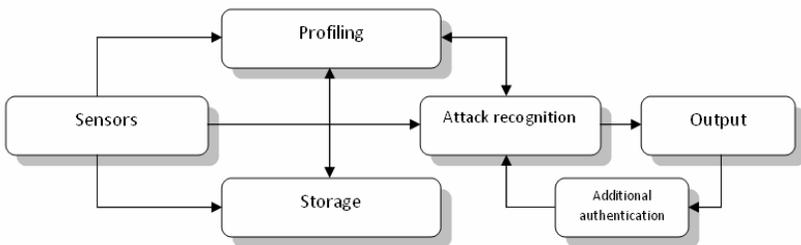


Figure 1: The plan of the IDS

Sensors are implemented at the mobile device monitoring all the network activities. Parts for monitoring telephony, user's roaming, exchanged SMS/MMS, wireless network activity, e-mail activity, file accessing applications, Internet activity, Bluetooth connections, Infrared activity, web browser activity, new applications installation. What is in common in all the above sensors is the common way an attack is recognized. By having kept a statistical amount of previous activity, which is hypothesized as normal, it is comparing with the received data and a threshold. In case the difference is higher than the allowed then the IDS is activated (Samfat&Molva, 1997).

Data are collected from sensors and then they are forwarded to storage and profiling session. At the second one, user's profile is created based on the data collected from the training period. This time is approximately one month but it is variable and it depends on the time that the device is used on daily basis. In other words the training period lasts as time as it is needed for enough data to be captured for profile creation. An alternative solution, more efficient for powerful devices as it requires more resources is when the profile is created from the activities of the last time interval, large enough in order to calculate users profile, average one month.

Indeed, when the users starts his/hers device for first time some brief questions are asked in order to be categorized in one of the default profiles. These are:

- Student
- Domestic
- Business
- Roaming

The categorization in these profiles except from providing data for protection during training period, it gives the opportunity for variable thresholds in attack recognitions. Different users do not have the same security expectations and for example thresholds in business profile can be tighter than the home one. This categorization gives the opportunity to the system to treat each user with a different way, more suitable to his/hers needs.

3.1 Hardware capability

This project is designed to be implemented to a wide range of devices. This of course is not possible as not all devices have the same capabilities. For example is there is no need to implement Bluetooth sensor in a device that does not support this protocol.

Group	1	2	3	4	5
Sensors	Telephony, Roaming, Exchanged SMS/MMS	All the previous plus Bluetooth, Infrared	All the previous plus Wifi connections, Internet activity	All the previous plus Web browser, E-mail activity	All the previous plus monitoring new applications installation

Table 1: Groups and sensors

Now, the mobile devices have to be categorized into these five groups according to their hardware capabilities. However, this is not simple as many vendors do not publish any details about their products. Actually, devices are categorized basically according to their CPU speed or memory. Table 1 presents a proposed categorization of the sensors into 5 groups. Of course this is not a necessity a device may be very powerful to have all sensors however some of them may be useless and not implemented. On the other hand, in case a devices performance is slowed down, some of the installed sensors can be removed. What is achieved with this method is that the designed IDS can be implemented to a wide range of devices.

3.2 Output

When a potential threat is recognized from the IDS then there are two levels, the warning and the alarm one. By the first time something unusual is recognized, the first (warning) level warns the users about potential abnormal activity. Then, he/she has the opportunity to authenticate again his/her self to continue that activity. In case this goes on, then the IDS is activated at the second layer, the alarm one. Then the network administrator is informed that a continuous abnormal activity takes place at the specific device. Now, additional authentication is mandatory, not optional as it was at the first layer. What is proposed for this a biometric authentication algorithm.

That on (Lodi et al 2002) is a very good one for voice recognition, not only by requiring low system resources, but also by its very effective results.

4 Discussion

This system is developed in order to recognize and identify abnormal activity on a mobile device. This recognition however, is based on the way user behaves with his/hers mobile. A potential vulnerability of the system is the fact that in case of an emergency, when user performs more calls and generally the whole activity of the system is increasing, the system is activated by false. Indeed, this reduces the reliability of the system and makes it more irritating for the user. This is the reason the optional additional authentication process is added at the first warning level.

What is more, the system needs a reset function/button. In case the device is sold again as second hand, all the previous settings in the IDS (profiles etc) need to be deleted and the training period needs to be started again for the new user. However, lots of care has to be given at this point as an attacker may be capable of resetting the IDS and perform his/hers malicious acts without a problem.

In general, it can be argued that the system is not able to follow any changes at user's attitude. For example, if someone changes his/hers job, then possibly the way he/she is using the device is changed as well. This of course is not similar with the attitude at the training time and consequently the IDS starts generating false alarms, something irritating. Once more, a reset function is necessary but with paying lots of attention at the authentication process of the person that performs the reset.

Furthermore, this system acts by capturing data from all kind of activities of the device. This however may be irritating for the user. Many people may feel that they are monitored from their own device, that their actions are captured. As a result, lots of care has to be given in order the system is not used for malefic purposes. For example, it can be modified and used by a company manager, to watch and monitor employees' activities, especially by monitoring their roaming.

In addition, the system is not capable of protecting the owner from attackers that know users attitude. For example, when an attacker knows the way that the victims uses his/hers device and the fact that this IDS model is implemented, then he/she is capable of stealing it and not be identified just by using it as its owner used it.

5. Conclusion

This framework is developed in 2006. As it can be argued that the area of mobile devices is a rapid developing are, the current situation is possible to be changed in a sort time. A future work can update this project by recategorizing them to new groups according to device resources of the exiting time. As devices are getting more powerful new sensors can be added and of course new groups covering all kind of activities. The author would be grateful to see this work updated and developed in future.

6. References

- Barber, R. (2001). "Security in a mobile world – Is Bluetooth the answer?" *Computers and Security*, v 20, n 5, 2001, p 374-379
- Clarke, N.L, Furnell, S.M., Rodwell, P.M., Reynolds, P.L. (2002). "Acceptance of subscriber authentication methods for mobile telephony devices" *Computers and Security*, v 21, n 3, 2002, p 220-228
- Clarke, N.L, Furnell, S.M.(2005). "Authentication of users on mobile telephones - A survey of attitudes and practices" *Computers and Security*, v 24, n 7, October, 2005, p 519-527
- Clarke, NL., Furnell, SM. (2005). "Biometrics - The promise versus the practice Source: *Computer Fraud and Security*", v 2005, n 9, September, 2005, p 12-16
- Elison, D. (2006). "Intrusion Detection, Theory and Practice" <http://www.securityfocus.com/print/infocus/1203> accessed 1/2006
- Farshchi, J. (2006). "Wireless Intrusion Detection Systems" <http://www.securityfocus.com/print/infocus/1742> accessed 1/2006
- Furnell, S. (2005). "Handheld hazards: The rise of malware on mobile devices" *Computer Fraud and Security*, v 2005, n 5, May, 2005, p 4-8
- Hager, C., Midkiff, S. (2003). "Demonstrating Vulnerabilities in Bluetooth Security" *Conference Record / IEEE Global Telecommunications Conference*, v 3, 2003, p 1420-1424
- Innella,P. (2006). "The evolution of Intrusion Detection Systems, Tetrad Digital Integrity", LLC, 16 November 2001, <http://www.securityfocus.com/print/infocus/1514> accessed 1/2006
- Ivo, P. (2003). "Bluetooth and security" *Proceedings of SPIE - The International Society for Optical Engineering*, v 5445, *Microwave and Optical Technology 2003*, 2004, p 55-59
- Hynninen, J. (2006). "Experiences in Mobile Phone fraud" *Helsinki University of Technology* <http://www.niksula.hut.fi/~jthynnin/mobfra.html> accessed 7/2006
- Kachirski, G. (2006). "Effective intrusion detection using multiple sensors" <http://doi.ieeecomputersociety.org/10.1109/PDCAT.2005.34> accessed 1/2006
- Kafi H., Conner, M. (2003). "Identifying security threats in ad hoc wireless network" *Proceedings of the International Conference on Security and Management*, v 1, *Proceedings of the International Conference on Security and Management, SAM 2003*, 2003, p 34-38
- Kemp, M. (2005). "For whom the bells toll: Effective IDS deployment strategies" *Network Security*, v 2005, n 5, May, 2005, p 16-18
- Kitsos, P., Sklavos, N., Papadomanolakis, K., Koufopavlou, O. (2003) "Hardware Implementation of Bluetooth Security" *IEEE Pervasive Computing*, vol. 02, no. 1, pp. 21-29, January-March, 2003
- Krugel, C, Toth, T. (2006). *Applying Mobile Agent Technology. To Intrusion Detection*. www.infosys.tuwien.ac.at/Staff/tt/publications/Applying_Mobile_agent_Technology_to_Intrusion_Detection.pdf accessed 1/2006

Lodi, A.; Toma, M.; Guerrieri, R (2002). “Very low complexity prompted speaker verification system based on HMM-modeling Acoustics, Speech, and Signal Processing” Proceedings. (ICASSP '02). IEEE International Conference on Volume 4, 13-17 May 2002 Page(s):IV-3912 - IV-3915 vol.4

Lundin, E., Jonsson, E. (2000). “Anomaly-based intrusion detection: Privacy concerns and other problems” Computer Networks, v 34, n 4, Oct, 2000, p 623-640

McGraw, G. (2005). “Are cell phones the next target?” Network Magazine, v 20, n 6, 2005, p 82

Mell, P., McLarnon, M. (2006). “Mobile Agent Attack Resistant Distributed. Hierarchical Intrusion Detection Systems” www.raid-symposium.org/raid99/PAPERS/Mell.pdf accessed 1/2006

Mobitedia. (2006). “Cell phones – Features Specs and user Reviews” <http://www.mobitedia.com/phones/> accessed 7/2006

Papadaki, M., Furnell, S. (2004). “IDS or IPS: What is best?” Network Security, v 2004, n 7, July, 2004, p 15-19

Rawat, S., Gulati, V., Pujari, A. (2004). “Frequency- and ordering-based similarity measure for host-based intrusion detection” Information Management and Computer Security, v 12, n 5, 2004, p 411-421

Samfat, D., Molva, Refik. (1997). “IDAMN: An intrusion detection architecture for mobile networks” IEEE Journal on Selected Areas in Communications, v 15, n 7, Sep, 1997, p 1373-1380

Stamouli, I., Argyroudis, P., Tewari, H. (2006). “Real-time Intrusion Detection for Ad hoc Networks” <https://www.cs.tcd.ie/~htewari/papers/wowmom05.pdf> accessed 1/2006

Telecommunications magazine (2006). “Bluetooth still needs security bite”, July 1, 2004, http://telecomtest.bvdep.com/International/article.asp?HH_ID=AR_646 accessed 1/2006

Teresa, L. (2006). “Detecting Intruders In Computer Systems” ,Computer Science laboratory SRI International Menlo Park California <http://citeseer.ist.psu.edu/lunt93detecting.html> accessed 1/2006

The Register. (2006a). “Mobile Devices and Users Quocirca Insight Report” www.theregister.co.uk/2005/06/29/mobile_management_report.pdf accessed 1/2006

The Register (2006b). “Mobile web access on the up” http://www.theregister.co.uk/2006/04/19/mobile_ipsos/ accessed 8/2006

Westhoff, K., Paul, D. (2006). “Context Aware Detection of Selfish Nodes in DSR based Ad-hoc Networks” http://www.iponair.de/publications/Paul_Globecom02.pdf accessed 1/2006

Wiens, R. (2001). “Realistic Expectations for Intrusion Detection Systems” 19 March 2001, <http://www.securityfocus.com/print/infocus/1206> accessed 1/2006

Yap, T., Ewe, H. (2005). “A mobile phone malicious software detection model with behavior checker” Lecture Notes in Computer Science, v 3597, Web and Communication Technologies

and Internet-Related Social Issues - HSI 2005: 3rd International Conference on Human.Society@Internet. Proceedings, 2005, p 57-65

Yongguang Z., Wenke L. (2006). "Intrusion Detection in Wireless Ad-Hoc Networks" ACM MobiCom'2000 <http://citeseer.ist.psu.edu/zhang00intrusion.html> accessed 1/2006

Keystroke analysis as an authentication method for thumb-based keyboards on mobile handsets

S.Karatzouni and N.L.Clarke

Network Research Group, University of Plymouth, Plymouth, United Kingdom
e-mail:info@network-research-group.org

Abstract

The evolution of mobile networking has opened the door to a range of possibilities for mobile devices, increasing at the same time the sensitivity of the information stored and access through them. Current PIN-based authentication has proved to be an insufficient and an inconvenient approach. Biometrics that have proved to be a reliable approach to identity verification can provide a more robust mean of security as they rely on personal identifiers. Amongst various biometric techniques keystroke analysis combines features that can offer a cost effective, non-intrusive and continuous authentication solution for mobile devices. This research is being undertaken in order to investigate the performance of keystroke analysis on thumb-based keyboards that are being widely used in PDA's and Smartphone devices. The investigation was based on the scenario of authenticating users while typing text messages, using two keystroke characteristics, inter-key latency and hold-time. The results showed to be promising achieving an EER=12.2% with the inter-key latency, whereas unusually hold-time did not prove to be a feasible feature to utilise in such tactile environment.

Keywords

Keystroke analysis, Biometrics, Authentication, Mobile

1. Introduction

The proliferation of mobile devices and mobile networking has introduced new challenges for the protection of the subscribers' assets. The security risks are no longer associated only with safeguarding the subscribers account. With the introduction of 3rd generation mobile networks the services and information accessible through mobile handsets have increased in sensitivity, as micro-payments, mobile banking, location-based services are all reality for the mobile world and more potential is arriving in the future. But moreover the attraction that high-tech devices can result places a further concern for enhanced security, as underlined by looking at the statistics for mobile theft, which in the UK accounts the 45% of overall theft (British Transport Police, 2006).

Current authentication, mainly achieved by PIN's, is not enough to substantially safeguard today's mobile handsets and the data access enabled. As a secret knowledge technique it has several drawbacks, as it can be shared or written down, but also being a 4-6 digit number is not difficult for a potential impostor to acquire (Lemos, 2002). Furthermore as survey results show, subscribers consider it as an inconvenient method and as such they do not use them in the first place leaving their device unprotected (Clarke *et al*, 2002). Even a secondary measure, SIM cards due to

their functionality it is unlikely to be removed from the device, thus provide no protection in case a device is stolen or lost.

Alternative authentication based on biometrics could provide an enhancement on the security currently provided. Biometrics rely on the personal identifiers and therefore they can provide authentication based on something a person is, a fact that introduces a unique level of security that other approaches do not meet as it relates the process to a person and not to a possession of knowledge or token. A biometric method that can provide a cost-effective and a non-intrusive solution for mobile handset authentication is keystroke analysis, which is based on the typing dynamics of a user. The purpose of this research is to investigate keystroke analysis in thumb-based keyboards based on text messaging input, looking at the feasibility of applying this technique as an authentication method for mobile handsets that offer that tactile interface.

2. Keystroke analysis

Keystroke analysis is a behavioural biometric that attempts to verify identity based on the typing pattern of a user looking at certain characteristics of his interaction with a keyboard. A lot of research has been undertaken on the method since first introduced in 1980's, identifying two main characteristics to provide valuable discriminative information:

- Inter-key latency, which is the interval between two successive keystrokes, and
- Hold-time, which is the interval between the pressing and release of a key

The majority of the studies have looked at the feasibility of keystroke analysis on full QWERTY keyboards (Umpruss & Williams, 1985; Joyce & Gupta, 1990; Brown & Rogers, 1993; Obaidat & Sadoun, 1997), showing satisfactory results for both of the characteristics mentioned. In general inter-key latency has showed to provide better information for the classification in comparison to hold-time.

As in all biometrics the way to access the performance of keystroke analysis, two measures are used. The False Acceptance Rate (FAR) that indicates the percentage of an impostor falsely granted access to the system, and the False Rejection Rate (FRR), which represents the percentage of a legitimate user getting rejected. There is a trade-off between increasing security (and therefore decreasing the FAR) and increasing user convenience (and thus decreasing the FRR). As of the different security requirements for each system, the point that those two rates cross - the Equal Error Rate (%), is used as a more objective mean for the comparison of different biometrics.

For the assessment of keystroke analysis traditionally statistical approached were used, though more recently the use of neural network pattern recognition proved to provide better performance. A summary of the literature results underlying keystroke analysis on PC keyboards is provided in Table 1.

Study	Users	Input	Inter-Key	Hold-time	Approach	FAR	FRR
Umpress & Williams	17	Alphabetic	●		Statistical	11.7	5.8
Joyce & Gupta	23	Alphabetic	●		Statistical	0.3	16.4
Brown & Rogers	25	Alphabetic	●	●	Neural N.	0	12
Obaidat & Sadoun	15	Alphabetic	●	●	Statistical	0.7	1.9
					Neural N.	0	0
Ord & Furnell	14	Numerical	●		Neural N.	9.9	30

Table 1: Literature summary results on keystroke analysis on PC keyboards

Although the extensive research on keystroke analysis, it was not till recently that the method was assessed on interfaces provided on mobile phones where the tactile environment differs. A series of studies (Clarke & Furnell, 2006) accessed the method on regular mobile phone keypads with promising outcomes, achieving an EER= 8% based on numerical input. Nevertheless, the performance of keystroke analysis for thumb-based keyboards was undocumented. Thumb-based keyboards constitute an interesting gap in research as they provide the extensive interface of a PC keyboards and the thumb-based keystrokes of a mobile phone.

3. Methodology

This study looked into the feasibility of authenticating a user while typing text messages. Two different types of analysis were used in the context of this research- static and pseudo-dynamic accessing inter-key latency and hold-time respectively. A number of thirty messages, comprised the input of the experiment, which were designed to fulfil certain requirements.

Keyword	# Inter-key latencies	#Samples after outliers' removal	Training Set	Testing Set
everything	10	27	18	9
difficult	9	26	18	8
better	6	27	18	9
night	5	27	18	9
the	3	26	18	8
and	3	27	18	9

Table 2: Keywords used for inter-key latency

For the static analysis six varying sized keywords were included in the text messages providing a static component to use. The keywords were selected based on the criteria that it should be likely to appear often in a text message, while no abbreviations could be used as substitutes. Thirty repetitions of each keyword were included, a number of which though were removed as outliers. The words selected are listed in Table 2, along with the number of inter-key latencies that they involve and the number of samples used for training and testing.

The pseudo-dynamic analysis was based on the hold-time of the six most recurrent letters in the English language – e t a o n i, adequate number of repetitions of which were included. Literature has showed that attempts to perform dynamic analysis on keystroke dynamics (Leggett, Napier) did not yield satisfactory results. As such an attempt was made to utilize a static component – the recurrent letters, in a dynamic form of analysis.

Fifty participants were recruited to type the series of the text messages, using an XDA II handset that deploys a representative example of today’s thumb-based keyboard, as illustrated in Figure 1. In order to capture the keystroke data, appropriate software was implemented using Microsoft’s Visual Basic .NET, and deployed on the handset. A screenshot of the software is provided in Figure 2. As usual in keystroke analysis studies, corrections were not permitted in case the user misspelled a word as this would undesirably interfere with the data of the inter-key latency (Umpruss & Williams, 1985). Instead the whole word should be retyped in the correct form. The data collection was performed in a single session, although it would be preferred to collect the data during multiple sessions, as thus a more indicative typing profile of the users could be captured.



Figure 1: An XDA II's thumb-based keyboard



Figure 2: Screenshot from experiment software

4. Results

4.1 Inter-key latency

An initial analysis of the input data showed a fairly large spread of values on the inter-key latencies, even for the smaller keywords that were expected to be more concise because of the commodity and length. Additionally to that the difference of the values attributed to each user was not large, so that many of the users overlapped. This puts a burden on the classification algorithm, as those two factors make the definition of limits to differentiate between users very difficult as the values are interfering. Figure 3, illustrates the mean and standard deviation for the larger keyword across all users as an example of the problem.

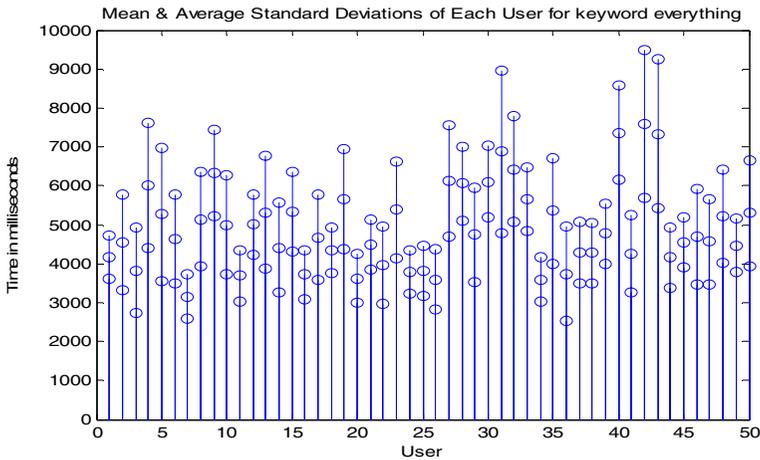


Figure 3: Mean & Standard deviation for keyword “everything”

A number of tests took place, using Feed Forward Multilayer Perceptron neural network as it has showed very good performance in previous research (Clarke & Furnell, 2006). Different configurations were tested, changing the network size and weights but also the training time, looking for optimum performance. The best results were outcome of the keyword ‘everything’ as expected because of providing a larger input vector, giving an EER=23.4% with FAR=19.3 and FRR=27.5, the last of two are indicated in Figure 4.

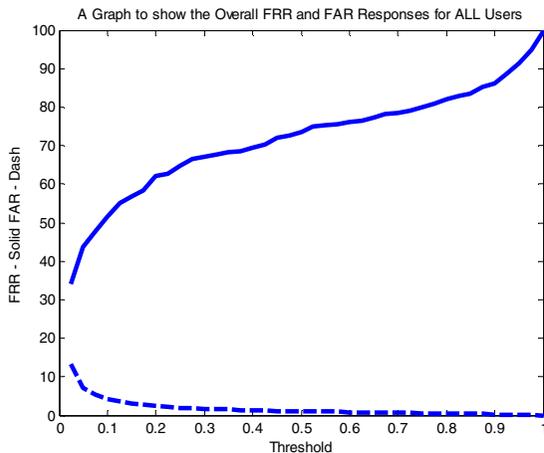


Figure 4: Overall FAR and FRR for best case network for keyword “everything”

As can be seen, but also for all of the tests, the results showed an FRR much higher from the FRR which can be explained by the large amount of 49 impostors extensively training the network versus the one authorised user. Furthermore the number of samples assigned to the testing of the classification was small, resulting to the FRR encountering large steps in its transitions.

The error rate is fairly high, nevertheless, there were cases of users reaching an EER below 10% with best case user 1 achieving an EER of 0.3% that shows a good ability of classification. The rest of the keywords resulted even higher error rates, as it was though expected as they provide a smaller input vector. The best results for each keyword are listed in Table 3.

Keyword	FAR (%)	FRR (%)	EER (%)
everything	12.8	34.2	23.5
difficult	13.2	43.0	28.1
better	18.0	43.1	30.5
night	21.3	45.8	33.5
the	23.7	41.5	32.6
and	24.3	43.6	33.9

Table 3: Best results for each keyword

The results of different networks showed minimal change in the EER's, though the FAR and FRR showed much variation. This indicates the fact that the network tries to optimise for the population of users, averaging the performance and as such each user can not train to the best suited way.

To overcome that problem a different approach was utilised based on the improved results it gave on previous study (Clarke & Furnell, 2006). A gradual training was performed, training the network for an extensive amount of time but periodically evaluating the performance. The results showed a noticeable decrease on the error rates with best case achieving an EER=12.2% for the larger keyword. The summary of the gradual training results are listed in Table 4.

Keyword	FAR (%)	FRR (%)	EER (%)
everything	15.8	9.1	12.2
difficult	16.8	12.0	14.4
better	23.5	14.4	18.9
night	24.2	14.4	19.3
the	29.3	19.5	24.4
and	28.7	17.6	23.1

Table 4: Gradual training results for all keywords

For the keyword “everything”, 20 users achieved an FRR=0% with the respective FAR below 10%, which provides a very promising result, with the best user achieving an FAR=0.7% and FRR=0%. The list of best and worst case users for all keywords are listed in Table 4. The results underline the requirement of different training intensiveness for each user, but mainly that inter-key latency offers the discriminative data to classify users in the specific tactile interface.

Keyword	Best Case				Worst Case			
	User	FAR	FRR	EER	User	FAR	FRR	EER
everything	2	0.7	0	0.4	6	42.6	22.2	32.4
difficult	11	2.6	0	1.3	46	18.1	50.0	34.1
better	49	3.2	0	1.6	27	35.1	33.3	34.2
night	34	4.5	0	2.3	25	25.6	55.5	40.5
the	26	12.8	0	6.4	39	41.6	50	45.8
and	11	10.9	0	5.4	5	32.2	66.7	49.4

Table 5: Best & Worst Case results from gradual training

As due to time limitations the network was not optimised it is believed that further testing will be able to provide even lower results.

4.2 Hold-time

In the contrary to inter-key latency, hold-time did not seem to be able to provide any data to help classify different users. A series of tests on different network configurations using all six letters (as to provide the larger possible input vector) resulted in an EER of around 50%, showing that little classification could be performed. The same error rate derived using different size subsets of the letters with smaller input vectors but with more repetitions of each letter, but also when a larger input of eight letters was used adding in the set also the letters ‘r’ and ‘s’, as next on the reoccurrence list.

In order to further access the performance of hold-time, a group of only 20 users was used aiming to help the classification as the population to discriminate against would be less, though with no change in the results. Even when gradual training was tested, using the six letters set, no improvement came. Sample results from various tests are provided in Table 6. Although there were users with FRR or FAR of 0% the respective FAR or FRR was reaching over 80%. Even though there was a 10% decline on the EER using gradual training, the results are still very high to suggest that hold-time can offer valuable discriminative information.

Set	Training	Users	FAR	FRR	EER
6 letters	normal	20	49.5	49.4	49.5
6 letters	normal	50	31.3	69.0	50.2
8 letters	normal	50	26.7	72.9	49.8
3 letters	normal	50	22.1	77.6	49.9
6 letters	gradual	50	34.2	36.8	36.8

Table 6: Sample results from various tests on hold-time

5. Discussion

As the results showed inter-key latency can provide a mean of differentiating between users, when based on a latency vector of 10, being able to achieve a 12.2%

EER with the gradual training approach. Using a smaller input vector, although classification was able to be performed there were increased error rates, though it must be noticed that no network optimization was researched for the smaller keywords.

In regards to the inter-key latency, the results did not have the low rates that research on regular keyboards has showed, though there are a number of factors that differentiate this study. An issue to underline is that the keyboard used provides a more restricted keystroke interface as the distance between the keys is smaller in comparison with a PC, but also the number of fingers likely to be used is two in contrast with ten in the respective case. Both of these factors limit the typing dynamics as the combinations of the fingers in conjunction with the timing of the keystrokes and movement to achieve them, are restricted. This results in a smaller value area for the keystrokes of the users, making the distinction between them more difficult. Furthermore, although the layout was familiar to all users as it shares the same layout with a PC keyboard, some of the participants experienced difficulty in identifying the placement of the keys due to the different way of typing.

Hold-time did not provide any proof that it can be utilised in the specific typing interface though there are a number of factors that may explain the inability of the keystroke feature.

Firstly the keys that the thumb-based keyboard deploys are very small related to the chunky tactile environment that a normal keyboard offers, restricting the interval length between the pressing and release of a key and thus not providing much differentiation in values. Although hold-time has performed well on regular mobile phone keypads (Clarke & Furnell, 2006), where still the keys were larger than the keyboard used in this experiment, a further factor was that, in a mobile keypad in order to access the preferred letter more than one pressings are often required, with the hold-time being calculated from the first keystroke till the last key release, increasing immediately the range of values and thus allowing an easier distinction between them.

Furthermore in a thumb-based keyboard, fingers stay almost static due to the limited area, thus keystrokes hardly differentiate, as no other factors such as hand movement appears as in PC keyboards which may affect the pressing of a key. What must be also noticed is that some participants complained about the feedback from the keyboard, as they could not at all cases be sure if they had pressed a key, which might led to a continual change of the hold-time.

6. Conclusion

This research was a feasibility study on the utilisation of keystroke analysis as an authentication method in devices that offer the tactile environment of a thumb-based keyboard. The results showed that from the two traditionally used keystroke characteristics- inter-key latency showed promising results, whereas hold-time gave no clues of a potential use in that kind of keystroke interface, though research must be undertaken to further access them.

Future work will search upon an optimised network configuration that was not extensively research during this study, in regard to the inter-key latency. Furthermore the use of different keywords will be investigated as also the combined use of more than one, looking also to use abbreviations as keywords as they are more likely to appear in a text message more often. In respect to hold-time, further tests are required before concluding to its ineffectiveness, exploring the use of longer input vectors and different letter subsets. A future experiment will also look to utilise thumb-based keyboards that offer a slight different tactile environment than the one used in this study, to have a mean of comparison, of the performance of the keystroke characteristics and an insight on the factors that may affect it.

Nevertheless, the study showed promising results for the use of keystroke analysis in thumb-based keyboards. Although the accuracy of the method does not compete in distinctiveness with other biometrics such as fingerprints, the nature of keystroke analysis can provide a monitoring authentication mechanism, transparent to the user that is not feasible for other techniques. In that basis it can provide continuous authentication based on the regular use of the device, and if used in conjunction with other authentication approaches that can fulfil the lack of the method in accuracy, a more enhance security can be achieved.

7. References

British Transport Police (2006): “Mobile phone theft”, <http://www.btp.police.uk/issues/mobile.htm>

Brown, M., Rogers, J. (1993): “User Identification via Keystroke Characteristics of Typed Names using Neural Networks”, *International Journal of Man-Machine Studies*, vol. 39, pp. 999-1014

Clarke, NL., Furnell, SM.,(2006) : “Authenticating Mobile Phone Users Using Keystroke Analysis”, *International Journal of Information Security*, ISSN:1615-5262, pp1-14,2006

Clarke, N., Furnell, S.M, Rodwell, P.M, Reynolds, P.L (2002): “Acceptance of subscriber authentication method for mobile telephony devices”, *Computers & Security*, vol. 21, no.3, pp220-228

Joyce R., Gupta, G. (1990): “Identity Authentication Based on Keystroke Latencies”, *Communications of the ACM*, vol. 39; pp 168-176.

Lemos, R. (2002): “Passwords: The Weakest Link? Hackers can crack most in less than a minute”, *CNET.com*, <http://news.com.com/2009-1001-916719.html>

Obaidat, M. S., Sadoun, B. (1997): “Verification of Computer User Using Keystroke Dynamics”, *IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics*, Vol. 27, No.2.

Ord, T. (2000): “User Authentication for Keypad-Based Devices using Keystroke Analysis”, *MSc Thesis*, University of Plymouth, UK.

Umphress, D., Williams, G. (1985): “Identity Verification through Keyboard Characteristics”, *International Journal of Man-Machine Studies*, Vol. 23, pp. 263-273

Strengthening the Human Firewall

G.C.Tjhai and S.M.Furnell

Network Research Group, University of Plymouth, Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

Employees' complacency, ignorance and unawareness of security are amongst the biggest obstacles to maintaining IT security within an organisation. Indeed, technical security controls alone are not enough to provide a real protection if there is no human participation acquired in this stage. As a consequence, it is desirable for many organizations develop an effective security awareness programme; by which user awareness could be enhanced. The objectives of this paper are to explore the extent of security awareness problems and to ideally specify and develop methods by which security culture could be cultivated (through training and awareness initiatives)

Keywords

Security awareness, User behaviour, Security culture, Training, Education

1. Introduction

Corporate security breaches are no longer new issues, and are now a reality to be faced by many modern organisations. With the numerous security incidents being reported in recent times, the need to secure and protect corporate sensitive information and networks is of greater importance than ever before.

The human element is always thought to be the key as well as the weakest link of security chain. Firmly focusing on human factor in security practices is the first step to achieve a successful corporate security culture. Unfortunately, the human-related security issue is not a straightforward problem to deal with. There is a need to make all employees and end users aware of the need for security, and to educate or train them to do their part in securing the enterprise. An organisational security awareness programme is aimed to make all the employees understand and appreciate not only the value of the company's information assets, but also the consequences if the assets are compromised. A good security programme should be able to change the employees' behaviours or activities into more secure habits, by convincing them why being more security conscious is important.

This paper attempts to benchmark the level of security awareness and culture within two organizations, enabling views to be formed on the effectiveness of the methods in use.

2. Background

As discussed previously, end-user security behaviour is one of the underlying issues of information security practices. Identifying significant factors of corporate security culture would be of value since they could have strong influences on user security behaviours. Besides, it is worth remembering that improper security habits are the major determinant of the level of security incidents experienced, and a good security programme is required to improve user security behaviours across the organisation (Leach, 2003).

Since user security behaviour is the main factor that could determine the level of security incidents, analysing or investigating staff behaviours are considerably crucial to perceive the state of security awareness within an organisation. Having a further investigation on end user security behaviours in a systematic view point of different kind of security behaviours are deemed to have an important role in influencing and enhancing the effectiveness of information system security (Stanton et al. 2005).

Security awareness is another important security component that needs to contribute into security culture. The neglect of information security practices is a result of having no full security culture and policy implementation within the organisation. Therefore, in order to ensure the proper organisational behaviours, information security obedience is the best solution. Information security obedience highlights the combination of corporate governance, culture and information security (Thomson & von Solms 2005), and in order to achieve it effective security awareness training and programmes are of prime importance (especially focusing upon specific groupings of employees within the organisations, such as top management, IT personal and end users).

So, in order to enable a more understanding about the level of security awareness and the extent of security problem faced by organisations, the results of an IT security awareness survey is presented below.

3. IT security awareness survey

The survey was mounted online for around 20 days, during the end of July until the middle of August 2006. Since the survey was targeted at corporate employees only, it had finally been promoted to two organisations within two different industry sectors; namely telecommunication and local government. A total of 134 responses (24 from telecommunication and 110 from local government) were received during the survey period, providing a suitable basis for the subsequent analysis. The sections that follow outline the areas covered by the questionnaire, and the associated results.

3.1 Respondents backgrounds

In this section, twelve questions were presented and basically aimed to gather information about respondents' backgrounds and to assess their basic understanding of current security issues.

3.1.1 Telecommunication sector

From this respondent group, the survey findings resulted in an unequal split between male and female; with approximately twice as many male respondents. From an age perspective, there was a significant focus within the 25-34 category, suggesting that most opinions came from respondents who would have grown up with information technology. It is also worth noticing that more than three quarters of respondents were high educated people, with half of them holding IT-related qualifications.

In respect of employees' security threat awareness, nearly all respondents were threat-aware in general, especially in relation to more common security issue such as viruses and spam, as illustrated in Figure 1. However, the survey result revealed that half of the respondents holding IT-related qualifications were not aware of the existence of social engineering attacks.

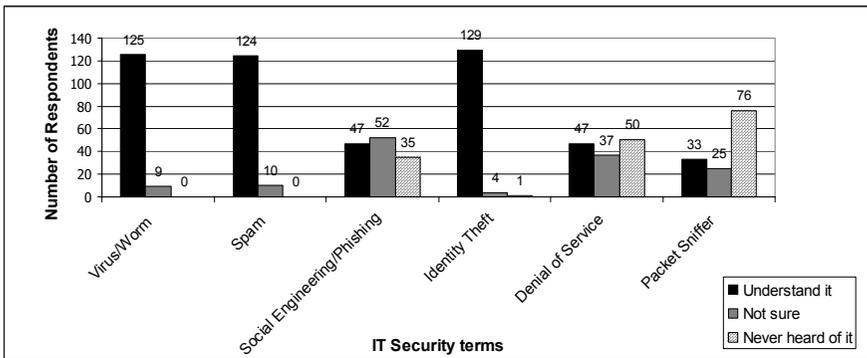


Figure 1: Level of understanding of IT security terms (from both respondent groups)

With regard to respondents' awareness of security initiatives, only one eighth of participants claimed to have security awareness programme in their organisation, while 66% showed that they were not made alert of any security programme in their organisations. Moreover, quite a significant number (one fifth of respondents) had declared to have no any security programme or related training being implemented by their organisations. In terms of their frequency in joining security awareness programme, it is surprisingly enough that from three respondents claiming to have a security awareness programme, none of them had attended a regular security programme, as shown in Figure 2. Two of them never joined any security programme implemented, while another one had had a security education once in an induction day. As such, this result gives a clear illustration that even where a security programme has been implemented in an organisation, it may not be widely adopted by those that may need it.

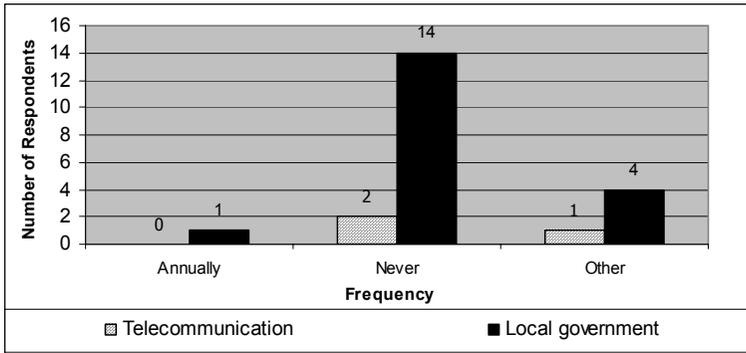


Figure 2: Respondents joining security awareness programme/training

3.1.2 Local government

As to the respondents' educational background, the majority were holding a middle level of education. Similar to what had been found in telecommunication sector, almost all respondents (more than 92%) were reported to have a good understanding and awareness of viruses, spam and identity theft. On the other hand, social engineering was still not a common term for the majority of employees, as illustrated in Figure 1. Only 30% of them had a good perception of social engineering attack.

From 110 respondents received from this respondent group, only 17% (19 people) had declared to have security programme or related training in their organisation. Conversely, a significant proportion (57%) reported having no idea of any security programme in their organisation, while 26% definitively claimed not to one. Worse still, from the 19 people who claimed to have a security programme, only one of them had participated in it (annually). A total of 18 people never attended any regular security programme or only received security education once in the induction day or via security messages, as shown in Figure 2.

3.2 Respondents security attitudes and decision making

In order to focus more upon users' understanding and perception of security issues, the survey presented 6 questions aiming to investigate users' security attitudes and their decision making in their current employment.

3.2.1 Telecommunication sector

Questions were designed to evaluate the effectiveness of password policy enforcement within the organisation. From this sample group, only approximately 8% of respondents (2 out of 24 people) had firmly adhered to the password policy by applying all strong password characteristics in their password selection. In terms of their frequency in changing passwords, almost 42% of respondents hardly ever changed them unless the system prompted them to do so.

The survey results also point out the most common security mistakes made by respondents in this sample group. The highest proportions of respondents, accounting for 42%, were inclined to leave their computers unattended without logging off in an open office. In term of respondents' understanding of corporate system security, the

findings show that almost 88% of respondents (21 persons) believed that personal data is not a good choice of information for password selection. However, 18 out of 21 people who declared to disagree with the usage of private data in password selection were still using that information as their password choice. Surprisingly, two thirds of respondents considered that the IT department has the sole responsibility for securing corporate IT systems. The role of IT administrator in controlling and managing corporate IT system has been misinterpreted by the majority of employees

3.2.2 Local government

From this local government sector, only 3% of respondents had firmly conformed to the password policy by applying all strong password characteristics in their password choice. Through this questionnaire, it is clear that several characteristics of strong password have been widely applied by many corporate employees, for example the usage of 8 characters, combination of letters, symbols and numbers and so on. Significantly, almost 38% of staff did not use private data as their password choice. With respect to their password changing habits, more than half of employees in this respondent group never changed their password unless the system had forced them to do so.

Nearly 28% of respondents admitted to sharing their password with other colleagues and to writing it down as a way to remember. Interestingly, leaving the computer unattended without logging off is again the most frequent mistake made by corporate employees, with 46% (51 people) being likely to do this.

A large number of respondents (84%) agreed that personal data is a bad choice for password selection. Despite this, however, 56 out of 93 persons used the personal data as their password. The usage of private data as password is undoubtedly more usable and practical for the users, but this choice could definitely render the system vulnerable to the attack since personal data is easier for hackers to guess. Significantly, the majority of respondents from this respondent group again believed that IT administrator is the only one who is responsible for securing the corporate IT system.

3.3 The influential factors of employees' security behaviours

After focusing on the employees' perception of organisational security awareness and their security habits, the next aspect of the questionnaire was directed to evaluating the factors that could possibly affect the way employees behave in their organisations.

3.3.1 Telecommunication sector

The survey finding has shown a clear point that the vast majority of respondents (92%) installing antivirus software were influenced by their awareness of security threats. Approximately 70% and 50% claimed to be affected by other two factors, namely policy enforcement and usability respectively. Significantly, less than 9% of participants asserted that installing security software could be strongly influenced by the environment surrounding them.

As to good security practice in taking back-up of data, around 83% of respondents in the telecommunication company claimed that having an experience of losing data could serve as a good motivation to take a back-up of their data. Very surprisingly, only 15% of them declared that the environment could effectively influence them to follow this security practice.

3.3.2 Local government

Interestingly, similarly to the findings from telecommunication sector, 77% of respondents who installed the security software were influenced by their security awareness. Quite a significant number of employees (67% and 47% of participants) installed antivirus software as a response to policy enforcement and usability factors. Interestingly enough, only a small number of them claimed that installing anti virus software could be deeply influenced by the environment or people surrounding them. For example, few thought that looking at other colleagues firmly following security practice (e.g. installing security software to protect their work PC) could strongly motivate them to do the same thing.

With regard to the second security practice, 71% of respondents were more likely to take back-up of their data because of their experiences in losing data and their awareness of the importance of having back-up data. Significantly enough, 62% of them asserted that policy enforcement was one of the strongest influential factor motivating them to adhere to this security practice, whilst two fifths of employees were affected by the usability issue. Only 15% of employees took back up of data because of their environment, such as the motivation from other colleagues.

3.4 Preferable security learning methods

The final question was focused on the evaluation of several familiar security learning methods. In this study, the respondents were asked to rate their preferences for each of the security learning techniques available.

3.4.1 Telecommunication sector

In general, the most popular learning methods reported among the employees in this respondent group was via presentation or face to-face training, with 83% of respondents rating it as an excellent learning method. Significantly, web-based awareness courses and inspection/audits were also well-liked among employees in this group; around 42% of employees viewing these methods as beneficial. Very interestingly, poster/screen savers, trinket/gifts and regular bulletins were rated as the least popular or helpful learning methods. Through this finding, it is clear that more educational methods (such as presentation sessions) are much more favourable than methods that are only intended to remind the users about security issues.

3.4.2 Local government

The findings here were slightly different to what had been found in the telecommunication company. Significantly, four learning methods were most welcomed by this respondent group; namely presentation or face-to-face training, web-based awareness courses, regular bulletins, and inspection or audits. Roughly 85% of respondents said that a presentation or web-based course was the best method used to enhance security awareness or deliver security materials.

Significantly, web-based awareness course and regular bulletins were also considered to be other beneficial methods used in security awareness programme, accounted for 60% respectively. Unlike the result found in telecommunication sector, regular bulletin is deemed to be a favourable method.

4. Discussion

Since the objective of conducting the IT security awareness survey was to investigate the security culture within the organisations, it is clear that the findings have revealed some interesting facts about the level of employee awareness within those respondent groups.

With regard to the employees' understanding about IT security terms, the majority of employees knew the common threats such as virus and spam. However, the findings show that the lack of understanding about more uncommon security threats, such as social engineering, could become one of the major challenges faced by organisations. Social engineering, which is also linked to the threat of phishing, is one of the most malicious attacks targeted on human element instead of technology. The lack of understanding from employees renders corporate system vulnerable to the attack. Due to this issue, the level of security awareness should be enhanced for people at every level in the organisation, regardless of their status. The employees ought to be educated well about corporate security issues, the potential risks, as well as their responsibility or participation in protecting company's assets.

From the survey results, it is also clear that the level of training given to the employees is variable within the organisations. Significantly, the survey finding has suggested that the level of security education or training given to the employees is deemed to be considerably low within both respondent groups. The low level of actual take-up suggests that the programmes have merely been thought to be an add-on activity, and that the value of security awareness has not been widely communicated to people in the organisation. For that reason, the lack of understanding about the importance of security awareness undoubtedly becomes a major obstacle in developing or building the ideal security culture.

Another significant point that could be drawn from the survey findings is that security is often traded-off against usability by employees. For example, even though more than three-quarters of respondents disagreed with the usage of personal data as password choice, 65% of the total group still used it in their password selection. This negative correlation is likely to have occurred as a result of the security and usability trade-off. Although, users believe that personal data is very sensitive information and easily guessed, they will yet apply this information as their password selection since it is easy to remember. This finding reveals the similar result as what had been discussed in prior research (Besnard & Arief 2003). In terms of their password-changing habits, the findings have shown that there is a lack of tendency or attentiveness from quite large number employees to change their password frequently. Commonly, instead of being seen as a security measure, a password is deemed to be a mere tool used to gain access to a system. As a result, the employees'

awareness of password security should be improved to some extent as an attempt to maximise the level of organisational system security.

Importantly, the survey findings also provide a clear picture that user security behaviours and attitudes are potentially influenced by environmental factors, usability issues, enforcement and self-persuasion (Leach 2003; Thomson & von Solms 1998). Through this result, it is clear that employees could adopt good security practices if they have the understanding of what behaviours are expected of them; for example what they are being told (enforcement), what they see being practiced by others around them (environment/social learning) and the experiences they had from the decision they made in the past (experience of failure).

Another critical component that needs to be seriously considered during the development of effective security programme is the technique used to deliver or promote the security culture. A good security programme should be able to attract the audiences to actively respond to the security initiative/material promoted. Here, several examples of effective security programme techniques are web-based awareness courses, presentation or face-to-face training, inspection and audit, handbooks, reference materials, and so on.

5. Conclusions

Overall, the survey has revealed some interesting facts about security awareness in organisations. Even if awareness programmes exist, there is still lack of understanding from the employees about the importance of corporate security awareness. Due to this issue, raising security awareness is becoming one critical practice for all companies; which then need to be tailored to minimise user-related faults and maximise the efficiency of security practices and procedures from the users' perception. Since only two industry sectors had been evaluated in this research, the study has not provided an enough basis for an extensive investigation of security culture across wide range of organisations.

After evaluating and developing an ideal security measure to increase the level of security awareness throughout the organisations, which is more focused on the measures at the strategic level, the future work now should be shifted to investigate the method used to measure the effectiveness of security programme. This is supporting by the fact that there is a lot less available in the literature on how to measure the effectiveness of security programme rather than how to deliver it. Since the security awareness programme is a dynamic process, it needs to be continually measured and managed to keep pace with changes in the organisation's risk profile.

6. References

Besnard, D. and Arief, B. (2003), 'Computer security impaired by legitimate users', *Computers and Security* (23). available online: <http://www.sciencedirect.com>, date visited: 6 August 2006.

Leach, J. (2003), 'Improving User Security Behaviour', *Computers and Security* 22(8). Available online: <http://www.citeulike.org/article/56424>, date visited: 21 January 2006.

Stanton, J. M., Stam, K. R., Mastrangelo, P. and Jolton, J. (2005), 'Analysis of end user security behaviours', *Computers and Security* 24(2). available online: http://www.elsevier.com/wps/find/journaldescription.cws_home/405877/description#description, date visited: 20 January 2006.

Thomson, M. and von Solms, R. (1998), 'Information Security Awareness: educating your users effectively', *Information Management and Computer Security* 6(4). available online: <http://www.emeraldinsight.com/Insight/ViewContentServlet?Filename=Published/EmeraldFullTextArticle/Articles/0460060404.html>, date visited: 25 January 2006.

Thomson, K.-L. and von Solms, R. (2005), 'Information security obedience: a definition', *Computers and Security* 24(1). available online: http://www.informatik.uni-trier.de/_ley/db/journals/compsec/compsec24.html, date visited: 22 January 2006.

Section 3

Computing, Robotics & Interactive Intelligent Systems

Evolution of musical lexicons by singing robots

E.Drouet and E.R.Miranda

Interdisciplinary Centre for Computer Music Research (ICCMR)
University of Plymouth, Plymouth U.K.
e-mail: eduardo.miranda@plymouth.ac.uk

Abstract

This paper introduces a model where a community of interactive autonomous singing robots programmed with appropriate motor, auditory and cognitive skills can evolve a shared lexicon of sonic intonation patterns from scratch, after a period of spontaneous creation, adjustment and memory reinforcement. Musical expectation is defined as a sensory-motor mechanism whereby the robots evolve vectors of motor control parameters to produce imitations of heard intonation patterns.

1. Introduction

We agree with composer Al Biles who stated that “everybody knows what music is, but that is not to say that everybody agrees on what music is. Rather, everybody has a personal conception of what music is, and that conception informs how they process the sounds they experience in their lives”. For the purposes of our research, we follow Al Biles’ definition of music as *temporally organized sound*. “This rather inclusive definition certainly covers the typical music we hear on the radio, which isn’t a bad definition of sounds that have acquired some societal consensus as music, but it also includes bird songs, babbling brooks, even the ambient sounds of daily life” (Biles, to appear). In this sense, it is fair to say that music is not uniquely human. A number of other animals also seem to have music (i.e., at least to our ears, even though animals use it primarily for communication purposes). Complex vocalizations can be found in many birds (Marler and Slabbekoon, 2004), as well as in a mammals such as whales (Payne and McVay, 1971) and bats (Behr and von Helversen, 2004). Recently Timothy Holy and Zhongsheng Guo (2005) demonstrated that the ultrasonic vocalizations that male mice produce when they encounter female mice or their pheromones have the characteristics of song. What is intriguing is that primates who are close related to humans are not as “musical” as those mammals that are far more distantly related to us. This intriguing fact suggests that music might have evolved independently among various types of animals, at various degrees of sophistication. In this context, it would be perfectly plausible to suggest the notion that robots might also be able to evolve music.

In order to build systems for the emergence of music one needs to establish the factors that may shape the course of musical evolution, such as physiological and cognitive factors, including models of interaction.

The physiological factors comprise the sensors and actuators of interacting individuals. These involve models of the hearing system, body, limbs, and so on. It may also involve models of the sensory-motor cortex and associated neural

mechanisms involved in sensory-motor tasks. The expertise for building these models is beyond the scope of this paper; it includes fields such as Biomechanics (Zinkovsky, Sholuha and Ivanov, 1996) and Biophysics (Glaser, 2001).

As for the cognitive factors, the brain certainly uses different mental modules to process music and these modules have varying degrees of independence from each other. Lawrence Parsons (2003) has conducted a number of brain-imaging experiments, which indicate that the neural system for processing music is widely distributed throughout our brain. This finding is confirmed by studies of patients with brain lesions (Peretz *et al.*, 1994). Isabelle Peretz and Max Coltheart (2003) proposed a functional architecture of the brain for music processing that captures the typical properties of such distributed modular organization. Basically, they have identified two main processing modules: one concerned with processing pitch and the other with rhythm. Both modules process incoming musical signals backed by a *musical lexicon*; a kind of memory bank of short musical sequences. Surely, this basic architecture can be refined into smaller components, depending on the level of detail at which one wishes to study its functionality. This is likely to become increasingly complex as research in the emerging field of Neuroscience of Music progresses. What is important, however, is that this modularity of the brain for music processing suggests a plausible methodology for building robotic models of musical evolution.

Why is it important to study the emergence of music with robotic simulations? A better understanding of basic mechanisms of musical evolution is of great importance for musicians looking for hitherto unexplored ways to create new music works with computers. Broadly speaking, current techniques for implementing generative music systems can be classified as *abstract algorithmic* or *music knowledge-based*. Abstract algorithmic techniques are suitable for generating music from the behaviour of algorithms that were not necessarily designed for music in the first instance, but embody pattern generation features that are suitable for producing musical materials. Such algorithms include Cellular Automata (Hunt *et al.*, 1991; Miranda, 1993) and Particle Swarms (Blackwell and Bentley, 2002) to cite but two examples. Music knowledge-based techniques generate music using algorithms derived from or inspired by well-established music theory. Most of these systems can learn compositional procedures from given examples, adopting either a symbolic approach (Steedman, 1984; Cope, 1996; Papadopoulos and Wiggins, 1998) or a connectionist (neural networks) approach (Todd and Loy, 1991; Mozer, 1994), depending on the way they store information about music. Hybrid systems also exist (Burton and Vladimirova, 1997).

Both classes of techniques have their merits and pitfalls. Abstract algorithmic techniques tend to produce rather complex music, most of which may sound too remote from what the majority of people, including expert listeners, would consider musical. This is possibly so because abstract algorithmic music tends to lack the cultural references that people normally rely upon when listening to music. Conversely, knowledge-based techniques tend to produce pastiches of existing musical pieces, which often are of little interest for composers aiming to create new music; that is, music that is not based on mimicking existing pieces or well-known musical styles. In our robotic simulations we aim to bring the merits of both

approaches closer to each other by offering the possibility of evolving new musical systems based upon the same principles that might have shaped existing musical styles. Inspired by John Casti's (1997) use of the term "would-be worlds", Artificial Life's goal of looking at "life as it could be," we refer to these emerging new musical systems as "would-be music", or "music as it could be".

In this paper we introduce a model where a community of interactive robots programmed with appropriate motor, auditory and cognitive skills can develop a shared lexicon of sonic intonation patterns from scratch, after a period of spontaneous creation, adjustment and memory reinforcement. The robots develop vectors of motor control parameters to produce imitations of heard intonation patterns. The robots thus expect to hear pitch sequences that correspond to their evolved motor vectors.

Intonation is generally defined as the melody of speech; it is characterised by the variations in the pitch of a speaker's voice. The rationale for attempting to model the development of intonation patterns comes from the fact that intonation is fundamental for the development of vocal communication and music. There have been a number of research reports giving evidence that babies are born with an acute sensitivity to intonation (Locke, 1993; Nazzi *et al.*, 1998). This ability probably evolved due to the need for enhanced mother-infant interactions. Baby talk, or infant-directed-speech, sounds like music due its exaggerated intonation, which helps babies and very young children to develop their linguistic ability. Mothers use baby talk to influence the behaviour and elicit emotions in pre-linguistic infants. Following this idea, the robots are programmed with a fundamental "instinct": to imitate what they hear.

In our model, imitation involves the task of hearing an intonation pattern and activating the motor system to reproduce it. The robots must form a common lexicon: a robot must develop a repertoire that is similar to the repertoire of its peers. Metaphorically speaking we would say that the intonations create some form of "social identity" for the robots, which can be assessed in terms of the similarity of their lexicons.

The importance of imitation for evolution has gained much attention after the discovery of mirror neurons in the frontal lobes of macaque monkeys. Mirror neurons are neurons which fire both when an animal performs an action and when the animal observes the same action performed by another animal, especially of the same species. Thus, the neurons mirror the behaviour of another animal, as though the observers were themselves performing the action. These neurons have subsequently been observed in some birds, and in other primates including humans (Rizzolatti and Craighero, 2004).

2. The Model

2.1 The Architecture

The robots (Figure 1) are equipped with a voice synthesiser, a hearing apparatus and a memory device.



Figure 1: The model uses DRK8000 robots, manufactured by Dr.Robot®, which were adapted by the authors for high-quality voice synthesis and analysis (with sampling rate at 8,000 Hz).

The voice synthesiser is essentially implemented as a physical model of the human vocal mechanism (Boersma, 1993; Miranda, 2002). The robots need to compute three vectors of parameters in order to produce intonations: lung pressure, the width of the glottis, and the length and tension of the vocal chords, represented here as *lung_pressure(n)*, *interarytenoid(n)* and *cricothyroid(n)*, respectively. As for the hearing apparatus, it employs short-term autocorrelation-based analysis to extract the pitch contour of a vocal sound (Miranda, 2001). The algorithm features a parameter that defines the sensitivity of the auditory perception of the robots. In essence, this parameter regulates the resolution of the hearing apparatus by controlling the precision of the short-term autocorrelation analysis.

Essentially, a robot's memory stores its lexicon of intonations, but it also stores other information such as probabilities, thresholds and reinforcement parameters. (These variables will be clarified when the algorithms are introduced below.) They have two distinct modules to store intonations in their memories: a motor map and a perceptual map. The motor map stores information in terms of three vectors of motor (vocal) parameters and the perceptual map stores information in terms of pitch contour, which is represented as a graph whose vertices stand for initial (or relative) pitch points and pitch movements, and the edges represent a directional path. Whilst the first vertex must have one outbound edge, the last one must have only one incoming edge. All vertices in between must have one incoming and one outbound edge each. Vertices can be of two types, initial pitch points (referred to as *p-ini*) and pitch movements (referred to as *p-mov*) as follows (Figure 2):

$$p\text{-ini} = \{SM, SL, SH\}$$

$$p\text{-mov} = \{VLSU, LSU, MSU, SSU, RSB, SSD, MSD, LSD, VLSD\}$$

where:

SM = start intonation in the middle register

SL = start intonation in the lower register

SH = start intonation in the higher register

and

VLSU = very large step up

LSU = large step up

MSU = medium step up

SSU = small step up

RSB = remain at the same band

SSD = small step down

MSD = medium step down

LSD = large step down

VLSD = very large step down

An intonation will invariably start with a $p\text{-ini}$, followed by one or more $p\text{-movs}$. It is assumed that an intonation can start at three different voice registers: low (SL), middle (SM) and high (SH). Then, from this initial point $\{t(n), n=0\}$ the next pitch at $t(n+1)$ might jump or step up or down, and so forth.

It is important to note that pitch frequency values or labels for musical notes are not relevant here because this scheme is intended to represent abstract melodic contours rather than a sequence of pitches (or musical notes) drawn from a specific tuning system. This is very important here because one should not assume that the robots must sing in any established musical scale, but should be given the ability to create their own scales.

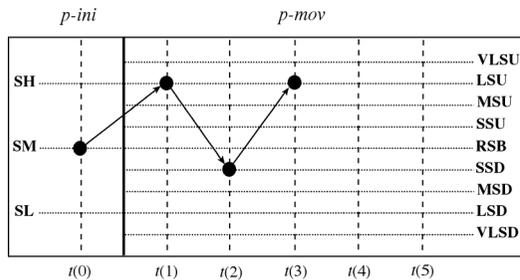


Figure 2: The representation of an intonation, where $t(n)$ indicates an ordered sequence of n pitches.

2.2 The Interactions

The interaction algorithms were largely inspired by the work of Luc Steels (1997) on evolutionary language games. All robots have identical synthesis and listening

apparatus. At each round, each of the robots in a pair plays one of two different roles: the *robot-player* and the *robot-imitator*. The main algorithms are given as follows:

Algorithm 1: Robot-player produces an intonation

1. motor_control[α] \leftarrow pick-any-motor-control in Motor-Repertoire(robot-player)
2. synthesise-sound(motor_control[α])

Algorithm 2: Robot-imitator produces an imitation

3. pitch_vector[β] \leftarrow perceive-intonation
4. intonation[β] \leftarrow perceptual representation(pitch_vector[β])
5. intonation[Δ] \leftarrow search-similar(intonation[β]) in **Perceptual-Repertoire**(robot-imitator)
6. motor_control[Δ] \leftarrow retrieve_motor_control (motor-control[intonation[Δ])
7. synthesise-sound(motor_control[Δ])

Algorithm 3: Robot-player hears the imitation and gives a feedback

8. pitch_vector[ψ] \leftarrow perceive-imitation
9. imitation[ψ] \leftarrow perceptual-representation(picth_vector[ψ])
10. intonation[ϕ] \leftarrow search-similar(imitation[ψ]) in **Perceptual-Repertoire**(robot-imitator)
11. intonation[α] = perceptual-representation(motor_control[α])
12. IF intonation[α] = intonation[ϕ]
13. THEN { feedback \leftarrow *positive*
14. reinforce(motor_control[α]) in **Motor-Repertoire**(robot-player)
15. reinforce(intonation[α]) in **Perceptual-Repertoire**(robot-player) }
16. ELSE { feedback \leftarrow *negative* }
17. output-signal(feedback)

Algorithm 4: Robot-imitator reacts to robot-player's feedback

18. IF feedback = *positive*
19. THEN { approximate(intonation[Δ] \rightarrow intonation[β]) in **Perceptual-Repertoire**(robot-imitator)
20. reconfigure_motor_control(intonation[Δ]) in **Motor-Repertoire**(robot-imitator)
21. reinforce intonation[Δ] in **Perceptual-Repertoire**(robot-imitator)
22. reinforce motor_control(Δ) in **Motor-Repertoire**(robot-imitator) }
23. ELSE IF feedback = *negative*
24. THEN IF success-history(intonation[Δ]) > success-threshold

```

25.   THEN { motor_control[λ] ← produce-new-motor-control
26.     intonation[λ] ← perceptual
representation(motor_control[λ])
27.     save-new(intonation[λ]) to Motor-Repertoire(robot-
imitator)
28.     save-new(motor_control[λ]) to Perceptual-
Repertoire(robot-imitator) }
29.   ELSE { distantiate(intonation[Δ] ↔ intonation[β]) in
Perceptual-Repertoire(robot-imitator)
30.     reconfigure_motor_control(intonation[Δ]) in Motor-
Repertoire(robot-imitator) }

```

Algorithm 5: End of interaction updates

```

31. interaction-updates(robot-player)
32. interaction-updates(robot-imitator)

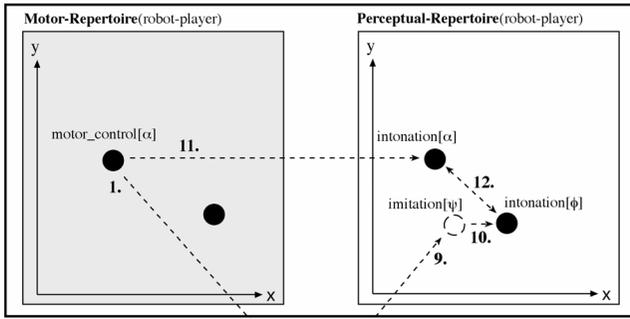
```

Glimpses at the functioning of these algorithms are given in Figures 3, 4 and 5. For didactic purposes, the co-ordinates of these figures do not correspond to the actual parameters of the model. For the sake of clarity, the plotting is in an idealized two-dimensional representation of the motor and perceptual repertoires. The numbers in the figures indicate actions corresponding to the line numbers of the algorithms.

The robot-player starts the interaction by producing an intonation α , randomly chosen from its repertoire. The robot-imitator then analyses the intonation α , searches for a similar intonation Δ in its repertoire and produces it. Figure 3 shows an example where the robot-player and the robot-imitator hold in their memories two intonations each. The robot-player picks the intonation α from its motor-repertoire and produces it (1). The robot-imitator hears the intonation α and builds a perceptual representation β of it (4). Then it picks from its own perceptual repertoire the intonation Δ that is most perceptually similar to the heard intonation β (5) and produces it as an imitation (6). Next, the robot-player hears the imitation Δ and builds a perceptual representation ψ of it (9). Then it picks from its own perceptual repertoire the intonation ϕ that is most perceptually similar to the imitation ψ (10).

If the robot-player finds another intonation ϕ that is closer to Δ than α is, then the imitation is seen as unsatisfactory, otherwise it is satisfactory. In Figure 3, the robot-player babbles the original intonation α to itself (11) and concludes that α and ϕ are different (12). Then, it sends a negative feedback to the robot-imitator (17). When an imitation is unsatisfactory the robot-imitator has to choose between two potential courses of action. If it finds out that Δ is a weak intonation in its memory (because it has not received enough reinforcement in the past) then it will move it away from α slightly, as a measure to avoid repeating this mistake again.

Robot-player



Robot-imitator

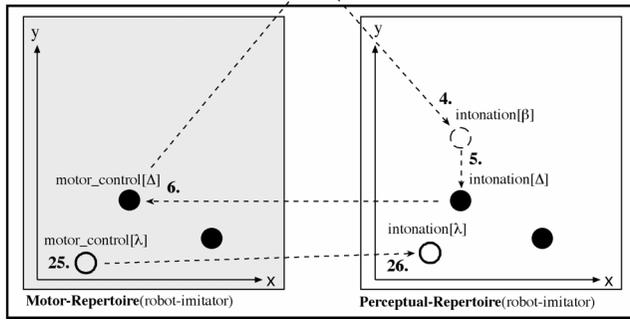


Figure 3: Example of an unsuccessful imitation.

But if Δ is a strong intonation (due to a good past success rate), then the robot will leave Δ untouched (because it has been successfully used in previous imitations and a few other robots in the community also probably consider this intonation as being strong) and will create a new intonation λ similar to Δ to include it in its repertoire; that is, the robot produces a number of random intonations and then it picks the one that is perceptually most similar to Δ . Let us assume that in Figure 3 the intonation Δ has a good past success rate. In this case, the robot-imitator leaves it untouched and creates a new intonation λ to include in its repertoire (25, 26).

Robot-imitator

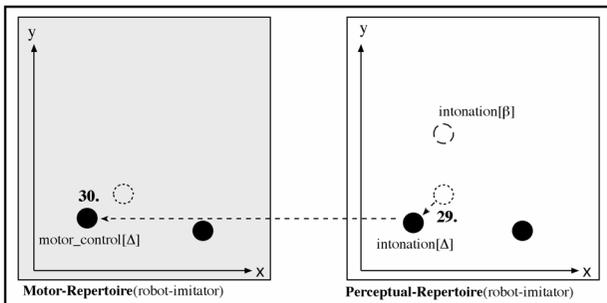


Figure 4: An example where the unsuccessful imitation involved an intonation that has a poor past success rate.

Figure 4 shows what would have happened if the intonation Δ did not have a good past success rate: in this case the robot-imitator would have moved Δ away from β slightly (29 and 30). Finally, Figure 5 shows what would have happened if the robot-player had concluded that α and ϕ were the same, meaning that the imitation was successful. In this case, the robot-imitator would have reinforced the existence of the intonation Δ in its memory and would have moved it slightly towards the representation of the heard intonation β .

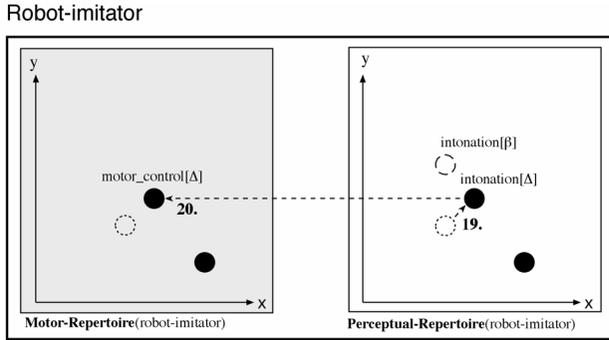
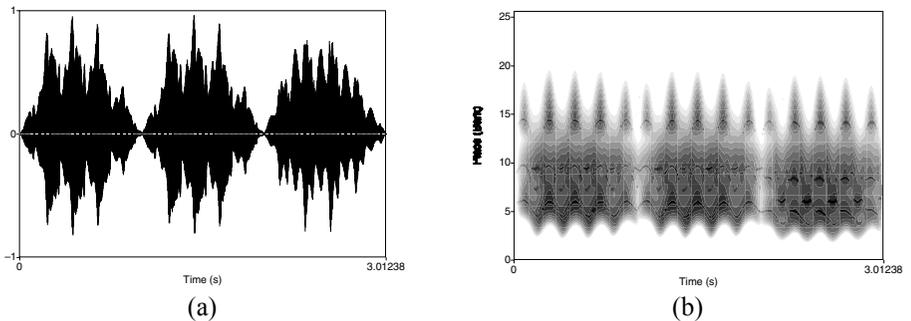


Figure 5: An example of a successful imitation.

Before terminating the round, both robots perform final updates. Firstly they scan their repertoire and merge those intonations that are considered to be perceptibly close to each other. Also, at the end of each round, both robots have a certain probability P_b of undertaking a spring-cleaning to get rid of weak intonations; those intonations that have not been sufficiently reinforced are forgotten. Finally, at the end of each round, the robot-imitator has a certain probability P_a of adding a new randomly created intonation to its repertoire; we refer to this coefficient as the “creativity coefficient”.

3. A Typical Simulation Example

Please note that although we can run the model in software simulation mode, the examples discussed below are not from software simulations, but from interactions with real sounds. Figure 6 plots an example of an intonation with three elements in the sequence and its respective cochleogram and pitch analysis.



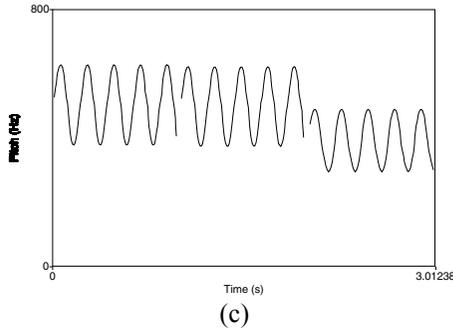


Figure 6: An example of an intonation produced by a robot. (a) Plotting of the intonation. (b) The spectrogram of the intonation. (c) The pitch analysis of the intonation.

The oscillations of the line representing pitch in Figure 6 are due to the vibrato nature of the singing voice.

The graph in Figure 7 shows a typical example of the evolution of the average repertoire of a group of five robots, with snapshots taken after every 100 interactions over a total of 5000 interactions. After a drastic increase of the repertoire at about 800 interactions, the robots settled to an average of seven intonations each until about 2200 interactions, when another slight increase took place. Then they settled to an average of nine intonations until about 3800 interactions. From 3800 interactions onwards the robots steadily increased their repertoires.

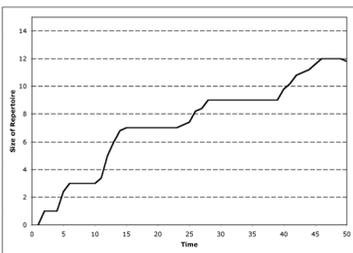


Figure 7: The evolution of the average size of the repertoire of intonations of the whole group of robots. In this case the group developed an average repertoire of 12 intonations. (The time axis is in terms number of interactions multiplied by 100.)

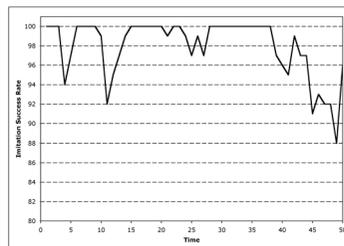


Figure 8: The imitation success rate over time. (The time axis is in terms of number of interactions multiplied by 100.)

The pressure to increase the repertoire is mostly due to the probability P_a of creating a new random intonation, combined with the rate of new inclusions due to unsatisfactory imitations. The size of the repertoire tends to stabilise with time because the more the robots use strongly settled intonations, the more these

intonations are reinforced in their repertoires, and therefore the more difficult for new intonations to settle in.

The graph in Figure 8 plots the imitation success rate of the community, measured at every 100 interactions. Note the decrease of imitation success rate during those phases when the robots were increasing the size of their repertoires. Although the repertoire size tends to increase with time, the success rate tends to stay consistently high. However, this is highly dependent upon the number of robots in the group: the higher the number of robots, the deeper the fall of the success rate and the longer it takes to re-gain the 100% success rate stability, if ever achieved. (Note: we have not run simulations with a large number of robots because we do not have the necessary resources, but we have run realistic software simulations with a large number of “agents”.)

Figure 9(a) portrays the perceptual memory of a robot randomly selected from the group after 5000 interactions. In this case, the length of the intonations varied from three to six pitches. (The minimum and maximum length of the intonation to be evolved is fixed beforehand.) This robot evolved eleven intonations; one below the average. Figure 9(b) shows only those intonations that are three pitches long.

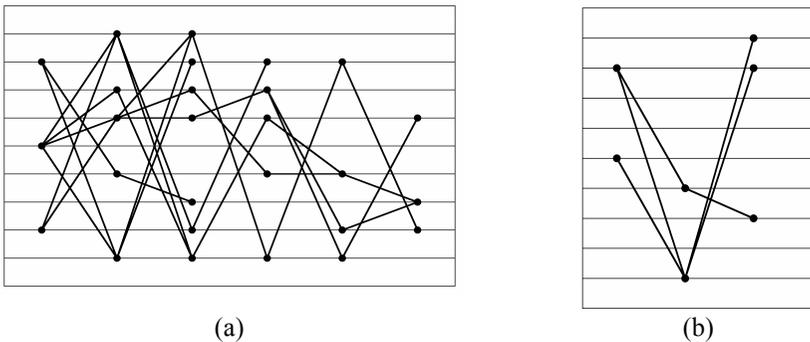


Figure 9: (a) The perceptual memory of one robot. (b) Only those intonations that are three pitches long. For the sake of clarity, the background metrics and labels of the graphs are not shown; see Figure 2.

An interesting feature of this model is that the robots do not necessarily have to evolve the same motor representations for what is considered to be perceptibly identical. Figure 9 shows the motor functions evolved by three different robots to represent what is essentially the same intonation.

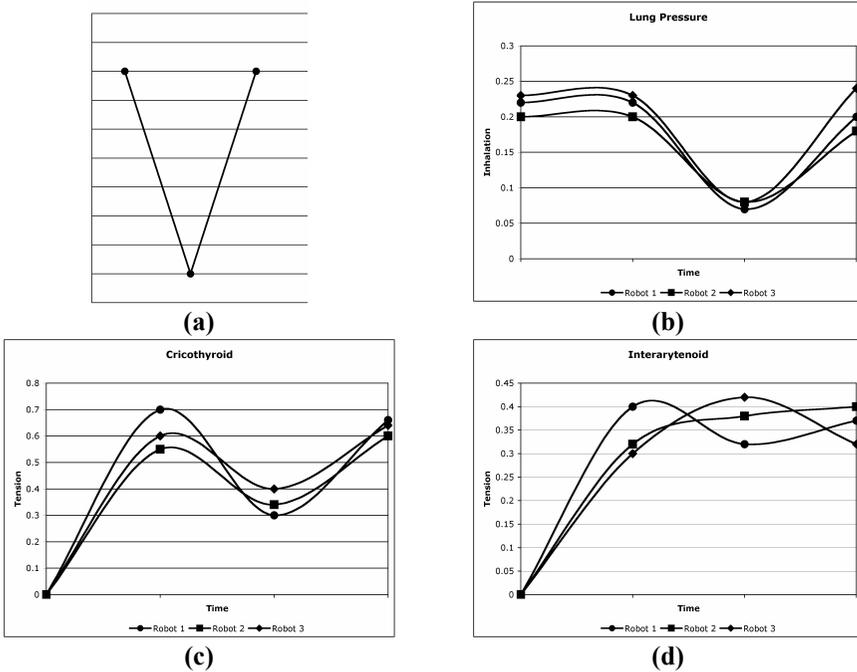


Figure 9: (a) One of the perceptual patterns from Figure 8(b) and its corresponding motor control vectors developed by three different robots: (b) the *lung_pressure* vector, (c) the *interarytenoid* vector and (d) the *cricothyroid* vector.

The imitation of an intonation pattern requires the activation of the right motor parameters in order to reproduce it. The robot-imitators assume that they always can recognise everything they hear because in order to produce an imitation a robot will use the motor vectors that best match its perception of the sound in question. It is the robot-player who will assess the imitation. Metaphorically speaking, we could consider that musical expectation is a “social convention” grounded on the nature of their sensory-motor apparatus.

4. Conclusions

In the introduction we suggested that it would be perfectly plausible to suggest that robots might be able to evolve music. Then we demonstrated how this might be done. However, we acknowledge that the model introduced in this paper cannot evolve proper music yet, but rather the rudiments of what one might refer to as *proto-music*.

One interesting hypothesis that is emerging from the research of a number of scholars is that there might have been a single precursor for both music and language: a communication system that had the characteristics that are now shared by music and language, but split into two systems at some date in our evolutionary history.

The model presented in this paper so far deals with short pitch sequences. The natural progression with this work is to devise a way to deal with longer pitch sequences and eventually with proper musical compositions.

Although the symbolic sensory-motor-like memory mechanism proposed for storing intonations served well the objectives of the present model, it is not efficient for storing longer pitch sequences, let alone fully fledged pieces of music. In order to increase the complexity of the model, it is necessary to improve the memory mechanism, which would probably be more efficient by storing information about generating the sequences rather than the sequences themselves.

Nevertheless, the model presented here is encouraging in the sense that it provides strong indications that music may indeed emerge from the overall behaviour of interacting autonomous robots.

Finally, it is obvious that this work is not about biological evolution in the strict Darwinian sense. Rather, we are interested in cultural evolution, which cannot be studied as a biological system. Nevertheless, our model does feature a selective mechanism. Moreover, some form of “mutation” does take place (intonations move closer to or away from other intonations in memory) and patterns are “born” (random additions) and “die” (spring-cleaning).

5. References

- Behr, O., von Helversen, O. (2004). “Bat serenades – Complex courtship songs of the sac-winged bat *Saccopteryx bilineata*”, *Behavioral Ecology and Sociobiology* 56:106-115.
- Biles, J. A. (to appear). “Evolutionary Computation for Musical Tasks”. In E. R Miranda and J. A. Biles (Eds.), *Evolutionary Computer Music*. Heidelberg: Springer
- Blackwell, T. M. and Bentley, P. J. (2002) Improvised Music with Swarms. In Proc of the 2002 Congr. On Evolutionary Computation, pp. 1462-1467.
- Boersma, P. (1993), “Articulatory synthesizers for the simulations of consonants”, *Proceedings of Eurospeech '93*, Berlin, Germany, pp. 1907-1910.
- Brown, S. (2000). “The “Musilanguage” Model of Music Evolution”, In N. Wallin, B. Merker and S. Brown (Eds.), *The origins of music*. Cambridge, USA: The MIT Press.
- Burton, A R & T Vladimirova (1997) A Genetic Algorithm Utilising Neural Network Fitness Evaluation for Musical Composition, In G.D. Smith, N.C. Steele, & R.F. Albrecht (Eds.), *Proceedings of the 1997 International Conference on Artificial Neural Networks and Genetic Algorithms* (pp. 220-224). Vienna: Springer-Verlag.
- Casti, J. L. (1997). *Would-be Worlds: How Simulation of Changing the Frontiers of Science*. New York: John Wiley & Sons.
- Cope, D (1996) Experiments in Musical Intelligence. Madison, WI: A-R Editions Inc.
- Glaser, R. (2001). *Biophysics*. Heidelberg: Springer.
- Holy, T. E. and Guo, Z. (2005). “Ultrasonic Songs of Male Mice”, *PLoS Biology* 3(12): e386.

Hunt, A, R Orton & R Kirk (1991) Musical Applications for a Cellular Automata Music Workstation. *Proceedings International Computer Music Conference (ICMC 91)*, Montreal, Canada, pp. 165-168.

Locke, J. L., 1993. *The Child's Path to Spoken Language*. Cambridge, MA: Harvard University Press.

Marler, P and Slabbekoorn, H. (Eds.) (2004). *Nature's music: The science of birdsong*. Boston, MA: Elsevier.

Miranda, E. R. (2002), *Computer Sound Design: Synthesis Techniques and Programming*. Oxford, UK: Focal Press.

Miranda, E. R. (2001), "Synthesising Prosody with Variable Resolution". *AES Convention Paper 5332*. New York, USA: Audio Engineering Society, Inc.

Miranda, E R (1993) Cellular Automata Music: An Interdisciplinary Project. *Interface (now Journal of New Music Research)* 22(1):3-21.

Mozer, M (1994) Neural network music composition by prediction: Exploring the benefits of psychophysical constraints and multiscale processing. *Connection Science* 6:247-280.

Nazzi, T., Floccia, C. and Bertoincini, J., "Discrimination of pitch contours by neonates" *Infant Behaviour*, No. 12, pp. 543-554, 1998.

Papadopoulos, G and Wiggins G., (1998) A Genetic Algorithm for the Generation of Jazz Melodies, Proceedings of 8th Finnish Conference on Artificial Intelligence, Jyväskylä, Finland.

Payne, R. S. and McVay, S. (1971). "Songs of humpback whales", *Science* 173:585-597.

Parsons, L. M. (2003). "Exploring the Functional Neuroanatomy of Music Performance, Perception, and Comprehension", In. I. Peretz and R. Zatorre (Eds.), *The Cognitive Neuroscience of Music*, pp. 247-268. Oxford, UK: Oxford University Press.

Peretz, I. and Coltheart, M. (2003). "Modularity of music processing", *Nature Neuroscience* 6:688-691.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, L., Hublet, C. and Demeurisse, G. (1994). "Functional dissociations following bilateral lesions of auditory cortex", *Brain* 117:1283-1301.

Rizzolatti, G. and Craighero, L. (2004). "The mirror-neuron system", *Annual Review of Neuroscience*, 27:169-192.

Steedman, M (1984) A Generative Grammar for Jazz Chord Sequences. *Music Perception* 2:52-77.

Steels, L. (1997). "The Origins of Syntax in Visually Grounded Robotic Agents", *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI '97)*, Nagoya, Aichi, Japan.

Todd, P.M., and Loy, D.G. (Eds.) (1991). *Music and connectionism*. Cambridge, MA: MIT Press.

Wray, A. (1998). "Protolanguage as a holistic system for social interaction", *Language & Communication* 18:46-667.

Zinkovsky, A. V., Sholuha, V. A and Ivanov, A. A. (1996). *Mathematical Modelling and Computing Simulation of Biomechanical Systems*. Singapore: World Scientific.

eGovernment take-up in the city of Plymouth, UK

G.Ford^{1,2} and A.D.Phippen¹

¹ Plymouth City Council, Plymouth, United Kingdom

² Network Research Group, University of Plymouth, United Kingdom
e-mail: info@network-research-group.org

Abstract

eGovernment presents new challenges to both central Government and local authorities in terms of service delivery and citizen engagement. In examining the socio-economic composition of Plymouth, we provide a baseline from which to measure take-up of local eGovernment in a specific region. Survey data collected through Plymouth City Council's citizen forum demonstrates good take up and good perception of the local authority site from its citizens. However, it does also raise some wide ranging issues regarding the methods of eGovernment engagement and how authorities can best go about getting best value from their web based service delivery channels to achieve the predicts savings eGovernment will bring.

Keywords

eGovernment, Engagement, Exclusion

1. Introduction

eGovernment is the electronic delivery of central and local government information / services by means of information and communication technology. In 1999 the UK government declared, "The information age should increase the choice of how citizens and businesses receive services, not restrict it ... We will develop targeted strategies to ensure that all groups have proper access to information age government" (Cabinet Office, 1999).

The UK strategy fits within a European Union (EU) framework, most recently the "Information Strategy i2010" which builds on its predecessor, launched in 2002, the "eEurope 2005 Action Plan". The EU strategies and action plans ensure a common framework of eGovernment services across Europe. Indeed in March 2001 the council of the EU identified 12 key services for its citizens, each of the eGovernment services are intended to be "standard" to all EU states.

Given that UK strategy works within an EU framework, in 2002 the National Audit Office (NAO) produced a report, "Better Public Services Through eGovernment" which included identifying five key benefits of eGovernment, i.e.

- Greater choice – to provide users with a greater range of services and delivery channels
- Better accessibility – giving citizens greater access to the range of services
- More convenience – providing services in a way which suits citizens and businesses, e.g. 24 hours a day 7 days a week

- Faster delivery – providing faster more accurate service
- Improved efficiency – replacing manual processing of routine high volume work with IT system

(NAO, 2002)

However, the NAO report also identified 6 key risks regarding the “take-up” of eGovernment services, i.e.

- Familiarity: the Internet which has yet to become a normal established part of everyday life
- Expectation: low expectations about IT and what it can deliver
- Ease of use: unless new services are easy to use there is a risk take up will be low
- Benefits: the benefits for the public must be clear or take up will be low
- Social exclusion: citizens will not take up services if they do not have access to a computer
- Cost: if the cost of accessing services on-line is expensive people will not want to use it

In order to achieve the goal of e-enablement, considerable investment has been made in central & local government eGovernment projects, between 2001/02 and 2005/06 £7.4bn has been spent (£3bn – local government, £4.4bn - central government) (Rogers, 2003).

In order to prioritise and standardise the development of local government services, Implementing Electronic Government (IEG) statements were developed, i.e. corporate plans for the goal of 100% “e-enablement” of particular local government services / information. The IEG statements logically fit within the EU framework, e.g. “local planning applications”.

The IEG statements have developed over time, i.e. IEG1 in 2001, ... IEG6 in 2006, the end of the programme evaluation is planned for April 2006.

The result of the IEG programme has been to furnish a suite of services and information common to all local authorities, the software or the method of delivery conceivably being different, however, the objective of delivering a service / information has been achieved.

However, “The Oxford survey” (Dutton at al., 2005) reported that only 24% of UK Internet users and 15% of the population have ever used any eGovernment service, whereas, 70% of Internet users rate the Internet as either “important” or “very important” to their current way of life. The variance between usage of eGovernment services and perceived “importance” of the Internet raises concerns with respect to “take-up” and perceived value.

In this paper we investigate eGovernment take-up in the city of Plymouth, in the South West of the UK to determine whether social exclusion is being addresses

through eGovernment delivery. Initial some core demographic information is presented as a baseline for investigation. Following this information, the analysis is provided via survey data collected for Plymouth City Council (PCC) regarding Internet and eGovernment usage.

2. The City of Plymouth

Plymouth is located in the south west of England and is the largest city west of Bristol. The population of the city is 241,000 and increasing; traditionally Plymouth has had strong links with the military, in particular the Royal Navy. Since the reduction in the navy's surface and submarine fleet and the privatisation of the naval dockyard, Plymouth has sought to diversify its economy, e.g. electronics, medical / healthcare, advanced engineering and call centres

A brief review of the 2001 Census (Census, 2001) provides in insight into the socio-economic composition of the electoral wards in Plymouth. Although Plymouth is the 14th largest city in the UK its retail ranking is only 29th, with a potential total high street expenditure of £1,094 million only £600 million is spent in the city centre. The city fails to attract relatively affluent shoppers in the outlying areas. (PCC-Experian, 2006)

The Neighbourhood Renewal Unit (NRU), which is part of the “Department for Communities and Local Government”, is responsible for overseeing the Government's neighbourhood renewal strategy. In 2004 the “Indices of Deprivation 2004” (ID2004) was published (revised June 2004), the ID2004 contains 7 key indices (income, employment, health deprivation and disability, education skills and training, barriers to housing and services, crime, living environment).

The indices are measured at local authority level and sub-areas - “Super Output Area” (SOA). The 32,482 low level SOA's each contain the seven indices of deprivation, thus allowing the SOA's to be ranked according to how deprived they are relative to each other. The base information has been collated to form the “Index of Multiple Deprivation 2004 (IMD2004)”. The “Indices of Deprivation 2004” has therefore identified,

- Plymouth is ranked 76th out of England's 354 local authority districts for its average deprivation and the extent of that deprivation across the city.
 - Overall, of the 160 Super Output Areas (SOA's) in Plymouth, 19 are amongst the top 10% most deprived in England.
 - The top 10% SOA's (16 in total) with the worst index scores are largely clustered in the south western corner of Plymouth, the wards of “St. Peter and The Waterfront”, “Devonport” and “Efford”.
 - The ‘least deprived’ ward of the city lies within “Plymstock Dunstone”.
- (PCC, 2006)

From the data presented above, we can conclude that Plymouth is a somewhat deprived city, compared to other cities of similar size, with higher rates of low skilled workers, lower levels of affluence, etc. than average. Therefore, we can

assume, based upon government literature (ODPM 2005) that a city such as Plymouth has a lot to gain from eGovernment. The following section considers both Internet usage and eGovernment engagement within the city, to test whether such claims can be demonstrated.

The remainder of this paper examines data collected from a survey carried out by PCC regarding Internet and eGovernment usage. It considers findings against the above discussion, and considers the state of eGovernment usage in the region compared to the national picture.

3. Usage Survey

In November 2002 PCC established a residents panel comprising of approximately 1,600 residents randomly selected to sit on a panel to complete surveys on various aspects of services supplied by PCC or affecting citizens of the city. Surveys are typically circulated 3 times a year, the scope of the surveys range from City Strategy to National Health Service (NHS) service provision.

In the June 2006 survey, a number of questions were posed to the panel regarding their Internet usage and use of the PCC website. The questionnaire resulted in 933 responses, indicating a response rate of 58%. Although the gender is nearly 50:50 upon further analysis age profile would appear skewed, whereby there are no female respondents aged 35-44 and no male respondents over 54:

Age	Respondents: Gender					
	Male		Female		Did not answer	
	Number	%	Number	%	Number	%
18 - 24	0	0%	27	3%	0	0%
25 - 34	85	9%	40	4%	0	0%
35 - 44	204	22%	0	0%	0	0%
45 - 54	114	12%	74	8%	0	0%
55 - 64	0	0%	169	18%	0	0%
65 - 74	0	0%	126	14%	0	0%
Over 75	0	0%	53	6%	0	0%
Did not answer	5	1%	5	1%	31	3%
Total	408	44%	494	53%	31	3%

Table 1 – Respondent demographic

The Office of National Statistics (ONS) estimate 13.9 million households (57%) in Great Britain could access the Internet from home (ONS, 2006). However, the percentage of respondents with Internet access at home was 71%. Similarly, the ONS estimates sixty nine per cent of households with Internet access have broadband connection (ONS, 2006). The proportion of respondents with broadband is 80%, however, a number still utilise “dial-up”.

The review identified a number of significant differences regarding the regularity of Internet usage, i.e., 67% of respondents said they used the Internet daily and 22% said they used it weekly.

Respondents who think it unlikely they will access the Internet cited a number of reasons, i.e.

- Not interested / prefer not to 35%
- Do not understand computers 25%
- Do not need to 20%
- Other 13%
- Too expensive 7%

When asked to detail “other”, responses varied, for example “Do not have a computer or time”, “Prefer face-to-face contact”, “No time”, “Dangerous place”.

The group that did not use the Internet is of particular interest as this challenges popularly perceived views about Internet non-usage. While it is generally felt the it is the excluded who can not obtain access to the Internet (ODPM 2005) the indications from this survey are that non-use is more by means of personal choice than factors of exclusion. Only 32% of non-users presented exclusion factors (price and confidence) as reasons not to use it.

Following consideration of Internet connectedness, the questionnaire continued to examine the number of respondents who have accessed PCC’s website in the previous 12 months has increased markedly, from 22% in 2004 (the last time the panel were surveyed about Internet usage) to 40% in 2006. Both of these figures are significantly higher

When asked about the frequency of visiting the website 3% of respondents replied at least daily, 10% at least weekly, there were minimal variances with respect to gender or age.

	Frequency of accessing PCC’s website			
	At least daily	At least weekly	At least monthly	Less often
All respondents	3%	9%	28%	60%

Table 2 – Respondent Access to Plymouth City Council Website

A series of questions were then asked to assess usability of the website, to each question the users response would be on a scale (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree).

The user assessment of the website was favourable, i.e.

- The website looks fresh and modern: only 9% “disagreed” or “strongly disagreed”
- The website is too cluttered: 31% “agreed” or “strongly agreed”
- It is difficult to navigate around the website: 28% “agreed” or “strongly agreed”
- Information on the website is current and up to date: 10% “disagreed” or “strongly disagreed”

- There is too much information on the website: 12% “agreed” or “strongly agreed”
- The website provides relevant information: 8% “disagreed” or “strongly disagreed”
- Overall, the website is good to use: 9% “disagreed” or “strongly disagreed”

The questionnaire asked “Have you used any of Plymouth City Council’s online services?”, i.e.

Have you used any of Plymouth City Council’s on-line services		
Yes	No	Can not remember
34%	63%	3%

Table 3 – Respondent Use of Plymouth City Council Online Services

The users were asked which on-line forms they had used, the intention being to identify the most widely used forms, i.e.

PCC On-line forms	Used	Not used
Refuse issues form	39%	61%
Bulky waste	32%	68%
Council tax	26%	74%
Have your say	18%	82%
Job application	14%	86%
School admissions	9%	91%
Street lighting fault	9%	91%
Benefits calculator	6%	94%
Other	32%	68%

Table 4 – Respondent Use of Online Services by Type

4. Further Discussion

From the data presented from our survey, we can conclude that the level of engagement regarding eGovernment is certainly ahead of those figures from the Oxford Study (Dutton at al., 2005). In addition, those that do engage with eGovernment services from Plymouth CC have a generally good opinion about the site and its services. The only criticisms levelled at this site are similar to those determined by other work examining citizen engagement in the UK (Lacohee et. al. 2006) – too much information that is difficult to locate. One might consider these to be problems imposed on the local authority from the IEG strategy, rather than problems with the specific local authority site.

The data does provide a couple of interesting statistics from the engagement viewpoint. Firstly, the frequency of usage is fairly low – the majority of people will use the site less than once a month. This reflects the general engagement with council services – the authority is not viewed as something with which an individual wishes to become familiar and regularly engage. However, the authority, and its website, do provide a valuable service which people will return to when they have an

information or service requirement. If we look at the types of services that are the most popular – they are all fairly mundane, but necessary aspects of everyday life. A local authority site is not something to which people will visit to find out community news, current affairs, etc. However, once they are aware that services exist, they will return to the site.

This leads us onto our second observation from the data – the people that do engage with the site tend to have a high satisfaction rating with the site. Therefore, if people find out about the site, they will tend to engage with it. So, the issue for the local authority is how to get people engaged. Work complementary to that presented here (PSF 2006) has shown that central government efforts to use mass media engagement have been largely ineffective, so questions remain regarding how to engage citizens with local eGovernment services.

Additionally, we wished to consider the issues of social exclusion and the potential of eGovernment to address this. We have certainly demonstrated that in a region without high levels of affluence and with high levels of low skilled workers take up of eGovernment is high compared with other statistics and well regarded. However, when considering the classic premises of social exclusion (lack of intellectual ability, low income, etc.) we cannot reflect this in our data. One of the most interesting findings from the respondents is that it seems that people are not using the Internet through choice, rather than barriers. This is a particularly interesting finding, and does challenge a lot of thinking in addressing issues of exclusion and perceived digital divides. It also has implications for the development of eGovernment – other service delivery channels must remain open, as it is unlikely that take up of eGovernment will reach 100%. Even with availability and intellectual ability, there will always be a minority that chose not to use the Internet because they prefer other approaches.

5. Conclusions

At the start of this paper we stated that we wished to investigate the take up of local eGovernment within a specific region, considering the socio-economic aspects of this region to determine the implications for eGovernment take up specific to that area. We have considered the rationale behind eGovernment, and the concerned voice by parts of the UK Government at the start of the eGovernment implementation process. Through data collected by Plymouth City Council we have demonstrated that take up is improving and the engagement and perception are both good, and that the original NAO concerns are not borne out in this region. However, we have also discovered a some of important aspects that merit further investigation:

1. Usage tends to be specific and low volume. This is not to say the site isn't popular, as we have data to demonstrate that it certainly is. However, local authorities should not view their sites in the same way that, for example, a commercial organisation might consider theirs (i.e. an advertising/marketing channel).
2. How does a local authority best approach engaging the disengaged aspects of their region?

3. Internet non-usage is perhaps not as simple as sometimes believed in social exclusion literature.

What this does demonstrate is the immaturity of grass roots eGovernment as a subject for study and evaluation, and there is still a considerable amount of work to do if take up will reach levels when predicted cost savings in service delivery can be achieved.

6. References

Cabinet Office (1999) *'Modernising Government'*, Cabinet Office, Cmd 4310 London

Census (2001), *'Census 2001'*, <http://www.statistics.gov.uk/census/>

Dutton, W.H., di Gennaro and A.M. Hargrave (2005) *'The Internet in Britain'*. Parliamentary Affairs Online. <http://pa.oxfordjournals.org/cgi/reprint/gsl003v1.pdf>

IDABC (2005) *'Overview of the eGovernment situation and progress in EU Member States'*. <http://ec.europa.eu/idabc/en/chapter/203> (Nov, 2005)

Lacohee, H, Phippen, A., Crane, S. (2006). *'Trustguide Final Report.'* <http://www.trustguide.org/>.

NAO (2002). *'Better Public Services through eGovernment'*. National Audit Office (4/April2002):

ODPM (2005). *'Inclusion Through Innovation: Tackling Social Exclusion Through New Technologies'*, Social Exclusion Unit, Office of the Deputy Prime Minister. <http://www.socialexclusion.gov.uk/downloaddoc.asp?id=768>

ONS (2006), *'Office of National Statistics, Internet Access Households and Individuals.'* <http://www.statistics.gov.uk/pdfdir/inta0806.pdf>

PCC (2006). *'Plymouth City Council'*. <http://www.plymouth.gov.uk/homepage/environmentandplanning/planning/planningpolicy/ldf/ldfbackgroundreports/ldfbackgroundinformation.htm>

PCC-Experian (2006) *'Plymouth's Potential Catchment (Experian May 2003)'* <http://www.plymouth.gov.uk/homepage/communityandliving/citycentre/bid/bidbusinessplan/bidplymouthspotential.htm>

Public Sector Forums (2006). *'Why Take-up Campaign Claims Are Complete Rubbish'*, <http://www.publicsectorforums.co.uk/page.cfm?pageID=2641>

Rogers, J (2003) *'EGovernment costs will outweigh savings as the UK focuses on improving services.'* Computer Weekly (17/06/2003)

The ‘Speech Music’ application and language variant analysis

C.C.Ford and S.L.Denham

Centre for Theoretical and Computational Neuroscience, University of Plymouth,
Plymouth, United Kingdom
e-mail: s.denham@plymouth.ac.uk

Abstract

The direct link between speech and music can be seen as a top priority within musicology research, with Neuroscience, Psychology and Music based research in this area answering many questions that will ultimately show whether this link truly exists. Both speech and music are derived through sound and deliver emotion and semantic meaning (to a greater or lesser extent) within their human communication boundaries. From a compositional point of view, western music has utilised speech throughout its development to ‘tell a story’ or ‘send a message’ and to affect the emotion of its listeners. This paper presents an assessment of the links between speech and music through a software engineering based update of a musical compositional tool entitled the ‘Speech Music’ application (Denham, 2005), which utilises at its core a computational model of the human cochlea (Coath *et al.* 2005 and Coath and Denham, 2005) to derive note sequences based upon frequency band energy and transient onset / offset data extracted from speech audio recordings. The research section ultimately aims to prove whether Carol Krumhansl’s ‘Tone Profile’ key finding algorithm (Krumhansl and Kessler 1982) is a robust tool for musical key finding.

Keywords

Speech, Language, Music, ‘Speech Music’, Cochlea, Musical Key, Tone Profiles.

1. Introduction

Music creation and performance has been practiced by cultures throughout the globe both as an art form for recreational purposes and also as a means of emotional and semantic communication for many centuries. Western music is historically shown to have begun circa 450AD along with the spread of Christianity, wherein music was performed as an integral part of Christian worship. The ‘Gregorian Chant’ (a polyphonic vocal musical style) was thus derived, taking it’s themes from the stories of the bible, with its composers being supported by the church. By the 15th Century, music notation for polyphonic music had been highly developed by composers from the Church and the newly formed Universities, and this bore music to a wider audience within the royal courts of many European countries. The ideas and principles of communication within music stayed and grew in this environment with the advent of plays and then operas composed largely for the royalty of the land.

The communication of semantic meaning and emotion has been borne through the history of music and as such there is a large belief of a direct link between both music and speech, the main form of direct human communication.

Today, musicologists based in Neuroscience, Psychology and Music based research centres around the globe, with their greater access to knowledge of the workings of the human brain, are performing research which is getting ever closer to answering the question of whether this direct link between ‘Speech and Music’ actually exists.

One such project, developed at the University of Plymouth’s ‘Centre for Theoretical and Computational Neuroscience’, has seen the creation of a computational model of the human cochlea (Coath *et al.* 2005 and Coath and Denham, 2005) which was then further implemented within a music composition software program entitled the ‘Speech Music’ application (Denham, 2005). The ‘Speech Music’ application allows for the creation of between 1 to 4 tracks of ‘Speech Music’ fragments, essentially musical note sequences, based upon the frequency band energy and transient onset / offset data derived from the computational cochlea models analysis of a source sound file of human speech, thus allowing for the creation of ‘Speech Music’ based compositions.

The goals of this study defined an update to the ‘Speech Music’ application that included the addition of Musical Instrument Digital Interface (MIDI) functionality, which allowed and inspired pure research on a number of language variant ‘Speech Music’ fragments (created through translation of a composed poem) with respect to testing Carol Krumhansl’s ‘Tone Profile’ algorithm (Krumhansl and Kessler, 1982) which utilises 24 tone based profiles of the Western chromatic keys (12 Major and 12 Minor) based upon probe tone ratings from multiple research subjects. The ‘Speech Music’ application was then further utilised to compose a professional composition based around the same language variant ‘Speech Music’ fragments, to show the newly implemented compositional possibilities within the application.

2. Background

Historically, there have been strong links between music and speech, to the point where it is believed that prior to the invention of language, humans communicated via sound, with differing sequences of pitches holding differing emotional meanings. This type of communication would likely have sounded like a primitive song (Arbo, 1998).

Recent studies strengthen this link with specific research showing direct chromatic links within emotional expression through speech. Happy and sad emotions are generally expressed in western musical styles through Major and Minor keys and this same function is found within everyday language based communication (Schreuder et al, 2005) through analysis of the pitch contours of sad and happy speech from multiple speakers.

As an extension of this link, strong correlations have been found between human vocal processing and musical pitch division (Schwartz et al, 2003) showing that the main frequency bands utilised in human vocal communication (due to the biological design of our vocal tract and the frequency / tonal patterns it creates), tally with the 12-notes of the western ‘chromatic’ scale, no matter an individual’s language / culture (Figure 1).

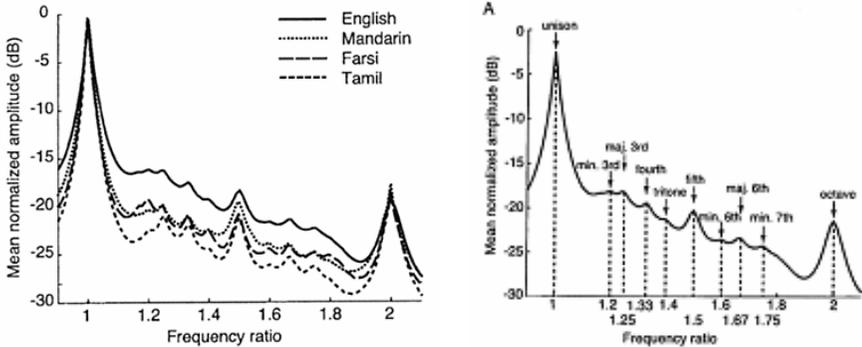


Figure 1: ‘Vocal tone link to chromatic scale’. (Schwartz et al, 2003)

“It is an ancient, and very pervasive, idea that music expresses emotion”. (Scherer and Zentner, 2001).

The author would deem that this is the reason for music and the reason why he himself composes. The way it influences the human psyche by creating extremes of happiness or sadness, dependant upon the music style and the current emotive state of the listener, makes music very special to humanity in many ways.

This link to emotion is now becoming a large area of study within the neurological and psychological sciences. This area of research is likely to have a huge impact on the music of tomorrow, with respect to writing music that triggers specific emotional responses from its listeners.

Speech and music are both forms of human expression and (loosely, with respect to music) communication and their human psychological links are born out through terms such as ‘the language of music’ and the fact that the use of voice within music composition has been strong throughout history, both as the sole instrument (choir / Gregorian chants) and as part of a band (i.e. instruments and voice). Vocals have become so important musically due to the fact that they deliver ideas / meanings / semantics (generally story based) as well as a high conveyance of emotion.

With the recent advent of speech synthesis and sampling technologies, many contemporary composers are utilising vocal effects within their compositions in many abstract ways, sometimes even as percussion or rhythmical fragments (and as such not necessarily within the context of language semantics).

With the direct links to our biological make-up (i.e. our vocal tract) and emotional communication, the emotional and historic compositional links between speech and music seem ever more important.

Even within an evolutionary aspect, speech and music seem inextricably linked. Vocal and Aural systems seem to have developed side-by-side with the transformation to bipedalism around 1.8million years ago. A vocal system similar to our own appeared within ‘Homo Heidelbergensis’ around 300,000 years ago. Structural and prosodic aspects within speech and music processing and production

are performed very similarly within the human brain which strongly suggests shared evolutionary foundations (Morley, 2002).

3. Methodology

To show the usefulness of the ‘Speech Music’ application as a scientific tool, a pure research study was performed to analyse ‘Speech Music’ output from the application in a number of language variants and then analyse their key structures to test the robustness of Krumhansl’s ‘Tone Profile’ algorithm (Krumhansl and Kessler, 1982).

The test pre-processing involved the creation of a simple 4-line poem entitled ‘Life’s Song’:

*A word spoken, from the tip of your tongue
The sound of your voice, sending more than its sum
The language of life, defining dreams and desires
From your first to your last breath, echoing life’s song*

The poem was then translated into five other language variants, Chinese (Mandarin), German, Russian, Spanish (Catalan) and Thai.

The six language variants were then recorded as Wave (.wav) audio file format utilising Steinberg ‘Cubase SL’ sequencer software, through a Shure SM-58 Vocal Microphone, via an M-Audio ‘Audiophile’ Audio / MIDI interface at CD-Quality audio settings (i.e. 16-bit, 44.1KHz Sampling Rate) and as mono track output.

The outputted Wave (.wav) audio files were then edited within ‘Cubase SL’ with the following audio editing processes:

- Trim (to cut excess silence).
- Fade In / Fade Out (to erase any noise at the beginning / end of the audio).
- Normalisation (to maximise the amplitude level of the final output).

3 Male and 3 Female subjects were used in the language variant recordings:

- Male: English, German and Spanish (Catalan).
- Female: Chinese (Mandarin), Russian and Thai.

These recordings were then processed through the ‘Speech Music’ application to output their associated ‘Speech Music’ fragments which were then analysed through modules from the MIDI Toolbox (Eerola and Toiviainen, 2006), utilising techniques for analysis of:

- Note Distribution.
- Key Finding (utilising Krumhansl’s ‘Tone Profile’ algorithm).
- Tonal Strength (SOM Modelling).

A professional composition was also produced to show the new music composition abilities of the ‘Speech Music’ application. The composition utilised the derived language variant ‘Speech Music’ fragments, exported as Standard MIDI File (SMF) data and further imported into professional sequencer software for compositional creation.

4. Results

4.1 Note Distribution

Language	Catalan	Chinese	English	German	Russian	Thai
Key	<i>C# Minor</i>	<i>F Major</i>	<i>F# Major</i>	<i>D# Major</i>	<i>F Major</i>	<i>F Major</i>

Note distribution results would, in a generic piece of western music, give direction to the key signature of the piece in question, however, this seems somewhat difficult where the information is derived from a purely energy based model, as with the Gammatone filter (Slaney, 1994) based ‘Cochlea’ model (Coath *et al.*, 2005 and Coath & Denham, 2005) used within the ‘Speech Music’ application.

The keys were derived visually from the results for each language based upon a mean average of all triads within the octave, but most were close to falling into a number of key signatures concurrently throughout or falling into a convoluted key signature, such as seen within the English structure (Figure 2) which falls more closely into ‘F# Major or C# Minor 7th’ (non-triad based) key signatures than it’s ‘Tone Profile’ algorithm derived A# Minor.

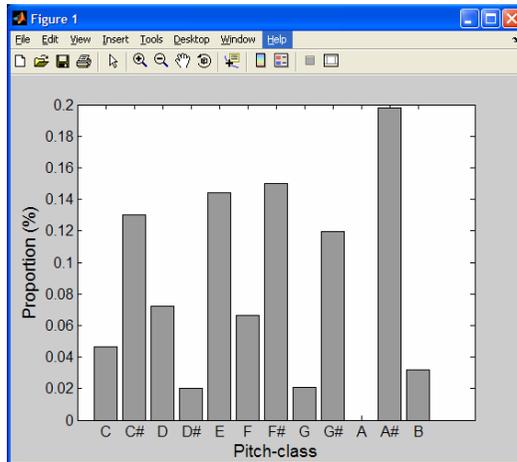


Figure 2: ‘Note Distribution Results for English Language Variant’.

There were many whole tone steps between the highest mean triad based note proportions which do not fit the **T**, **T**, s**T**, **T** (Major) or **T**, s**T**, **T**, **T** (Minor) triad sequences (where **T** = whole tone, s**T** = semitone and emboldened entries are the 2nd and 3rd notes that make up the Triad Chord from the tonic (root) note), hence the closest fit based upon the triad sequences was chosen.

4.2 Key Finding

Language	Catalan	Chinese	English	German	Russian	Thai
Key	<i>C# Major</i>	<i>F Major</i>	<i>A# Minor</i>	<i>G Minor</i>	<i>F Major</i>	<i>F Major</i>

Key Finding results were derived utilising Carol Krumhansl’s ‘Tone Profile’ algorithm (Krumhansl, 1990) via a function within the MIDI Toolbox (Eerola and Toiviainen, 2006).

Chinese, Russian and Thai equate to the same key of F Major and also gain the same key results from both the Note Distribution and ‘Tone Profile’ Algorithm, whereas Catalan, English and German garner differing results. This leads to a deliberation of whether Krumhansl’s Tone Profiles’ perform their role of key finding adequately. The problem here is that the output from the ‘Speech Music’ fragments is not based upon specific key structures, in that our analysis shows that many notes occur in and around differing keys. This brings up the effect through key analysis that at points in the ‘Speech Music’ a specific key does not exist, or that keys change on concurrent beats within the fragment. Keys are seemingly dependant upon the strengths of these notes at each beat point (as covered in Krumhansl’s ‘Tone Profiles’) and as such there seems no clear way with Note Distribution values to clearly judge any inherent key within a fragment

Also, and a lot more controversially, through creating the ‘Life’s Song’ composition, the keys used throughout the triad based Choir backing and the Cello solo within the piece were taken from those derived from each ‘Speech Music’ fragment through Krumhansl’s ‘Tone Profiles’ and these work within a Western Tonality perspective, sounding musical and linking well with the ‘Speech Music’ fragments within the composition.

Conversely though, the note distribution values for Catalan show a definitive key structure of C# Minor (C#, E, G#), whereas Krumhansl’s ‘Tone Profile’ algorithm chooses C# Major just over C# Minor. Admittedly, the Major 3rd of C# Major (i.e. F) is the 5th highest peak at 0.075, but it is still very low compared to the Minor 3rd (i.e. E) at 0.175, which is more than double its distribution (Figures 3).

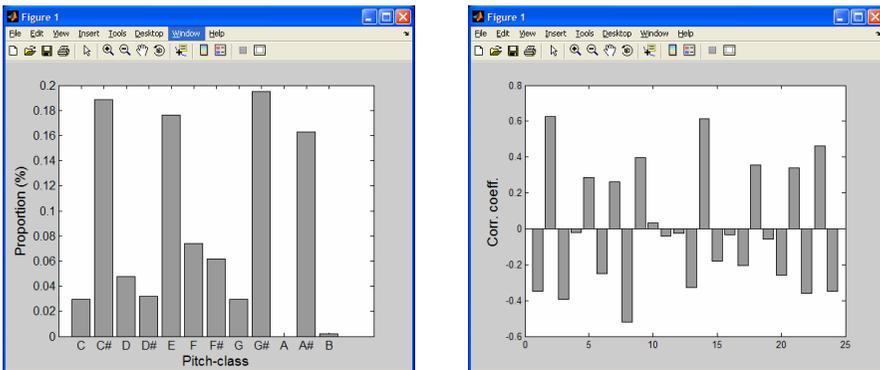


Figure 3: ‘Note Distribution and Key Finding Results for Catalan’.

4.3 Tonal Strength

Language	Catalan	Chinese	English	German	Russian	Thai
Key 1	<i>C# Major</i>	<i>F Major</i>	<i>A# Minor</i>	<i>G Minor</i>	<i>F Major</i>	<i>F Major</i>
Key 2	<i>C# Minor</i>	<i>F Minor</i>	<i>C# Major</i>	<i>D# Major</i>	<i>F Minor</i>	<i>F Minor</i>
Key 3		<i>C Major</i>	<i>F# Major</i>	<i>C Minor</i>	<i>G Minor</i>	<i>A# Major</i>
Key 4		<i>G Minor</i>	<i>C# Minor</i>	<i>C Major</i>		

The SOM model correlated for each language gave a strength rating over the time of the performance, with the following tonal strength responses per language based within the +0.5 to +1 (upper 25%) area. The first two keys were by far the strongest and as such these were taken as the analysis data.

5. Conclusions

5.1 Note Distribution and Key Finding Hypotheses

The results reveal Chinese, Russian and Thai as all being within the key of F Major. These were the three poems read out by females, whereas Catalan, English and German were read by males. This could potentially indicate a gender specific profile.

Also and somewhat more interestingly, the following note frequencies were missing in the following languages:

Language	Catalan	Chinese	English	German	Russian	Thai
Missing Notes	A	D, F#	A	D, F#	D, F#	D, F#

These ‘always missing’ notes could indicate a potential link between the style of language being spoken.

English and Catalan do not possess tonality, whereas the others (with the possible exception of German) do. “Tonal languages contain changes of the fundamental frequency pattern within a phonemic segment which determine the lexical meaning” (Koelsch *et al.*, 2004).

Also, English and Catalan are based upon Latin as a base language whereas the others are not. This could potentially be linked to a remnant of historical language style.

F, G and A are predominant where D and F# are missing (i.e. Chinese, German, Russian and Thai), whereas C#, E, G# and A# are predominant when A is missing (i.e. Catalan and English). These ‘always present’ notes lead strongly to some kind of reference between these two language groups, be it Tonality, Initial Base Language or some other hypotheses.

5.2 Tonal Strength Hypotheses

Catalan and English show definite similarities, both containing C# Major with the English main key of A# Minor being the dominant 5th from the Catalan 2nd of C# Minor. Chinese, Russian and Thai again show their tonal links with the same F Major / F Minor footprint.

German stands alone in this research segment with its keys being one tone away from all other languages in either direction (i.e. F→G, C#→D#).

Again, even with the exception of German in this case, these results point strongly to some kind of reference between these two language groups, as stated, be it Tonality, Initial Base Language or some other as yet unobvious hypotheses.

5.3 Overall Conclusions

These results show an obvious link between Chinese, Russian and Thai through both Note Distribution and Krumhansl's 'Tone Profile' results.

Also, the links between 'always missing' and 'always present' notes, shows a definitive group structure of Catalan and English as one group and Chinese, German, Russian and Thai as another. Key strength results show German standing alone and leaving its close links with the second group, but all other group languages again show a close relation. Links here between language style (i.e. tonality), historical base language structure or some other hypotheses seem very strong indeed.

Going back to our initial pure research idea, no definitive proof could be found that Krumhansl's 'Tone Profiles' are either a reliable or ineffectual way of indicating key structure within a musical piece.

Overall, this analysis in every area shows that further research and analysis needs to take place before any real proof of any of the hypotheses from these results can be deemed to be pertinent. The results however do strongly indicate a language based link and as such give great reason to perform this extended research analysis in the future.

6. References

Arbo, A. (1998) *'La truce du son. Expression et intervalle chez Condillac'*. In J.-M. Chouvel and M. Solomos (eds.) *L'espace: musique/philosophie*. Paris: Editions L'Harmattan.

Coath, M., Brader, J. M., Fusi, S. and Denham, S. L. (2005) *'Multiple views of the response of spectro-temporal features supports concurrent classification of utterance, prosody, sex and speaker identity'*, *Network: Computation in Neural Systems*, Vol 16 2/3: 285-300.

Coath, M and Denham, S. L. (2005) *'The role of onsets in auditory processing'*. Centre for Theoretical and Computational Neuroscience, University of Plymouth.

Denham, S. L. (2005) *'The Speech Music System'*, Centre for Theoretical and Computational Neuroscience, University of Plymouth.

- Eerola, T. & Toiviainen, P. – MIDI Toolbox Website (2006) ‘MIDI Toolbox Product Home Page’, <http://www.jyu.fi/musica/miditoolbox/>. (Accessed 27th July 2006).
- Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T. and Friederici, A. D. (2004) ‘*Music, Language and meaning: brain signatures of semantic processing*’. *Nature Neuroscience*, Vol 7, No. 3:302-307.
- Krumhansl, C. L. and Kessler, E. J. (1982) ‘*Tracing the dynamic changes in perceived tonal organisation in a special representation of musical keys*’. *Psychological Review*, 89:334-368.
- Morley, I. (2002) ‘*Evolution of the Physiological and Neurological Capacities for Music*’, Cambridge University Press.
- Scherer, K. R. and Zentner, M. R. (2001) ‘*Emotional Effects of Music: Production Rules*’, *Music and emotion: theory and research*. Oxford ; New York : Oxford University Press, Chapter 16, 361-92.
- Schreuder, M., Eerten, L. v. and Gilbers, D. (2005) ‘*Speaking in Minor and Major Keys*’, University of Groningen.
- Schwartz, D. A., Howe, C. Q. and Purves, D. (2003) ‘*The statistical structure of human speech sounds predicts musical universals*’, *The Journal of Neuroscience*, 23(18): 7160-7168.
- Slaney, M. (1994) ‘*Auditory Toolbox Documentation: Technical Report 45*’, Apple Computers Inc., www.Slaney.org/malcolm/apple/tr45/AuditoryToolboxTechReport.pdf. (Accessed: 15th December 2005)

Autonomous robot navigation in buildings: a case study using the Evolution ER1 robot

D.A.E.Harewood-Gill and T.Belpaeme

University of Plymouth, Plymouth, United Kingdom
e-mail: tony.belpaeme@plymouth.ac.uk

Abstract

This paper focuses on the work carried out on a Evolution ER1 Robot, the aim of which was to create an autonomous robotic application that was capable of exploring an office environment without human intervention or without a planned path. The robot only relies on images from a USB web camera for its sensory input. Two implementations are described, one for obstacle avoidance and one for detecting humans. The obstacle avoidance application uses edge detection to locate objects. A control system is used to move away from obstacles depending on where the obstacle has been identified in the camera image. The human recognition system works around the principle of SIFT to identify footwear. The input image is subjected to a blurring function followed by a sharpening stage to generalise the shape. This image is then compared with other generalised images stored in a library. If matched, the robot moves away from the shoe with the assumption that there is a human present. Footwear that has been stored in the library was 10% more likely to be recognised than footwear that was not.

Keywords

Evolution ER1 Robot, navigation, SIFT, obstacle avoidance.

1. Introduction

Derived from the Slavic word which meant worker or slave, stories created delighted and terrified people who set out to see these metallic workers run amok or to rebel against their human masters (McKerrow, 1993). Eighteenth century Europe saw the creation of mechanical puppets that moved the same way again and again, however, until the early part of the 20th century, robotics was a source of fiction and amusement. Now, for the last 50 years robots have become widely deployed. Robots are used in many of the worlds leading industries, mainly the automotive industry and have also critical in carrying out tasks that would be too dangerous for humans to perform such as tasks involving the disposal of explosives, the control of nuclear materials, search and rescue, or exploration.

However, in most of these cases, the systems mentioned were either programmed to repeatedly perform the same task not allowing any flexibility in their operation or had human operators controlling their every move. Robots are not required to reason or act for themselves. This ability will be required to some greater or lesser extent if robotic systems are to evolve to a point where tasks of increasing complexity can be given to them. Technology is also a lot cheaper than it used to be, and it expected that robots aimed at the edutainment market will be made for only a few hundreds of

dollars. This paper discusses the work carried out on autonomous robot navigation in buildings.

2. Platform

The robot used during the course of this research was the Evolution ER1 Robot from Evolution Robotics (Evolution Robotics 2006). The ER1 is a robot aimed at robot enthusiasts and is as such not designed for one specific task, but can be used for tasks ranging from teaching to research to edutainment. The ER1 robot is essentially an aluminium frame that is slotted together. Attached to this frame are two stepper motors, two wheels, a drive module containing two microcontrollers, and a USB web camera. The robot frame also supported an Intel Pentium M laptop. The laptop was connected to the robots drive module and the web camera, both by USB allowing the control of both of these systems. The overall system could either be controlled from the laptop, or through the laptop via TCP/IP (Transport Control Protocol/Internet Protocol) from another computer (figure 1).

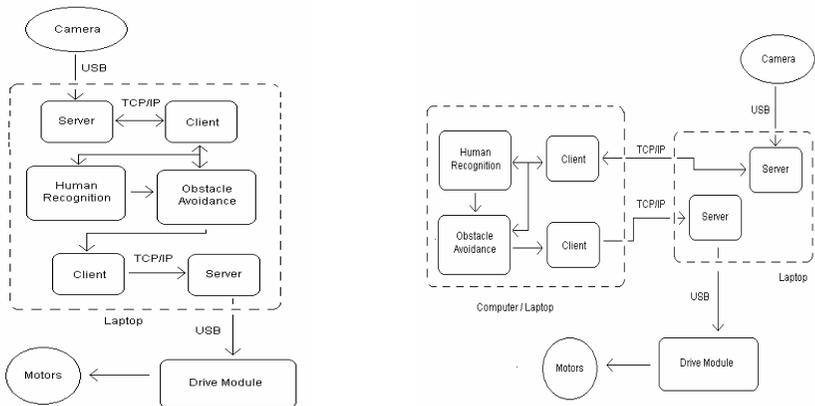


Figure 1: Showing the two possible system configurations. Left: Using one computer. Right: Using two computers, one sitting on the ER1 robot, the other a remote computer.

TCP/IP is a very useful protocol. It allows the client written by one computer language to talk to a program written in another either on the same computer or spread across several. In both configurations, TCP/IP was used to transport data amongst the running programs and systems.

3. Design

The ER1 Robot as previously stated is required to move around autonomously while avoiding obstacles and avoiding humans where possible. For both of these programs, the following assumptions are made:-

- For both the obstacle avoidance and human recognition applications it is assumed that the ground will be level (also known as the *ground level constraint*).

- The robot will run indoors in a typical office building.
- For the obstacle avoidance program, we assume that the floor will always have the same colour and texture.
- For the human recognition program, we assume that humans will always be wearing shoes or footwear of some sort.

It is necessary to define the environment and conditions that the robot would be exploring, and with these assumptions it is possible to design certain aspects of the program.

3.1. Obstacle avoidance

The first step is to take an image which is filtered with a median filter to remove any noise that may have appeared. The camera is aimed at the floor. This means that any obstacle appearing in the cameras field of vision would either appear from the side of the image or the top. This is an important consideration. For obstacle detection, a vision technique called edge detection (Gonzalez and Woods 1992, pp 414 - 428) is used.

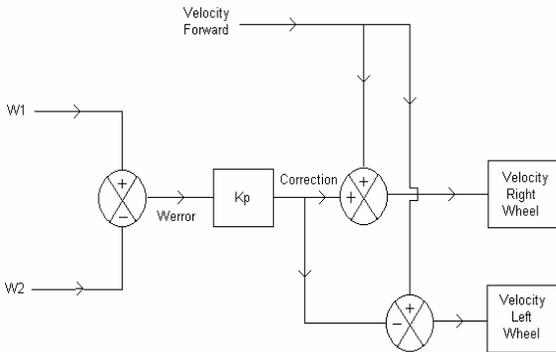


Figure 2: Showing the control system used for steering during obstacle avoidance.

It allows the distinction between edges by measuring spatial gradients and applying a series of masks. Once an edge is detected, anything below the edge was considered clear space; anything above including the actual edge is considered an obstacle. Movement instructions are then given using a control system approach, see Figure 2. The rate of turning depends on how much of the screen the object covered on one side or another, determined the amount of free space on the left and right hand side from the object. These movement instructions are then sent to the left and right stepper motors. Three threshold controlled functions are also added. The first is designed to check to see how much of the screen the obstacle covered; if it covers too much then the robot was instructed to turn around. The second program is aimed at seeing if there are any obstacles directly in front of the robot. If there are then the robot will again turn around. This was added to take into account the fact that if an obstacle is located dead centre of the screen and spread evenly over both sides from the centre, the robot will still consider going forward as the optimal answer.

3.3. Human recognition

This program is based upon the assumption that people in the operating area of the ER1 Robot would be wearing footwear of some sort. The heart of this algorithm is a program called Scale Invariant Feature Transform or SIFT (Lowe, 2006). SIFT works by taking an image that had been blurred and then sharpened to generalise the shape, blurring it with a Gaussian function, comparing it to several other blurred versions of the same image at different scales, comparing the maxima and minima of these blurred images to find key points via the gradient and then using a set of criteria to filter out the key points that did not measure up.

A series of images of footwear are taken at different angles, processed using SIFT and then the key points of these images are stored into a library. When a new image of a shoe is taken, it is processed and then compared to those stored in the library. If any matches are found, the mean of the sum of the x co-ordinates of all matching key points found are calculated, and this is fed into an equation.

$$d = \text{magicnumber} \left(\frac{\text{width}}{2} + \text{mean} \right)$$

Where d is the distance in degrees to rotate, width was the width of the image, mean is the mean x co-ordinate found and magicnumber is a user set value that affects the severity of the turn, this is set to 0.1.

3.4. Supporting Programs

There are several programs used to supplement the control of our main applications. The first of these is a replacement program the default ER1 Control Centre for controlling the ER1 Robot. Although the ER1 Control Centre has a wide range of options, it does not allow low-level commands to the motors. The program used instead is written in C++ called ER1MoveIT! It allows direct access to the ER1's microcontrollers and makes it possible for the robot to perform fluent motion patterns, instead of the straight moves or on-the-spot rotation provided by the ER1 Control Centre. This program could also receive commands through an TCP/IP connection.

To obtain images from the USB web camera, there were two options available. The first was to use a camera server, again connected via TCP/IP. The second option was to use a program that took images periodically and saved them to a specific location. For the latter method, a program called Coffee Cup (Coffee Cup 2006) was used.

A Matlab toolbox called ER1 toolkits created by the University of California had several small programs for employing Matlab with the ER1 Control Centre. These programs could be used to establish a TCP/IP connection, send movement commands, or acquire images from the camera server.

4. Experiments and Results

To assess the obstacle avoidance and human recognition programs written, several tests were conducted.

4.1. Obstacle Avoidance

Initial tests conducted were aimed at calibration of the obstacle avoidance system and for familiarising with the system as a whole. The criterion for the test was the length of time the ER1 Robot could avoid obstacles without colliding with one. The initial set up of the system was aimed at calibrating the system via the camera position for the correct angle of elevation and centring it along the x-axis, the first and second tests were aimed at doing the same thing, but only doing it in software. The first tests performed were aimed at experimenting with the threshold values of the control system. The constant Kp visible in Figure 4 and the value for the forward velocity were the first to be altered the aim of course was to increase overall performance. Test 2 was aimed at optimising the threshold values for the two movement functions mentioned previously.

Test 3 was carried out in two locations, a crowded robotics lab, and a corridor. Test 3 had the same criteria as the first two tests. The test was run 20 times in each location to ensure a fair result and each time placed in the same location from which to start. The results for Test 3 showed that performance was better in open spaces with an average time of 39.1 seconds for the corridor compared to 29.1 for the robotics lab. After the completion of Test 3, two further tests were carried out. In each of these tests, the Forward Velocity was decreased to 8.5 cm/s for Test 4, and 8.0cm/s for Test 5 where the constant, Kp was increased to 0.0009 for Test 4 and 0.00095 for Test 5. Test 4 and Test 5 were started in 4 different orientations 90 degrees from the previous one; the test was conducted five times in each orientation to obtain an average.

Positions	Lab (Sec's)	Corridor (Sec's)
1	45.7	42
2	32.3	45.1
3	33.3	57.9
4	42.2	42
Average	38.8	46.7

Positions	Lab (Sec's)	Corridor (Sec's)
1	53.7	61.4
2	52.2	62.2
3	55.6	56.6
4	47.2	58.9
Average	52.2	59.8

Table 1: Left, showing the results of Test 4, Right Showing the Results of Test 5

The results were similar to that of Test 3, the corridor results were better on average, however, the gap between the corridor results and the robotics lab results were decreasing. The final overall performance of the obstacle avoidance program was respectable but left room for improvement.

4.2. Human Recognition

As previously mentioned, shoes were used as a means for checking to see if there were humans within the local area (Figure 3).



Figure 3: Showing some of the different shoes used viewed at different angles.

The initial test for the human recognition program was aimed at the effect the resolution size of the image would have on SIFT. The camera when taking images was set to a resolution size by the user. However, as **Table 2** shows, the resolution size directly affected the processing time of SIFT and the number of key points found.

Resolution	Response Time (secs)	No of Key Points
160 x 120	4.06	12
176 x 144	5.44	15
320 x 240	18.72	28
352 x 288	26.56	47
640 x 480	95.48	54

Table 2: Showing the performance processing time and number of key points depending on resolution size.

The larger the resolution, the more key points found. However, the larger the resolution time, the longer the processing time. The default resolution eventually chosen was 352x288 as it returned a fair number of key points for a reasonable processing time. To go down one resolution side would have nearly halved the key points found.

The next test focussed on the performance of the system as a whole. Five shoes that had been entered into the library and five shoes that hadn't were tested with the Human Recognition program. Each shoe was tested in 12 different orientations in relation to the camera to see how well the system worked. Table 3 shows the results.

Table 3 shoes that shoes known to the system were recognised better on average than those not known to the system, though unknown shoes scored did fairly well. Table 4 shows the number of matches on average for the twelve orientations found for each

of the five shoes and Table 5 shows the number of key point's matched from the subject shoe to the first shoe found on average for the twelve orientations of the five shoes.

Shoe	Known Shoes %	Unknown Shoes %
1	83	50
2	50	67
3	58	42
4	67	67
5	58	50
	63.2	55.2

Table 3: Showing the average success rate for each of the five known and unknown shoes for each of the 12 orientations and the overall average for each.

Shoe	Known Shoes	Unknown Shoes	Shoe	Known Shoes	Unknown Shoes
1	1.25	0.58	1	2.67	0.58
2	0.67	0.83	2	1.08	0.92
3	1	0.41	3	1.33	0.5
4	1	0.75	4	1.42	1
5	0.75	0.5	5	0.75	0.75
	0.93	0.61		1.45	0.75

Table 4 Left, and Table 5 Right: Showing the number of matches to database shoes and the number of key points for the first shoe matched.

It can be seen from these tables that the number of matches and key points found for known shoes is a lot higher than those unknown to this system which really isn't that surprising. What is surprising however is while the number of matches on average is respectable, the number of key points matched with those of the first shoe to be matched with the subject shoe is very low when considering each image on average generates 50 key points when first analysed.

5. Discussion

The outcome of the tests described in section 4 shows that for the most part, the system worked as expected. The obstacle avoidance system has several problems. The first and foremost is the USB web camera. Calibrating the position of it is a tedious process but needs to be performed to stop bias to one side or the other. It was also noted that the robot has a blind spot, if it is close to being parallel to the wall, then there is a chance it will not see the wall and run into it. Another characteristic that developed is when the robot was placed in the same place facing the same way; it did not always follow the same path as expected. Roughly 50% of the time it would follow a different course. The rest of the time it will follow the same course; this was part of the reason that Tests 4 and 5 were altered so the robot would face four different directions. The test results from the obstacle avoidance test reveals several things. Firstly and unsurprising obstacle avoidance was easier in an open spaces like a corridor. Secondly, the performance improved over time as shown by

Table 1 after some minor calibrations to each of the three threshold controlled movement functions and the constant Kp . Finally the average time gap between the corridor tests and robotic lab tests decreased with further calibrations of the three threshold controlled movement functions and the constant Kp .

The performance of the human recognition system was fairly high for an experimental system. Shoes that were known to the system were more readily recognised than those which were not with an average of 63.2 % to 55.2% respectively. This was not really surprising. The processing time for SIFT to calculate key points was fairly high, too long for a mobile autonomous robot acting in real-time. However, the number of key points being matched is very low and to decreasing the image resolution will greatly reduce the results of library images being matched with newly captured images. To increase the overall systems performance it is possible a larger image library would result in more matches. A higher resolution would also result in more matches, as long as an effective method was found to reduce the processing time required for a larger resolution.

6. Conclusion

To conclude, the object detection program worked by taking an image, filtering out any noise then applying an edge detection function. Depending on the obstacles location, depended on how the robot avoided the obstacle. The system can avoid obstacles on average for 52 seconds in enclosed spaces and up to a minute in open areas. The human detection program works by assuming that humans would wear shoes at all times. When an image was taken of a shoe, the image would be blurred then sharpened to generalise the shape, and then key points are extracted from this image using SIFT. These key points are compared to a library of already processed images. If a match is found, an angle is calculated and the robot will turn away from the shoe. Shows known to the program, e.g. those images of shoes already stored in the library are more likely to be recognised. The success rate of shoes in the library matching with those retrieved from the camera in real time also goes up when the size of the camera resolution is increased.

There is still room for improvement. The recognition of shoes is not high and could certainly be improved by tuning the parameters of the SIFT feature detector. Also, the reactivity of the robot could be improved by compiling the software; right now the image processing and detection software runs as interpreted Matlab code and several concurrent applications are used to access the camera hardware and the robot hardware. These could all be combined into one single compiled program, which would drastically improve the performance.

7. References

- Coffee Cup (2006), <http://www.coffeecup.com/webcam/>
- Evolution Robotics (2006), <http://www.evolution.com/er1/>
- Gonzalez, R. and Woods, R. (1992) Digital Image Processing. Addison Wesley.

Lowe, David G, (2004), “Distinctive Image and Features From Scale Invariant Features”, International Journal of Computer Vision, Vol. 60, No. 2, PP 91-110.

Mckerrow P.J(1993), Introduction to Robotics, Published by Addison-Wesley Publishers Ltd, ISBN 0-201-18240-8, pp 2.

Information systems integration in virtual learning environments

N.Mcilree and A.D.Phippen

Network Research Group, University of Plymouth, Plymouth, UK
e-mail: info@network-research-group.org

Abstract

This work seeks to present an analysis of the methods used for integrating information systems within a managed learning environment. The research is based on a series of online surveys and interviews with FE and HE institutions in the South West of England conducted during August and September of 2006. The results of the research indicate that whilst interconnectivity between information systems within MLEs does exist, it is on a principally ad hoc basis and differs from institution to institution. Likewise, the full range of integration methods and technologies currently available are not being used to their full extent.

Keywords

MLE, VLE, MIS, library management systems, integration, JISC, interoperability, RSC, student records, information systems, education, e-learning.

1. Introduction

A longitudinal study undertaken in 2005 on behalf of UCISA (Universities and Colleges Information Systems Association) found that out of 85 HE institutions surveyed 95% currently used Virtual Learning Environments (Jenkins et al, 2005). With the implementation of VLEs approaching a state of near ubiquity within the HE/FE sector a great deal of focus has now turned to the examination of MLE (Managed Learning Environment) implementation within UK educational sector. In particular the government advisory body JISC (Joint Information Services Committee) has funded many studies into the area of MLEs. This study intends to add to the existing research by providing a technical review of how FE/HE organizations have chosen to integrate their information systems with regards the provision of an MLE.

The term MLE refers to the information systems whose combined actions enable the provision of an integrated electronic learning and administrative platform. Rather than a single application or product an MLE should be viewed as a collection of individual information systems whose combined interactions provide the functionality attributable to an MLE.

“Managed Learning Environment (MLE) refers to the whole range of information systems and processes of an institution (including its VLE if it has one) that contribute directly or indirectly to learning and the management of that learning.”

Social Informatics Research Unit et al (2003)

Typical functionality one might expect from an MLE consists of;

- Access to course material and a variety of differing learning resources through user centric single point of entry.
- Tracking of learner progress and learner utilization of above resources.
- Simple management of above resources administrative and academic staff.
- Electronic assessment and coursework receipting.
- Provision of learner and institutional information to staff members and learners.
- Electronic enrollment of learners and suitable provisions for electronic payment.
- Widening of access and participation.
- Access and management of various I.T resources such as internet, email, print and network storage.
- Access to library services both on and off-site.

2. Research Method

The research that informs this study was undertaken during August and September of 2006. It is comprised of a series of interview based case studies of four FE/HE institutions, and an online survey distributed through mailing lists run by the South West RSC (Regional Support Center). The RSC is a JISC funded body that offers support to FE/HE institutions on a regional basis.

The online survey presented a series of questions focused on six information system deemed to be commonplace in most MLEs.

- Virtual Learning Environments
- Management Information systems or Student Record Systems
- Library Management Systems
- Timetabling Systems
- Intranets
- Computer Networks

Each section requires respondents to identify if integration occurs between the different systems, and if so, what method was employed. In tandem with this survey four HE/FE institutions were directly interviewed to gain a more detailed understanding of what integration methods were being employed.

3. Results

The research indicated that all information systems reviewed can be considered both consumers and producers of information to other information systems. However, MIS systems were found to output significantly more information than any of the other information systems. This has led to the view that most MLEs can be characterized as being MIS centric, in that where data exchange does take place the highest instance is from the MIS to the other information system.

Networks were classified as the next highest producer of exchanged information. In the context of the case studies this can be characterized by their role as provider of authentication services via network directories. In the case studies all of institutions interviewed used network directories to provide user credentials to other information systems – usually VLEs or intranets. In most instances this revolved around the VLE or Intranet employing LDAP lookups to verify user identity against a network directory. Contrary to what one would expect from the case studies, the online survey indicated that networks were lowest consumer of resources from MIS. In the case studies all the network directories received their user information from MIS. The response from the online survey did not reflect this indicating only 45% of institutions surveyed passed information from MIS to the network.

Overall Intranets were the largest consumers of information across the range of systems reviewed. This can be attributed to the wide range of services associated with intranets. Unlike MIS, timetabling, or library management systems, intranets are not as function specific. In addition the usage of intranets as a portal or gateway to different resources supports the survey results in indicating they consume the most information.

Library Systems and Timetabling had the most limited range of integration with other systems. In fact library systems had interactions almost only with MIS. Similarly timetabling showed limited interactions with any systems other than MIS. With regards library systems one might assume this is student information being passed from MIS to library system. However, in the case of timetabling there is also a 45% instance of information being passed back to MIS.

The rates of interaction and average consumption of information are summarized in the two tables below.

Output >	MIS	VLE	Intranet	Network	Timetabling	Library	Average
MIS	NA	51%	43%	56%	56%	60%	53%
VLE	6%	NA	6%	18%	6%	6%	9%
Intranet	6%	0%	NA	18%	6%	0%	6%
Network	25%	25%	47%	NA	13%	6%	23%
Timetable	56%	0%	18%	12%	NA	0%	16%
Library	8%	8%	16%	8%	0%	NA	8%

Figure 1: Summary of interactions

	MIS	VLE	Intranet	Network	Timetabling	Library	Average
Average Consumption	20%	17%	26%	22%	16%	14%	19%

Figure 2: Average consumption

With regards the information systems used by the targeted institutions it is very easy to identify the market leaders. The area of library management systems is principally the domain of two products – Heritage and OLIB. ebs and UNIT-e are the most widely used MIS systems, whilst for networks and servers Microsoft is definitely the most commonplace. Once again Microsoft enjoys a wide share of database usage with Oracle and MySQL following closely. In the context of this study MOODLE is the most used VLE, though taking into account a previous UCISA longitudinal study (Jenkins, 2005) it would be foolhardy not to acknowledge the popularity of Blackboard and WebCT (which subsequent to that particular work is now owned by Blackboard).

Of the respondents that confirmed integration took place the methods used are indicated below.

Method of Integration	
Automated export feature of this System	2%
Manual	2%
Automated Export Feature	4%
CSV	4%
Third Party Integration	10%
Shared Data Source	10%
Manual export feature of this System	14%
Automated data transfer	25%
Direct Data link between this and target System	29%

Figure 3: Integration Methods

Automated data transfer and direct data link are the two most widely used techniques accounting jointly for 54% of the methods used. In the context of the case studies, and additional information supplied by institutions via the online survey it is evident that the actual implementations of these techniques are developed in house and on a generally ad hoc basis.

4. Conclusions

Other than provision for the manual import or export of standardized data formats vendors do not seem to have put in place any explicit functionality for data exchange with other specific information systems within a MLE. Or if they have this functionality is not being used by educational organizations. Given the limited number of information systems that exist within the MLE sector it would seem that vendors have overlooked a valuable opportunity to gain a commercial advantage by partnering with other providers of related systems. By including this sort of functionality into proprietary products interactions within MLEs would benefit from a higher degree of consistency and a reduction in the need for bespoke developments to be undertaken from department to department.

The research also indicates that despite the many methods that exist to facilitate system integration, SQL and LDAP are used almost exclusively as the sole means of transferring information. Likewise the many data aggregation and management

services and applications used in industry do not appear to have penetrated significantly into the FE/HE sector other than for the provision of authentication services.

This may be that the organizations surveyed are too small to benefit from these applications, or lack the technical skills or financial resources to implement them. However further investigation may be merited to examine to what degree FE/HE organizations would benefit from the implementation of industry level integration technologies.

With regards authentication there does appear to be an uptake of third party applications to manage these services. Three of the organizations in the case studies and one in the online survey have indicated that they intend to, or currently implement a centralized authentication method with regards their MLE. Though this is still a relatively small number, as the range of systems and resources included in MLEs increase, so too is it likely will the need for centralized management of user identity.

There is ample evidence to indicate that integration within MLEs is very much active within the FE/HE sector. However many organizations do not appear to be using the full range of methods and services currently utilized outside of this sector. Likewise vendors do not appear to be engaging with other suppliers of products that would typically be used in conjunction with their own. Obviously there are many constraints both with regards vendors and educational institutions, a study of which is not within the scope of this work. However, as previously stated investigation into bringing information systems together through vendor supported development would be a major step in increasing both the consistency and usability of MLE system integration.

5. References

Jenkins, M., Browne, T. and Walker, R. (2005), “A longitudinal perspective between March 2001, March 2003 and March 2005 for higher education in the United Kingdom”, UCISA 2005

Social Informatics Research Unit, (2003), “Managed Learning Environment Activity in Further and Higher Education in the UK”, University of Brighton, Education for change Ltd, and The Research Partnership

Real-time granular synthesis with spiking neurons

J.Murray and E.Miranda

Interdisciplinary Centre for Computer Music Research (ICCMR)
University of Plymouth, Plymouth UK
e-mail: eduardo.miranda@plymouth.ac.uk

Abstract

A new digital musical instrument based on granular synthesis using a pulse-coupled network of spiking artificial neurons to trigger the generation of sound grains is presented. The synthesis engine is connected to a network of spiking neurons, where the behaviour of a number of sub-groups in the network is used to control oscillators. A sound grain is generated from each oscillator when any one of the sub-groups neurons fire. The parameters used to control the oscillators and network of spiking neurons are controlled through a "Lemur" which is a touch-screen controller.

Keywords

Granular Synthesis, Spiking Neurons, Touch-Screen Controllers, Real-time Control

1. Introduction

A granular synthesiser that uses a pulse-coupled network of spiking neurons to trigger the creation of the sound grains has been developed. There are three levels of control used in the synthesiser: the Network Layer, the Granular Layer and the Performance Layer. The lowest level of control is the Network Layer. The Network Layer is concerned with the control and setup of the pulse-coupled neural network. The network connections, control over the networks and the behaviour of the neurons are all controlled in this level by providing control over the network's parameters.

The second level of control over the synthesiser, the Granular Layer, is concerned with connecting the oscillators used to generate the sound grains with the neural network setup under the Network Layer. The user has control over the oscillators at this level and can change each oscillator's parameters by selecting the amplitude and frequency respectively.

The highest level of control is the Performance Layer. This layer is for the general control over the synthesiser and all the other parameters used to generate and control the sound grains. This layer is for the overall control over live performance parameters such as grain duration, delay and spatial location. The pitch of the grains is also controlled here, but it is also related to the rate of firing of the neurons. This level provides the interface to the neural firing control. This model is essentially a polyphonic granular synthesiser in the sense that we have sub-groups of spiking neurons firing different streams of grains.

2. Overview of the System

An image displaying the relationships between the layers is given in Figure 1.

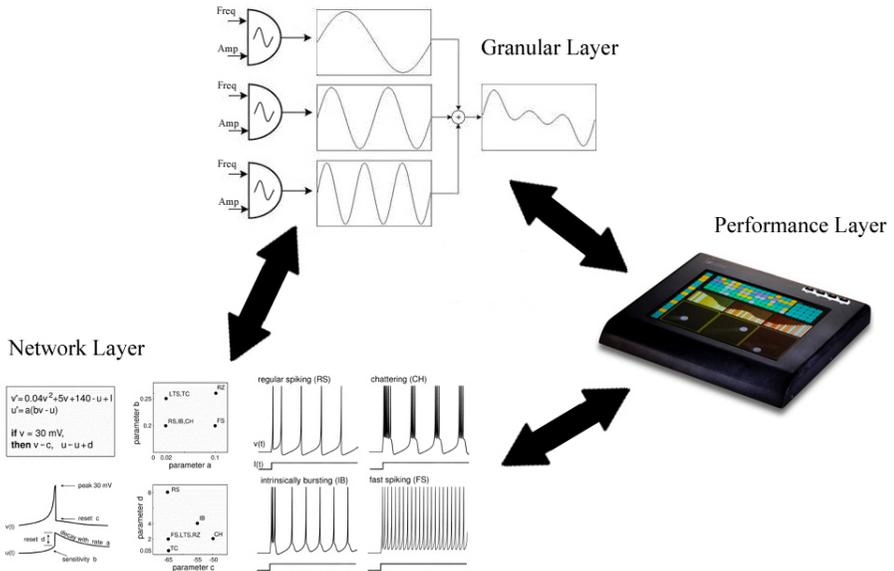


Figure 1: The three layers of the system. The lowest level, the Network Layer is connected to the second level of control, the Granular Layer, which links up the oscillators with the neural networks. The third level, the Performance Layer provides control over the global parameters for the granular synthesiser and the neural firing control.

The Network Layer controls the behaviour of the pulse-coupled network of spiking neurons. This is the lowest level of functionality in our model. The networks firing patterns are independently controlled here.

The Granular Layer connects each neuron in the pulse-coupled network to an individual oscillator. The network is used as a trigger to fire the oscillators when each of the neurons in that network fire. The summed oscillator outputs from each network form the sound granules.

The Performance Layer provides control over the model as an instrument. Five individual firing controls are provided through balls connected to each of the networks. These balls are tracked in a two dimensional space and provide control over the amplitude and scaling frequency variable of each of these firing controls. Each firing control controls one network of spiking neurons so we are essentially a polyphonic granular synthesiser.

3. The Synthesis Engine

Granular synthesis works by generating very small bursts of sound known as grains. These grains are typically between 10 and 35 milliseconds long and together form complex sounds (Figure 2). Our synthesis engine uses additive synthesis to produce these sound grains by summing the outputs from various oscillators.

Granular synthesis was inspired by a sound representation method by Dennis Gabor in the 1940s. Gabor acknowledged that the ear has a time threshold for discerning sound properties. Any sound heard below this threshold is heard as a click..

The ear can not discern the difference between two short sounds that are close together. This can be thought of similarly to the eye when watching a movie, where the speed between images is too short for the eye to discern and a continuous flow of motion is observed.

Similarly the ear has a time threshold for discerning sound properties. The ear naturally performs Fourier analysis on the sounds we hear and then processes the resulting sinusoidal waveforms, but the ear needs several cycles to discern the properties of these waves and so will not discern the difference between sounds if they are too close together.

Single grains are heard as clicks, whilst a collection of grains overlapping one another forms a rich sonic sound sometimes known as a sonic cloud. Denis Gabor argued that any sound could be broken down into a collection of smaller sonic sounds.

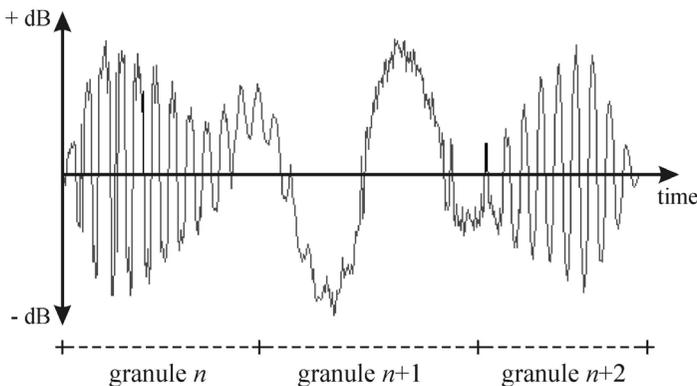


Figure 2: A sequence of three very short sound grains. A rapid succession of thousands of such grains forms larger complex sounds. (Miranda, E. 2005)

Most granular synthesisers used stochastic methods to control the production of the grains, such as probability tables holding waveform parameters which can be called to provide synthesis values for each grain during the synthesis process.

Spiking neural networks present an interesting control mechanism for Granular Synthesis as they exhibit very rich dynamics and the spike frequencies emitted by the neurons are of in the same timescale of those relevant to granular synthesis (i.e. on the level of milliseconds and tens of milliseconds). This makes them very suitable to

use as a triggering mechanism for a granular synthesiser. Another advantage of spiking neural networks is that it provides a good mechanism for controlling the complexity of granular synthesis in real time.

4 Spiking Neural Networks

4.1 Pulse-Coupled Neural Network Overview

A Neuron is essentially an object that sends out an electrical charge once the input voltage has exceeded a certain threshold. The charge sent out by real neurons is known as a spike signal, and has an amplitude in the order of 100mV (millivolts) with the duration of the spike being between 1 and 2 milliseconds. This spike is sent out to all the other (post-synaptic) neurons to which this (pre-synaptic) neuron is connected. It takes a matter of milliseconds for this spike to reach the post-synaptic neurons, and the post-synaptic neurons will in turn fire a spike to all the neurons it is connected to if their current voltage state plus the signal of the spike are above the threshold and so on. Most of this processing takes place in the cortex with each neuron being typically connected to up to 10 000 other neurons. One can quickly see how complicated resulting dynamics would be from simply sending out a single spike to just one neuron.

Figure 3 shows the twelve of the most common forms of mammalian cortical neuron firings. When a neuron receives a spike it updates its connection with all the neurons that it is connected with.

Spiking Neural Networks simulate firing behavior of Neurons in the brain. A neuron sends an electrical charge to all the other neurons to which it is connected to if its threshold reaches a certain limit. The neuron will be inactive for a certain duration while it recovers from sending out a voltage spike. The timing between spikes greatly influences the effect of these Networks.

4.2 Izhikevich's Pulse-Coupled Neural Model

Eugene Izhikevich has developed a simple model of spiking neurons with random connections that can produce realistic organised collective behaviour. The model contains enough detail to produce the rich firing patterns found in cortical neurons (Figure 3), yet it is also computationally very efficient.

The model proposed by Eugene Izhikevich produces rich spiking and bursting behaviour which mimics the known behaviour of cortical neurons. The model consists of two differential equations.

$$\frac{dv_i}{dt} = 0.04v_i^2 + 5v_i + (140 - u_i) + I_i \quad (1)$$

$$\frac{du_i}{dt} = a(bv_i - u_i) \quad (2)$$

Where the first (1) or v represents the membrane potential of the neuron, and the second (2) or u represents the membrane recovery.

There are four variables used in controlling the system; a , b , c , and d . a relates to the recovery variable u , and smaller values result in slower recovery. b describes the responsiveness of the recovery variable u . c is the after-spike reset voltage of the neuron which is typically -65mV . d represents the after-spike reset value of the recovery variable u and is typically equal to 2. Changing the values of these parameters changes the behaviour of the network as seen in Figure 3.

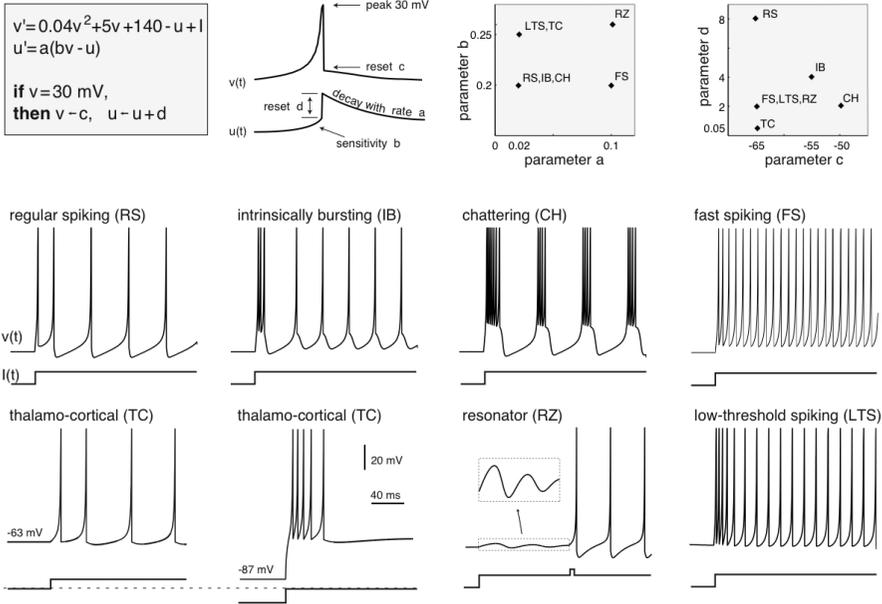


Figure 3: Known types of neurons correspond to different values of the parameters a , b , c and d in Izhikevich's model. (Izhikevich, E. 2005)

The firing patterns of the Izhikevich's model have a collective firing range of between 1 and 40Hz. This is in a similar range as Granular Synthesis which generates sounds between 10 and 50 milliseconds in length, and so presents an interesting case for using the model for controlling a granular synthesiser.

5. Controlling the System

Control over the synthesiser is carried out using a touch-screen controller; the "Lemur". The Lemur is a modular graphical touch-screen interface. It has touch-screen technology that can track multiple fingers simultaneously and the user can create their own graphical interface through modular editing software.

The Lemur uses OSC (Open Sound Control) to communicate over a normal T-100 Ethernet cable. Open sound control communicates using 32-bits and is more efficient than MIDI. Granular synthesis can be computationally demanding as there are a number of parameters that need to be controlled. The Lemur's unique design along

with the advantages of using OSC provides a very good modular interface for controlling a polyphonic granular synthesiser.



Figure 4: The Lemur modular touch-screen controller. You can navigate through any number of screens designed for your interface and track multiple fingers simultaneously.

Figure 5 shows a large area in the middle of the picture (GrainCtl). This is the “Multiball” object window and it allows you to track multiple balls around its area. The balls in this area do not come into existence until the user touches the multi-ball area, and the balls fade out and disappear once they have been released. The manner in which the balls fade in and out is controllable via the “attack” and “release” faders at the top of the screen. The rate at which the balls fade in and out controls a global gain parameter of that network and hence acts as an envelope which shapes the output of the different networks. The Multiball object triggers the output of the grains generated by each network and controls the amplitude and frequency of the grains.

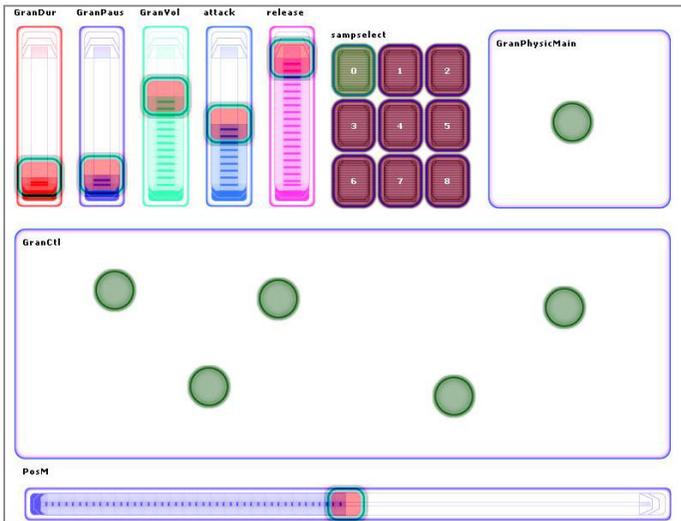


Figure 5: The GUI for triggering of the output for each network and the global parameters for envelopes, grain duration and interval length are controlled here.

The screen on the right of Figure 6 shows how the four parameters used to control the behaviour of the neural network are controlled through the Performance Layer. These parameters are mapped to two two-dimensional areas, with a mapped to the x-axis and b mapped to the y-axis in the one area, and c mapped to the x-axis and d mapped to the y-axis of the second area. This is done for each of the networks to control the firing behaviour of that network and the grains they generate. The other sliders control the other network parameters such as Gain, Frequency, Saptialization and Filter (Filter Frequency, Filter Q and Filter Gain).

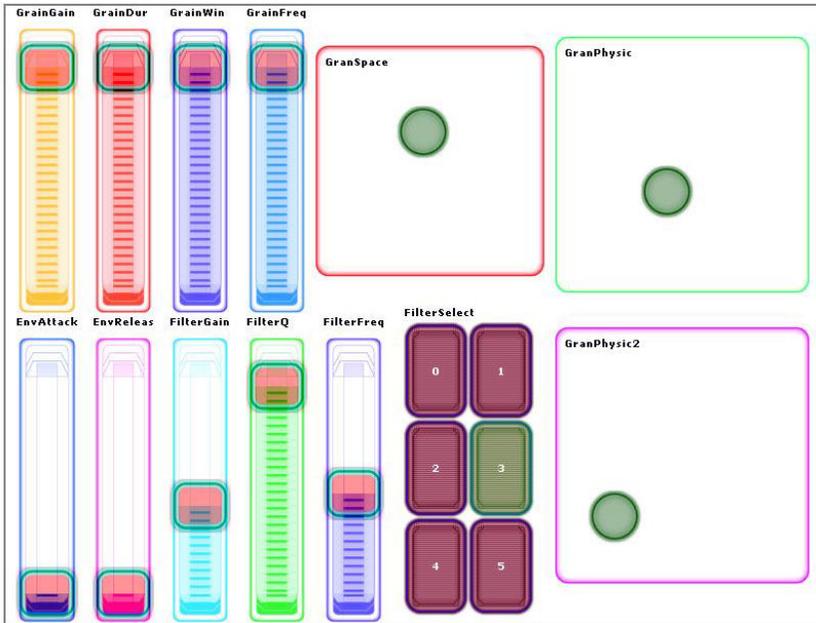


Figure 6: The GUI design for controlling the behaviour of the network of spiking neurons and the rest of the networks Granular Synthesis parameters.

6. Conclusion

The system is still being developed. The control layers for the second and third levels still need further refinement. Nevertheless, we have done some experiments to prove the systems functionality with positive results and found that the system works well, and we have produced some interesting sounds out of it.

The Lemur with its modular interface concept proves to be the perfect device to provide control over our granular synthesiser. Its flexible nature and the dynamics of the controllers from the objects library provide intuitive control over the firing of the grains and the rest of the parameters used in granular synthesis.

Using a pulse-coupled network of spiking artificial neurons to trigger the generation of sound grains proved to be a successful method of controlling the production of the granules.

The technique we have introduced in our model successfully fulfilled the aims of our investigation. The ability to control the temporal behaviour of our networks sub-groups proved to be a new and interesting controllable way of producing the grains used for granular synthesis.

We are currently investigating various settings for the parameters that control the spiking neurons to see if we can find relationships between the behaviour of the spiking neurons and the groups of sounds they produce. These sound categories are intended to be used as presets for our module.

7. References

Izhikevich, E. M. (2003). “A simple model of spiking neurons”, *IEEE Transactions on Neural Networks*, 14:1469.

Miranda, E. R. (1995). “Granular Synthesis of Sounds by Means of a Cellular Automaton”, *Leonardo*, 8(4):297-300.

Miranda, E. R. (2002). *Computer Sound Design: Synthesis Techniques and Programming*. Oxford (UK): Elsevier/Focal Press.

Miranda, E. R. and Matthias, J. (2005) *Granular Sampling Using a Pulse-Coupled Network of Spiking Neurons*. *EvoWorkshops*, (3449):539-544. Springer-Verlag Berlin Heidelberg.

Emotional speech recognition: a neural network pruning approach

S.Saqib, G.Bugmann and E.Miranda

School of Computing, Communications and Electronics, Plymouth, United Kingdom
e-mail: gbugmann@plymouth.ac.uk

Abstract

This paper explores acoustic features used to differentiate between emotions of anger and happiness from the speech signal, considering only the nonverbal information using Neural Networks as the principle classifier and Neural Network Pruning for feature selection/reduction. An acted speech database containing 3512 short (approximately 2-4 seconds) utterances recorded by a female was used to extract 68 acoustic cues for emotion from the speech signal. The features included 34 prosodic (basic and compound) and 34 quality (formant frequency based, harmonicity based and energy band distribution based) features. An over all recognition rate of 74% was obtained using all 68 features without pruning. Neural Network Pruning reduced the original feature set to 46, giving an overall recognition rate of 90% which shows that Neural Network Pruning has proven to be a good feature reduction/selection method to distinguish between happy and angry speech.

1. Introduction

It would be quite useful if a computer were able to recognize what emotion is expressed in a given utterance. Human computer interface, for example could be made to respond differently according to the emotional stats of the user. This could be important in situations where speech is the primary mode of interaction with the machine.

Listeners react to the speaker's emotive state and adapt their behavior depending on what kind of emotion the speaker transmit, e.g. showing empathy to sad people, or trying to help someone who hesitates to clarify what he means or wants. Classification of emotional states on basis of the prosody and voice quality requires classifying acoustic features in the speech as connected to certain emotions. This also implies the assumption that voice alone really carries full information about emotive state by the speaker (Gustafson-Capková, 2001). Most research papers reveal that this assumption can without doubt be taken for granted, but there are still studies that take this assumption questionable (Stibbard, 2001)

Despite the difficulties of achieving emotion recognition through the speech signal there are still attempts to cope with this task. Dellaert, Polzin and Waibel (Dellaert et al. 1996) recorded a corpus of over 1000 utterances from several different speakers, who were asked to read 50 short sentences in four different emotional states (happiness, sadness, anger and fear) plus the normal (neutral) state. Features were extracted per utterance based solely on pitch profile. The three pattern recognition methods used included Maximum Likelihood Bayes classifier (MLB), Kernel Regression (KR) and K-Nearest neighbours (KNN). The lowest recognition error,

performing leave-one-out (LOO) cross-validation was obtained for the KNN method and reaches approximately 36%.

The use of acoustic prosodic cues in order to classify angry vs. neutral speaking style is described in (Huber, 1998). 20 speakers were asked to produce 50 neutral and 50 angry utterances and multi-layer perceptrons were trained with these data. Results reach around 90% of accuracy in the simplified tasks of distinguishing emotional from non-emotional utterances.

Studies have shown that differentiating between anger and happiness has been the hardest. The study conducted by (Yildirim et al. 2004) investigates acoustic properties of speech associated with four different emotions (sadness, anger, happiness, and neutral) intentionally expressed in speech by an actress. The data analyzed in this study were collected from a semi professional actress and consist of 112 unique sentences uttered with four emotions, i.e., angry, happy, sad, and neutral.

Their aim was to obtain detailed acoustic knowledge on how speech is modulated when speaker's emotion changes from neutral to a certain emotional state. Fisher's linear discriminant analysis was performed to see how effectively acoustic cues could be used to discriminate emotions. Their results are summarized in the confusion matrix shown in table 1. This table shows the number of utterances recognized and confused out of one category of emotion containing 112 utterances. For example when 112 happy samples were tested, 68 were classified as happy 31 as angry, 1 as sad and so on. The values in bold shows that 31 of 112 happy samples were falsely recognised as angry, while 42 of 112 angry examples were recognised as happy. Happy and angry have higher confusion rates compared to all other emotional categories in the study.

Recognized As>	Neutral	Sad	Angry	Happy
Neutral	90	19	1	2
Sad	23	82	2	5
Angry	6	4	60	42
Happy	12	1	31	68

Table 1: Confusion matrix for five emotions: sadness, boredom happy, angry and neutral. Adapted from Yildirim et al. (2004)

Emotion recognition in this work consists of four stages: Acquiring of an emotional database, feature extraction, feature selection and feature classification. Section 2 describes the database used in recognizing between emotions of anger and happiness. In section 3 features extracted are briefly described. Section 4 explains the feature reduction/selection and classification methodology. Results of this work are discussed in section 5 and finally the paper concludes with some final conclusions

2. Emotional Database used in this study

Emotional database used in study was obtained personally from the University of Edinburgh. This database was used by Gregor O. Hofer in the fulfilment of his Master of Science Thesis "Emotional Speech Synthesis" in the year 2004. The

original database consisted of 400 long recordings in a female voice of happy, angry and neutral sentences each. Average duration of each sentence was 10 seconds. A female speaker with acting training was used to record the sentences. The script used was taken from magazines and newspapers. Each sentence was recorded portraying anger, happiness and without any emotions (neutral). All 400 recordings were done in a soundproof recording booth.

There are several advantages in using an actor to portray emotions. The recording sessions with the actress were a lot quicker and her session waveforms sounded a lot more “professional”.

In speech recognition emotions are considered as annoying noise that decreases the accuracy of recognition. Although it is true that some words and phrases are correlated with particular emotions, the situation usually is much more complex and the same word or phrase can express the whole spectrum of emotions. Since no linguistic information is used to recognize emotions, this leads to deal only with global acoustic features, computed for a whole utterance or command. This seems to be in the favor of many recent studies (Dellaert et al. 1996; Petrushin, 2000)

The same emotional database is used in this study after splitting each long sentence into short parts consisting of short phrases and even one word at some instances. The average duration of utterance in the modified database ranges from 2 to 4 seconds. PRAAT, a shareware program developed by Dr. Paul Boersma of the University of Amsterdam was used for this purpose. This yielded a total of 1756 angry and 1765 happy utterances for single female speaker.

3. Feature Extraction

A total of 68 global acoustic features were computed for a whole utterance using scripts written in the PRAAT software. 34 prosodic features were extracted: 14 basic prosodic, 7 energy and 7 frequency based; 12 speech rate, voiced unvoiced information based; 8 compound prosodic, associated with changes in the speech signal. 34 quality based features were extracted, 7 harmonicity, 8 energy and 19 format frequency based.

4. Feature Selection and classification

Neural networks were used as principal classifier during the project. Neural networks are able to eliminate redundant and irrelevant features when there is enough training data. A network that is too big for a particular classification task is more likely to over fit the training data and have poor performance on unseen examples (i.e., poor generalization) than a small network. Therefore, a heuristic to obtain good generalization is to use the smallest network that will learn to classify correctly the training data by eliminating all those nodes (features) and/or links that are not relevant for the classification task. This process is called Pruning. Reducing the number of irrelevant/redundant features drastically reduces the running time of a learning algorithm and improves generalization. Pruning also provided the following benefits:

- The price of hardware implementation is directly related to number of chip and therefore the network size. The cost of a net can be reduced by finding a minimal network topology (think of runtime, memory and cost for hardware implementation).
- The generalization can (but need not) be improved.
- Unnecessary input units can be pruned in order to give evidence of the relevance of input values.
- Software implementation for a minimal network topology is less than an oversized one.

For the experiments carried out during this study, the *Skeletonization* algorithm was employed for pruning. Skeletonization (Smolensky and Mozer, 1989) prunes units by estimating the change of the error function when the unit is removed. For each node, the attentional strength is introduced which leads to a different formula for the net input.

Feedforward Neural Networks with one hidden layer, 10 hidden nodes for some experiments, 20 for some and 15 for one, two output nodes, *logistic activation function* and *Standard Backpropagation* learning algorithms were used. The software employed is the SNNS (Stuttgart Neural Network Simulator), a simulator for neural networks on Unix workstations developed at the Institute for Parallel and Distributed High Performance Systems (IPVR) at the University of Stuttgart

The *402040* analyzing function was used decide the meaning of the output vector. 402040 stands for the '402040' rule which means on a range from 0 to 1, h will be 0.6 (upper 40%) and l will be 0.4 (lower 40%). The middle 20% is represented by h – l.

A pattern is classified correctly if:

- the output of exactly one output unit is $\geq h$.
- the 'teaching output' of this unit is the maximum teaching output (> 0) of the pattern.
- the output of all other output units is $\leq l$

A pattern is classified incorrectly if:

- the output of exactly one output unit is $\geq h$.
- the 'teaching output' of this unit is NOT the maximum 'teaching output' of the pattern or there is no 'teaching output' > 0 .
- the output of all other units is $\leq l$.

A pattern is unclassified in all other cases. Default values are: $l = 0.4$ $h = 0.6$, which were used in this study.

Six experiments were performed to study the role of pruning of acoustic features to recognize between emotions of anger and happiness in the speech signal. In the first experiment basic prosodic features, in the second experiment, basic prosodic features combined with compound prosodic features, in the third experiment formant frequency based features and in the fourth experiment energy based features combined with harmonicity based features were used as inputs to neural networks to study the

role of each feature set independently on the overall recognition. A reduced set of feature was obtained using pruning in each experiment. The combination of all reduced features after four experiments were taken as the final feature set which was used for recognizing between angry and happy speech in the fifth experiment. A final experiment was conducted by using all 68 features, without pruning, to compare the overall recognition rate with that of experiment five.

5. Results

	Hidden Nodes	Inputs	Inputs Removed	Inputs Retained	Average Over all Recognition	Avg. Un-recognised
Exp No. 1	10	14	8	6	77%	1.5%
Exp No. 2	10	34	11	23	87%	0.00%
Exp No. 3	20	19	10	9	83%	0.45%
Exp No. 4	10	15	7	8	77%	0.19%

Table 2: Results for experiment No. 1, 2, 3 and 4

The first experiment was conducted using 14 basic prosodic (7 energy and 7 frequency based) features as inputs to the Neural Network. Approximately 70% of the patterns were used for training (2546 patterns), 15% for evaluation (526 patterns) and 15% for testing (528 patterns). The second experiment was conducted using 34 prosodic features (14 basic prosodic, 8 compound prosodic and 12 speech rate) features as inputs to the Neural Network. Approximately 70% of the patterns were used for training (2546 patterns), 15% for evaluation (526 patterns) and 15% for testing (526 patterns). The third experiment was conducted using 19 formant frequency based quality features as inputs to the Neural Network. Approximately 70% of the patterns are used for training (2546 patterns), 30% for testing (1052 patterns). The fourth experiment was conducted using 15 quality (8 energy based and 7 harmonicity based) features as inputs to the Neural Network. Approximately 70% of the patterns are used for training (2546 patterns), 30% for testing (1052 patterns). The number of hidden nodes, number of inputs removed (pruned) and retained, average overall recognition (average of angry and happy speech recognized), and average unrecognized speech (average speech unrecognized as angry or happy) for the above four experiments are given in table 2. Confusion matrixes are given in table 3.

The fifth experiment was performed with the 46 reduced features that were obtained from the first four experiments. These features were once again be pruned to finally check whether it the best feature subset of the original 68 features. An over all recognition of 90% was achieved was approximately 70% of the patterns were used for training (2546 patterns), and 30% for testing (1052 patterns). Results are shown in table 4.

	Recognized As>	Angry	Happy	Un-recognised
Exp No. 1	Angry	77.35%	20.89%	1.76%
	Happy	21.00%	77.62	1.38%
Exp No. 2	Angry	88.62%	10.38%	0.00%
	Happy	15.00%	85.00%	0.00%
Exp No. 3	Angry	78.87%	20.79%	0.34%
	Happy	10.39%	89.05%	0.56%
Exp No. 4	Angry	79.62%	29.19%	0.19%
	Happy	25%	75%	0.00%

Table 3: Confusion matrix for experiment No. 1, 2, 3 and 4

In the final experiment all 68 input features were used without any pruning algorithm. An over all recognition of 74% was achieved, which was quite less than the recognition rate achieved by 46 input reduced features. Results can be seen in table 5. Confusion matrixes for experiment five and six are given in tables 6.

Hidden Nodes	Inputs	Inputs Removed	Inputs Retained	Average Over all Recognition	Avg. Un-recognised
20	46	0	46	90%	0.00%

Table 4: Results for experiment No. 5

Hidden Nodes	Inputs	Average Over all Recognition	Avg. Un-recognised
15	68	74%	26%

Table 5: Results for experiment No. 6

	Recognized As>	Angry	Happy	Un-recognised
Exp No. 5	Angry	90%	10%	0.00%
	Happy	10.76%	89.24%	0.00%
Exp No. 6	Angry	66.92%	1.14	31.94%
	Happy	10.76%	89.24%	17.87%

Table 6: Confusion matrix for experiment No. 5 and 6

Statistics using a combination of reduced features from all previous experiments reveal that this 46 feature set is the best subset of the original 68 that can be used to differentiate between the emotions of happiness and anger. After pruning no further feature is removed that further substantiates the validity of this reduced/selected feature set.

6. Conclusions

Results show that Neural Network Pruning has proved to be a good feature selection/reduction method to distinguish between the emotions of anger and happiness. Recognition rates for all 34 prosodic features used are higher than recognition rates of 14 acoustics features. This shows that emotional space increases recognition rates and reduces the number of unclassified patterns. One reason of this might be a sound example portraying none of the emotional categories(hot anger

instead of anger, or excitement instead of happiness) is not recorded accurately by the actor. Databases containing such acted speech may suffer from this problem. Un-classification rate is much lower than compared to when prosodic features are used. Furthermore degree of unrecognized speech improves with increased features .

In the experiment done with 14 basic prosodic features, only one of the 8 compound prosodic feature was eliminated. This shows that pitch variation related features prove to be more effective in recognising between angry and happy speech.

Happy emotion tends to raise both formant frequencies. A happy expression usually comes together with a slightly smiling, which causes lips movements and force the rounded vowels to sound unrounded. This can be seen as recognition rate for happy emotion is greater than angry emotions when formant frequency based features for classification between angry and happy speech.

Results using a combination of reduced features from all previous experiments show that this 46 feature set is the best subset of the original 68 which can be considered to used to differentiate between the emotions of happiness and anger for a female speaker. After pruning no further feature is removed that further substantiates the validity of this reduced/selected feature set.

This feature set can be a bench mark for distinguishing angry speech from happy or neutral from the short utterances typical of Interactive Voice Response (IVR) applications such as call centres in which an angry customer can be identified immediately and his call can be diverted to some experienced operator.

7. References

Alter, K.; Rank, E.; Kotz, S.A.; Toepel, U.; Besson, M.; Schirmer, A.; Friederici, A.D.: *Accentuation and emotions – Two different systems?* In ICSA Workshop on Speech and Emotion. Belfast, 2000.

Alter K.; Rank E.; Kotz S.A.; Pfeifer E.; Besson M.; Friederici A.D.; Matiasek J.: *On the relations of semantic and acoustic properties of emotions*. In Proceedings of the 14th International Conference of Phonetic Sciences (ICPhS-99), San Francisco, California, p.2121, 1999.

Dellaert, F.; Polzin, T.; Waibel, A.: *Recognizing Emotion in Speech* ICSLP'96 Conference Proceedings, Delaware. 1996.

Gustafson-Capková, S.: *Emotions in Speech: Tagset and Acoustic Correlates*. Speechtechnology, termpaper. Autumn 2001

Huber, R.; Nöth, E.; Batliner, A.; Buckow, J.; Warnke, V.; Niemann, H.: *You BEEP Machine – Emotion in Automatic Speech Understanding Systems*". TSD'98, Brno, Masaryc University.

Klasmeyer, G.: *The Perceptual Importance of Selected Voice Quality Parameters* in Proceedings of ICASSP'97, Munich, Germany, 1997.

Petrushin, V.A.: *Emotion in speech: Recognition and Application to Call Centers*". Artificial Neu. Net. In Engr. (ANNIE'99), pp. 7-10, Nov. 1999.

Saqib, S.: *Emotional Speech Recognition: A Neural Network Pruning Approach*, Project Report, School of Computing, Communications and Electronics, University of Plymouth, United Kingdom.

Smolensky, P; Mozer M,: Skeletonization: A Technique for Trimming the Fat from a Network via Relevance Assessment. In D. S. Touretzky, editor, *Advances in Neural Information Processing Systems (NIPS) 1*, pages 107--115, San Mateo, 1989. Morgan Kaufmann Publishers Inc.

Stibbard, R. M.: *Vocal Expression of Emotions in Non-laboratory Speech: An Investigation of the Reading/Leeds Emotion in Speech Project Annotation Data*. Unpublished PhD thesis. University of Reading, UK. 2001.

Yildirim, S.: Bulut, M., Lee, C. M., Kazemzadeh, A., Deng, Z., Lee, S.,/Narayanan, S., Busso,C., (2004):"An acoustic study of emotions expressed in speech", In *INTERSPEECH-2004*, 2193-219

Author Index

Al-Ghatam	157	Marston	89
Al-Tawqi	164	Mcilree	273
Asiwe	3	Michalopoulos	205
Aung	173	Miranda	233, 278, 286
		Mohyuddin	100
Belpaeme	264	Mued	130, 138
Bryant	182	Murray	278
Bugmann	286		
		Nwobodo	106
Chatziapostolou	11		
Clarke	60, 89, 205, 213	Papadaki	191
		Phippen	182, 199, 247, 273
Davy	19		
Denham	255	Rizvi	114
Dowland	3, 100, 114 ,146, 157, 173		
Drouet	233	Saqib	286
Edmonds	30	Tjhai IC	130
Edwan	38	Tjhai GC	222
Evangelatos	51		
		Vahedi-Sarrigani	123
Ford G	247		
Ford CC	255	Wang	38, 72, 80, 106, 123
Furnell	11, 30, 51, 164, 182, 191, 222		
		Yang	138
Ghita	19, 38		
		Zhang	146
Harewood-Gill	264		
Jaeger	60		
Jayakumar	199		
Jing	72		
Karakasiliotis	191		
Karatzouni	213		
Khan	80		

Distributor:

Network Research Group
University of Plymouth
Drake Circus
Plymouth
PL4 8AA
United Kingdom



<http://www.network-research-group.org>