EXPLORATION OF EMAIL SPAM, WITH A FOCUS ON ITS EFFECTS AND MITIGATION IN SAUDI ARABIA

By

Hasan Shojaa Alkahtani

School of Computer Science, Engineering and Mathematics

Faculty of Science and Engineering

2014

A thesis presented to the

Flinders University of South Australia

In fulfilment of the requirements for the degree of

Doctor of Philosophy

Adelaide, South Australia, 2014

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ABS	Australian Bureau of Statistics
Acc	Accuracy
ACMA	Australian Communications and Media Authority
Ads	Advertisements
ADTree	Alternating Decision Tree
AIS	Artificial Immune System
ANN	Artificial Neural Network
AOL	America Online
ASRG	Anti-SPAM Research Group
AT&T	American Telephone and Telegraph
B	Bachelor Degree
BART	Bayesian Additive Regression Trees
BGP	Border Gateway Protocol
BL	Black Lists
BMF	Bayesian Mail Filter
BP	Back Propagation
C	Central Region
CAPTCHA	Completely Automated Public Turing test to tell Computers and Humans Apart
CART	Classification and Regression Trees
CAUCE	Coalition against Unsolicited Commercial Email
CC	Clustering Coefficient
CDT	Centre for Democracy and Technology
CERT	Computer Emergency Response Team
CHI	Chi-Square
CIA	Communication Interaction Average

CITC	Communication and Information Technology Commission
CNNIC	China Internet Network Information Centre
COEIA	Centre of Excellence in Information Assurance
CR	Communication Reciprocity
CRS	Challenge Response System
CS	Consulting Services
CS&IT	Computer Science and Information Technology
D	Diploma Degree
DCC	Distributed Checksum Clearinghouse
DF	Document Frequency
DMA	Direct Marketing Association
DNS	Domain Name System
DNSBLs	Domain Name System Black Lists
DoS	Denial of Service
DSDBL	Domain Specific Dynamic Black List
DT	Decision Trees
E	Eastern Region
E&T	Education and Teaching
ECT	Electronic Communications and Transactions
EML	Employees
ENISA	European Network and Information Security Agency
EP	Educational Positions
EPCCI	Eastern Province Chamber of Commerce and Industry
ERP	Enterprise Resource Planning
Err	Error rate
ESP	Email Service Provider
EU	European Union

F&I	Finance and Investment
FAR	False Acceptance Rate
FB	Flexible Bayes
FN	False Negative
FP	False Positive
FTC	Federal Trade Commission
GA	Genetic Algorithm
GCC	Gulf Cooperation Council
GDP	Gross Domestic Product
GIF	Graphics Interchange Format
HS	High School
HS&M	Health Sciences and Medicine
HTML	Hyper Text Markup Language
HTTP	Hyper Text Transfer Protocol
ICTs	Information Communication Technologies
ID	Identification number
ID3	Iterative Dichotomiser 3
IDA	Infocomm Development Authority
IDC	International Data Corporation
IDS	Intrusion Detection System
IFC	International Finance Corporation
IFCC	Internet Fraud Complaint Centre
IG	Information Gain
IP	Internet Protocol
IRTF	Internet Research Task Force
ISC	Internet Society of China
ISP	Internet Service Provider

IT	Information Technology
JPEG	Joint Photographic Experts Group
KACST	King Abdulaziz City for Science and Technology
K-NN	K-Nearest Neighbours
LAN	Local Area Network
LINX	London Internet Exchange
LM	Learning Module
LP	Legitimate Precision
LR	Legitimate Recall
LRC	Logistic Regression Classifier
LVQ	Learning Vector Quantizers
M	Master Degree
MBL	Memory Based Learning
MED	Ministry of Economic Development in New Zealand
MEM	Maximum Entropy Models
MI	Mutual Information
ML	Machine Learning
MLP	Multi-Layer Perceptron
MM	Morphology Module
MN Boolean NB	Multinomial Boolean Naive Bayes
MN TF NB	Multinomial Term Frequency Naive Bayes
MP	Medical Positions
MTP	Management Positions
MV Bernoulli NB	Multivariate Bernoulli Naive Bayes
MV Gauss NB	Multivariate Gauss Naive Bayes
N	Northern Region
NB	Naive Bayes

NDRC	National Development and Reform Commission
OECD	Organisation for Economic Cooperation and Development
P&BS	Physical and Biological Sciences
P&M	Production and Manufacturing
PDF	Portable Document Format
PhD	Doctor of Philosophy
PMC IT	Prince Muqrin Chair for Information security Technologies
POP	Post Office Protocol
RBL	Realtime Blackhole Lists
REPTree	Reduced Error Pruning Tree
RF	Random Forests
RFID	Radio Frequency Identification
RIPPER	Repeated Incremental Pruning to Produce Error Reduction
S	Southern Region
S/MIME	Secure/Multipurpose Internet Mail Extensions
SA	SPAM Assassin
SAGIA	Saudi Arabian General Investment Authority
SAR	Saudi Riyal
SBK	Search Based Keyword
SBREC	Social and Behavioural Research Ethics Committee
SE	Self-Employed
SMEDC	Small and Medium Enterprises Development Centre
SMS	Short Message Service
SMTP	Simple Mail Transfer Protocol
SP	SPAM Precision
SPASTIC	Simple Procmail Anti-Spam Templates (Improved Code)
SPF	Sender Policy Framework

SR	SPAM Recall
SS	Social Sciences
STD	Student
SVM	Support Vector Machine
T&T	Technology and Telecommunication
TCP	Transmission Control Protocol
TCR	Total Cost Ratio
TEOS	Trusted Email Open Standard
TF-IDF	Term Frequency-Inverse Document Frequency
TP	Technical Positions
TREC	Text REtrieval Conference
UBE	Unsolicited Bulk Email
UCE	Unsolicited Commercial Email
URL	Uniform Resource Locator
UUNET	Unix to Unix Network
V	Variance
VPN	Virtual Private Network
W	Western Region
WA	White Lists
$\mathbf{W_{ACC}}$	Weight Accuracy
$\mathbf{W_{Err}}$	Weighted Error rate
WL	White Lists

Abstract

Introduction

Email spam is an international issue that has caused many challenges in different countries. In Saudi Arabia, the volume of email spam is high compared to other countries. This research investigated the nature of email spam in Saudi Arabia and the awareness of email users about it and efforts to combat it; and provided suggestions for strategies mitigate it. The study was conducted among three groups in Saudi Arabia: public users, businesses and ISPs.

Methodology

This research adopted a quantitative approach, using self-administrated questionnaires to collect data. In this descriptive and cross-sectional study, data was collected to answer the research questions from February 2011 to July 2011. Multiple cluster random sampling was used to select public users and businesses, and convenience sampling was used to select ISPs. A total of 1,500 public users from universities, schools, hospitals, and government departments, and 300 businesses were selected randomly from five regions; and all 27 ISPs. The validity of the questionnaires was examined through a pilot study.

During data collection, public users, businesses and ISPs were asked to forward Arabic and English email spam that they received in their email inboxes (i.e. email spam that was bypassed anti-spam filters) to a specific email address created for the purpose of this research. An email spam corpora was collected to investigate the tricks used in the Arabic and English spam to bypass filters, affecting their effectiveness. A total of 1,270 email SPAMs were analysed: 1,035 Arabic, 179 English, and 56 mixed Arabic and English spam. A taxonomy of email spam filters (mostly developed to detect English spam) was constructed to develop methods to counter the tricks used in Arabic spam. Using a phenetics approach, filters were classified according to similarity between the methods used to detect spam. Statistical tests such as chi-square and independent-samples t-test were used to analyse the data.

Results

Email users in Saudi Arabia had limited awareness of spam and ways to combat it, although a large portion of them were well-educated professionals. ISPs, businesses and public users believed that most of the spam was written in English, followed by a large minority in Arabic. The most common types of Arabic spam were related to forums, and religion and politics; and most English spam was pornographic, and phishing and fraud emails. Saudi Arabia was the greatest source of Arabic spam; whereas most of the English spam was sent from non-Arabic countries.

ISPs indicated that anti-spam filters were not completely effective, and these filters performed better in detecting English spam than Arabic spam. The highest percentage of Arabic spam originated from Saudi Arabia. Different tricks were used in Arabic and English spam to bypass the filters. More Arabic than English spam included attractive words in the subject line, contained an image in the body of the message, and was sent by obfuscated or fake email addresses. Malicious contents (e.g. viruses) appeared more often in English spam than Arabic spam.

The greatest effect of email spam on the performance of public users and organisations in Saudi Arabia was reduced productivity, which can affect the country's economic growth.

Conclusion

More work is needed to combat spam in Saudi Arabia. Recommended strategies for government and ISPs to reduce its effects in Saudi Arabia are: adopt an agreed definition; enact culturally fit anti-spam laws; investigate effective ways to educate email users; and refine and develop more effective filters, especially for Arabic spam.

Certification

I certify that this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

As requested under Clause 14 of Appendix D of the Flinders University Research Higher Degree Student Information Manual I hereby agree to waive the conditions referred to in Clause 13(b) and (c), and thus

- Flinders University may lend this thesis to other institutions or individuals for the purpose of scholarly research;

- Flinders University may reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Signed Dated

Hasan Shojaa Alkahtani

Acknowledgements

First and foremost, I would like to thank my God (Allah) Almighty for giving me patience, opportunities, guidance, and commitment to complete this research. This research would not have been carried out without his bounties and blessings.

I would like to express my great gratitude and sincere thanks to my Principal Supervisor, Dr. Robert Goodwin, for his patient guidance, persistent help, careful reading and feedback on my thesis, giving me the motivation that helped to pass the difficulties of this research, and encouragement all the way through my PhD research.

I would also like to sincerely thank my Co-supervisor, Dr. Paul Gardner-Stephen, for his guidance and advice, smart and sharp discussion about the research that provided me insightful ideas, constructive comments on research chapters and our joint papers, and cooperation on many issues that faced me at every stage of my research.

I am sincerely grateful to my colleague at Flinders University, Dr. Masoud Mohammadnezhad, for his expert guidance, valuable comments on study surveys and data analysis, and providing excellent ideas and consultation about writing this thesis.

I wish to thank the Head of the Transition office at Flinders University, Associate Professor Salah Kutieleh, for his review and support in translating research surveys from English to Arabic and vice versa, and excellent comments on wording of survey questions.

I would like to express my thanks and appreciation to the General Director of the Palestinian National Internet Naming Authority (PNINA), Mr. Marwan Radwan, for his support and assistance in providing some domain names that have been used to achieve purposes of this research.

I would also like to express my thanks and appreciation to the Executive Director of the SPSNET ISP, Dr. Fahad Al Hindi, and the Sales and IT Manager, Mr. Kassem Diab, for their assistance, and feedback and comments on the ISP survey.

I wish to express my most sincere gratitude for Staff of the School of Computer Science, Engineering and Mathematics at Flinders University for their assistance, and for always being friendly.

My warmest thanks go to my family for their support, and encouragement. Especially, I am indebted to my parents for their support, advice, prayers, encouragement, giving me the motivation, and confidence to complete this research. I would like to express my special thanks to them for being always proud of me. I wish to sincerely thank my dearest brothers and sisters for their support, encouragement, and concern throughout my study.

Last but not least, I would like to express my heartfelt thanks to my beloved wife, Mrs. Jamellah, for her patience, unconditional support, and encouragement me during the period of my PhD study in Australia. Without her support and love, this research would not have been possible. I would also like to thank my beloved sons, Talal and Fahad, and my beautiful daughter Wed for giving me happiness throughout the period of the stay and study in Australia.

I dedicate this thesis to my loving parents, my beloved wife and sons, and my dearest brothers and sisters for their love, support, prayers, encouragement at every stage of my life, and motivation to achieve my goals in the life. I wish that granting my PhD could paint joy and happiness on their faces.

Publications

The following publications were results of researches conducted as part of the preparation of this thesis.

1. A research paper titled "*A Taxonomy of Email SPAM Filters*" was presented at the 12th International Arab Conference on Information Technology, Riyadh, Saudi Arabia December 11-14, 2011; and published in proceedings of the International Arab Conference on Information Technology (ACIT2011) (Alkahtani, Gardner-Stephen & Goodwin 2011).

2. A research paper titled "*Email SPAM Related Issues and Methods of Controlling Used by ISPs in Saudi Arabia*" was presented at the 12th International Arab Conference on Information Technology, Riyadh, Saudi Arabia December 11-14, 2011; and published in proceedings of the International Arab Conference on Information Technology (ACIT2011) (Alkahtani, Goodwin & Gardner-Stephen 2011).

3. A research paper titled "*Email SPAM in Saudi Arabia and How do End Users Deal with it?*" was presented at the International Conference on Computing, Networking and Digital Technologies, Sanad, Bahrain November 11-13, 2012; and published in proceedings of the International Conference on Computing, Networking and Digital Technologies (ICCNDT2012) (Alkahtani, Gardner-Stephen & Goodwin 2012).

4. A research paper titled "*A Comparative Study of the Perceptions of End Users in the Eastern, Western, Central, Southern and Northern Regions of Saudi Arabia about Email SPAM and Dealing with it*" was published in the International Journal of Cyber-Security and Digital Forensics (IJCSDF2012) (Alkahtani, Goodwin & Gardner-Stephen 2012b).

5. A research paper titled "*Combating of Email SPAM by Different Businesses in Saudi Arabia*" was presented at the 13th International Arab Conference on Information Technology, Zarq, Jordan December 10-13, 2012; and published in

proceedings of the International Arab Conference on Information Technology (ACIT2012) (Alkahtani, Goodwin & Gardner-Stephen 2012a).

6. A research paper titled "*The Key Findings of Surveys Related to Email SPAM and Methods of Combating it in Saudi Arabia*" was published in the International Journal of Emerging Technology and Advanced Engineering (IJETAE2013) (Alkahtani, Goodwin & Gardner-Stephen 2013).

Chapter 1: Introduction

The first chapter of the thesis presents an overview of the research. This chapter is divided into seven sections as follows:

- Section 1.1: introduces the background of the research problem.
- Section 1.2: describes the scope of the research.
- Section 1.3: provides the aim and objectives of the research.
- Section 1.4: presents the research questions.
- Section 1.5: outlines the research methodology.
- Section 1.6: provides the contribution of the research.
- Section 1.7: describes the structure of the thesis.

1.1 Background to the Research Problem

Email spam is an international problem that causes many challenges in different countries. A number of studies have been conducted to investigate email spam and to provide effective educational, legal and technical suggestions to combat it and its effects (Al-A'ali 2007; Bujang & Hussin 2010; Chigona et al. 2005; Grimes, Hough & Signorella 2007; Leng 2006; Pallas & Patrikakis 2005), but a significant concern uncovered when studying the issues related to email spam is its definition. A literature review revealed that there is no specific universal definition for email spam. Some studies defined it as unsolicited bulk email (UBE); that is, the sending a large number of emails that are not requested by recipients (Cook et al. 2006; Polz & Gansterer 2009). Some defined it as unsolicited commercial email (UCE). This definition comprises promotional advertisements from different businesses sent to a large number of recipients (Boykin & Roychowdhury 2004; Carreras & Marquez 2001; Sakkis et al. 2003). The Centre for Democracy and Technology (CDT) in the United States has defined email spam as junk mail that includes jokes and chain letters from businesses, friends and family (Center for Democracy & Technology 2003). The definitions of previous studies of email spam are discussed and the definition adopted for the purpose of this study is been provided in Chapter 2.

Email spam has a negative impact on the performance of email users and organisations, and on the growth of the economy in different countries. Ferris Research has estimated that the cost of spam for companies around the world in 2004 was about US$14 per user per month in lost productivity (Everett 2004). The Singapore Infocomm Development Authority (IDA) has reported that the total cost of spam for consumers was about S$23 million in lost productivity each year (Leng 2006). Garcia, Hoepman and Nieuwenhuizen (2004) stated that email spam consumes the resources of email servers and this costs Internet Service Providers (ISPs) a lot of money as they have to increase the capacity of their email systems and buy extra bandwidth. The amount of Gross Domestic Product (GDP) loss caused by processing email spam in Japan was about 500 billion yen a year (Takemura & Ebara 2008).

Email spam can be a way to transfer malicious programs such as viruses, worms, trojans, spyware, fraud and phishing to users' computers and organisations systems (Cournane & Hunt 2004; Moustakas, Ranganathan & Duquenoy 2005; Pallas & Patrikakis 2005). Some of the malicious programs aim to steal users' identities and confidential information, such as credit card details (Hershkop & Stolfo 2004), and some aim to crash users' software and hardware (Hayati & Potdar 2008). The US law enforcement officials, federal agencies and experts at McAfee Corporation state that email spam is a way to steal the identities and money of unsuspecting consumers. The Internet Fraud Complaint Centre (IFCC) estimated that the cost to consumers of online fraud was $17.8 million in 2001, and the estimated number of Americans who are victims to identity theft is 500,000 –700,000 each year (Hinde 2002).

Many ways have been developed to combat email spam, including educational, legal and technical efforts (Chigona et al. 2005; Lugaresi 2004). The educational efforts are one of the important strategies for reducing email spam. According to Dantin and Paynter (2005), "an important step towards minimising unsolicited and unwanted emails is raising the awareness of employees". Some countries, such as USA and the member states of the European Union (EU), have taken action to increase awareness of email spam. Pfleeger and Bloom (2005) reported that "the US Federal Trade Commission (FTC) organisation has made the public more aware of the growing spam problem”. An example of the actions taken by the member states of the EU was conducting campaigns to make users aware of spam and the appropriate ways to deal with it (Pfleeger & Bloom 2005). Many studies have investigated users’ awareness of email spam and ways of combatting it. Bujang and Hussin (2010) studied the awareness of Malaysian email users and how they deal with it, and found that although the Malaysian people were aware of email spam, they did not know how to protect their email addresses. A study conducted in Singapore has shown that forty-two per cent of users did not know how to protect themselves from email spam (Leng 2006). Another study conducted in Bahrain revealed that seventy-four per cent of the participants did not know about anti-spam filters (Al-A'ali 2007).

Legislation is another way to combat email spam. According to Lev and Goldin (2006), “The spam to legitimate email ratio in Japan is much lower than average due to the strict attitude towards law enforcement”. Some countries have enacted special laws against spam to reduce its volume and effects. Examples of these countries include the USA, European Union (EU) member countries, Australia, and some Asian countries, such as Japan and Singapore (Sorkin 2009). However, none of the Arabic countries have enacted laws to combat spam (Al-A'ali 2007).

Another way to combat email spam is by the use of technical controls. Experts who work in the field of information and network security have undertaken much research and many projects to develop techniques to combat it, including origin-based techniques (Garcia, Hoepman & Nieuwenhuizen 2004) and content-based techniques (Cook et al. 2006). In addition, many ISPs and businesses have instituted technical policies to reduce spam. According to Sorkin (2001), some ISPs have applied clear policies that do not allow using their facilities to send email spam (Sorkin 2001).

In Saudi Arabia, email spam is a big issue and its volume is high compared to other countries. A report released by Symantec in 2012 indicated that Saudi Arabia remains the most spammed country in the world, with a spam rate of 75.5% of the total email traffic in the country (Symantec Corporation 2012). A spam-relaying countries report for the third quarter of 2012 published by information technology (IT) security and data protection firm Sophos indicated that Saudi Arabia was at the top of the list of spam-relaying countries in the Middle Eastern region (Sophos Ltd

2012). As well, Kaspersky reported that in 2011 Saudi Arabia was the largest source of spam in the Gulf Cooperation Council (GCC) countries (Kaspersky Lab 2012), with the potential for many repercussions in Saudi Arabia, including for the country's economic growth. This requires further efforts by Saudi Government or relevant agencies to combat it. In the literature, little or no evidence could be found of previous studies that have investigated the nature of email spam and its characteristics in Saudi Arabia, and this warrants a study on this issue. This thesis presents the results of such a study.

The discussion above concludes that, although studies have investigated email spam in many countries, assessed its effects and provided appropriate ways to combat it, no previous studies have been found that investigate spam issues and its effects in Saudi Arabia. The study reported in this thesis is the first. The increase in the volume of email spam could be attributed to the lack of awareness of it and the lack of legal efforts and technical measures to combat it (Dantin & Paynter 2005).

This study investigates email spam amongst three different user groups in Saudi Arabia – public users, businesses and ISPs. This focus was prompted by the researcher's understanding the nature of email spam. It addresses the following gaps in the literature on research into email spam in Saudi Arabia:

- awareness, definitions and characteristics of email spam
- the differences between Arabic and English email spam
- legal efforts to combat email spam as perceived by three groups
- technical efforts to combat email spam in Saudi Arabia, and the awareness of public users and businesses about appropriate measures
- an understanding of anti-spam filters used by Saudi ISPs, and their evaluation of the effectiveness of these filters in detecting Arabic and English email spam, including the tricks used by spammers in the headers and bodies of Arabic and English email spam to avoid detection by anti-spam filters.

Further, this study discusses the implications and makes recommendations that can be utilised by the government, ISPs and businesses to combat email spam in Saudi Arabia.

1.2 Scope of the Research

This research concentrates on the study of email spam in Saudi Arabia, and not on other forms of spam such as web, image, and short message service (SMS) spam. Public users (selected from universities, schools, hospitals and government departments), businesses, and ISPs were chosen as a representative sample of Saudi society. Public users were surveyed because they use email daily for personal communication with each other. Because one of the important uses of email is for commerce, to communicate with employees, customers, and other international and local businesses, businesses were also surveyed. And ISPs were surveyed because they have assumed a part of the responsibility to combat email spam in Saudi Arabia. The surveys sought to investigate public, business users' and ISPs' awareness and perceptions about spam and the efforts in Saudi Arabia to combat it.

This research focuses on the investigation of Arabic and English email spam, and not on email spam in other languages, because Arabic is the native and official language in Saudi Arabia (Chejne 2009), and English is the language that is most used in the world (Altbach 2004; Huddleston & Pullum 2002; Kirkpatrick 2007). Both Arabic and English are understood by the researcher and were used to achieve the objectives of this research.

1.3 Aim and Objectives of the Research

The research aim was to understand the nature of email spam in Saudi Arabia, including its volume, its languages and its types; to investigate the awareness of email users about it and the efforts to combat it; and to provide possible suggestions for its mitigation. In order to meet the aim of the research, the following objectives were addressed:

- To investigate the awareness of public users and businesses about email spam, anti-spam filters and the efforts to combat it in Saudi Arabia.

- To investigate the nature of email spam (volume, languages and types) received by public users and businesses, and blocked by ISPs.

- To investigate the differences between Arabic and English email spam.

- To investigate how public users, businesses and ISPs deal with email spam.
- To investigate the effects of email spam on the performance of public users, businesses and ISPs.
- To investigate the anti-spam filters used by Saudi ISPs, and their evaluation of the effectiveness of these filters in detecting Arabic and English email spam.
- To propose a taxonomy that includes most of anti-spam filters used to detect email spam, mostly in English; and then suggest which of these filters could be selected to develop new filters for Arabic email spam.
- To investigate the differences between spammers' tricks used in Arabic and English email spam to bypass anti-spam filters.

1.4 Research Questions

In order to achieve the research objectives that are described above, several questions were developed from the literature review findings (Chapter 2). The methods that were developed to answer these questions are described in Chapter 3. The research questions were:

Awareness of, filters for, and efforts to combat email spam

Q1: Are public users and businesses aware of email spam and anti-spam filters, what are the sources of their knowledge and how do they define email spam?

Q2: Are public users and businesses aware of government and ISPs efforts to combat spam in Saudi Arabia?

The nature of email spam

Q3: What is the volume of email spam received by public users and businesses and blocked by ISPs in Saudi Arabia; in which languages does it occur; and what are the sources or origins of Arabic and English email spam?

Q4: What are the differences between Arabic and English email spam?

Dealing with email spam

Q5: How do public users, businesses and ISPs deal with email spam?

The effects of email spam

Q6: What are the effects of email spam on the performance of public users, businesses and ISPs?

Anti-spam filters and their effectiveness in detecting Arabic and English spam

Q7: What anti-spam filters are used by Saudi ISPs to block email spam, and how effective are they in detecting Arabic and English email spam?

Spammers' tricks used in the headers and bodies of Arabic and English email spam

Q8: What is the extent of the following spammers' tricks used in the headers and bodies of Arabic and English email spam, respectively:

- attractive words or false statements in the subject line
- texts or texts embedded in images in the content
- malicious links and attachments, by type
- fake or obfuscated email addresses.

1.5 Research Methodology Overview

This quantitative study conducted among three different Saudi groups from 2009 to 2014. These groups included public users, businesses, and ISPs. The respondents comprised both males and females who were living in the eastern, western, central, southern and northern regions of Saudi Arabia at the time of collecting the data. They were aged 15 years and older, students and employees, private organisations and government departments. They used email regularly and were willing to participate in the study. Anybody who did not belong to these three groups and was not willing to participate in the study was excluded. Different types of sampling methods have been used to select samples of the study. A multiple cluster random sampling was used to select public user and business participants, and availability sampling (convenience sampling) was used to select ISP participants.

To cover all of the research objectives, three questionnaires were used: one for public users, one for businesses and one for ISPs. A pilot study was conducted to examine the validity of the three groups of questionnaires, and ethical considerations were taken into account. The questionnaire for public users was distributed to 1,500 participants in the central, eastern, western, southern and northern regions of Saudi Arabia, and completed questionnaires were collected from 1,020 participants. The participants were from universities, schools, hospitals, and government departments.

The questionnaire for businesses was distributed to 300 businesses in the central, eastern, and western regions of Saudi Arabia. Completed questionnaires were received from 92 businesses. The participants varied in size, sector, and establishment year. At the time, 27 ISPs were licensed by the Communication and Information Technology Commission (CITC) to provide Internet Service in Saudi Arabia (CITC 2012). The 27 ISPs were located in the eastern (Dammam), western (Jeddah) and central (Riyadh) regions. All were surveyed for this research, and completed questionnaires were collected from 11 ISPs.

One of the objectives of this study was to propose a taxonomy of email spam filters to help in developing methods for combatting it. The phenetics approach, or numerical taxonomy, which has been used in the field of informatics to classify objects based on their similarities (Nickerson, Varshney & Muntermann 2013), was used to develop a proposed taxonomy that classifies the filters on the basis of the similarity of the methods they use to detect email spam. The taxonomy requirements and the development of elements or constructs of email spam filter taxonomy have been considered in this study.

The investigation of tricks used by spammers in the headers and bodies of Arabic and English email spam was another objective of this study. To achieve this investigation, public users, businesses and ISPs (during the collection of questionnaires) were asked to forward Arabic and English email spam that they received in their email inboxes (i.e. email spam that was not detected by anti-spam filters) to a specific email address created for the purpose of this research. A total of 1,270 email spams were analysed. The 1,270 spam emails included 1,035 in Arabic, 179 in English, and 56 in a mixture of languages (i.e. Arabic and English). Internet security software (Kaspersky 2013) was used during the analysis to protect the researcher's computer from any potential malicious links or attachments.

Different statistical tests were used to analyse the data, using the Statistical Package for Social Sciences (SPSS) software (version 18) for Windows.

1.6 Contribution of the Research

This thesis, which addresses three different Saudi groups – public users, businesses and ISPs – is the first study to investigate the nature of email spam in Saudi Arabia and the efforts to combat it in that country. It provides a foundation for future studies on email spam in Saudi Arabia and the issues related to it. The main contribution of this research is an understanding of the nature of email spam, email users' awareness of it, and the efforts to combat it in Saudi Arabia. It will assist the Saudi Government, ISPs and relevant agencies to improve the current efforts to combat email spam, to develop new counter-measures, and to educate Saudi society about it.

The research results provide insights into how the three different Saudi groups (public users, businesses and ISPs) currently deal with email spam and could help them to improve their current efforts of dealing with email spam. This research provides an understanding of the anti-spam filters used in Saudi Arabia and their effectiveness in detecting Arabic and English email spam, as perceived by ISPs. It will assist in improving existing filters, or creating new, more effective filters for languages such as Arabic.

This study proposes a taxonomy of email spam filters based on a large number of filters, which are mostly in English. This taxonomy could help future researchers or anti-spam filter developers in choosing or suggesting appropriate filters for classifying Arabic email spam. This study also provides an understanding of the differences between Arabic and English email spam; a list of keywords and phrases, and an Arabic and English email spam corpora. These materials could help anti-spam developers to refine the existing filters to be more effective in detecting Arabic and English spam.

By providing an understanding of the differences in the tricks used by spammers in the headers and bodies of Arabic and English email spam, the study can help anti-spam developers in developing more effective filters. The research works carried out in this thesis have been published in peer-reviewed international conference papers and journals. Over the course of this research, six research papers have been published. These papers are listed in the publications section.

1.7 The Structure of the Thesis

This thesis comprises ten chapters, which are summarised in Table 1.1.

Chapter 1 introduces a background to the research problem, describes the scope of the research and its aim and objectives, provides the research questions, outlines the research methodology and describes the contribution and the outline of the thesis.

Chapter 2 is the literature review that focuses on the existing studies in the field of email spam, email users' awareness of it, its effects on the performance of email users, and the efforts to combat it. This chapter helped the researcher to identify the gap of the knowledge and to develop the theoretical concepts underpinning the research.

Chapter 3 describes in detail the methods used to conduct the research. It describes the development of the questionnaires for public users, businesses and ISPs, and examines their validity. It describes the data collection methods and procedures, including the sampling methods, the sample size, and the inclusion and exclusion criteria for the three participating groups. The methodology for the analysis of the email spam corpora received from the participants is set out, including how the data were coded and managed for SPSS analysis. Ethical issues for conducting the study, and how they were managed, are discussed.

Chapter 4 presents and discusses the results of the questionnaire given to public users: their perception of email spam, their awareness of anti-spam filters, and the efforts to combat spam. It also describes how the public users dealt with email spam, and its effects on their performance. It also analyses and discusses the results based on the public users' demographic factors.

Chapter 5 presents and discusses the results of the email spam questionnaire given to businesses in Saudi Arabia, to better understand the nature of email spam in Saudi

Arabia from their perspective, including the impact that it has on their performance. It presents the results of the questionnaire about businesses' awareness of email spam, anti-spam filters and the efforts to combat it in Saudi Arabia. It also analyses and describes the results according to the businesses' demographic factors.

Chapter 6 presents and discusses the results of the questionnaire given to Saudi ISPs. It includes ISPs' attempts to raise awareness of their customers and employees, the nature of the email spam blocked by the ISPs, and the nature of their attempts to prevent it. It also provides information about the anti-spam filters used by Saudi ISPs to block email spam and their effectiveness in detecting Arabic and English spam, as perceived by ISPs.

Chapter 7 presents the proposed taxonomy of email spam filters, which includes the different methods that have been proposed by other researchers to detect email spam in different languages, mostly in English. This taxonomy could be useful in identifying appropriate filters for particular spammers' tricks (described and discussed in Chapter 8), especially those used in Arabic email spam. It could also help the researcher or other developers in the future to improve or produce new filters for Arabic email spam.

Chapter 8 describes the results of the analysis of the headers and bodies of a collection of Arabic, English and mixed language (Arabic and English texts) email spam received from Saudi public users, businesses and ISPs. It investigates the tricks spammers used to bypass anti-spam filters.

Chapter 9 discusses the main findings of the research questions as revealed by the questionnaire responses of public users, businesses, and ISPs. It addresses about the nature of email spam, the awareness of it, how these user groups dealt with it, and its effects on their performance. This chapter also discusses spammers' tricks as revealed by the analysis of headers and bodies of Arabic and English email spam. It also provides possible suggestions for government, businesses and ISPs to combat spam in Saudi Arabia.

Chapter 10 concludes the research by revisiting the research aim and objectives, presenting the main findings, providing the novelty of the research, describing the research limitations, discussing the research implications and providing recommendations for future work for other spam issues in Saudi Arabia.

Table 1.1: A summary of thesis chapters

Chapter	Objective of the chapter	Structure of the chapter
1	To introduce the research and outline this thesis	• Background to the research problem • Scope of the research • Aim and objectives of the research • Research questions • Research methodology overview • Contribution of the research • Structure of the thesis
2	To conduct a broad literature review to identify the gap of the knowledge, and to develop the theoretical concepts underpinning this research	• Search strategy used to find relevant articles to identify the gap of knowledge • The definition of email spam • The awareness and education of email users about spam and anti-spam filters • The nature of email spam • How email users deal with spam • The effects of spam on the performance of email users and ISPs • The effort to combat email spam • Spammers and email spam
3	To explain the materials and methods used in the research	• Research aim, objectives and questions • Research philosophy • Research design • Research instruments • Pilot study • Procedures for data collection • Variables • Data analysis • Ethical considerations • Methodology followed in the analysis of the email spam corpora received from the participants
4	To analyse the public user questionnaire data, and to present, discuss and compare the results with those of other relevant studies	• Results of public users questionnaire • Discussion
5	To analyse businesses’ questionnaire data, and to present, discuss and compare the results with previous research studies	• Results of business questionnaire • Discussion
6	To present, discuss and compare the	• Results of ISP questionnaire

Chapter	Objective of the chapter	Structure of the chapter
	results with other relevant studies	• Discussion
7	To propose a taxonomy of major anti-spam filters	• The methodology followed in the development of the proposed taxonomy • The proposed taxonomy of email spam filters • The effectiveness of anti-spam filters in detecting email spam
8	To analyse the headers and bodies of a collection of Arabic, English and mixed (contains Arabic and English texts) email spam, and present the results of the analysis of the tricks used by spammers to bypass anti-spam filters and lure the recipients	• Results of the analysis about tricks used by spammers to bypass anti-spam filters • Discussion of the results of the analysis of tricks used by spammers to bypass anti-spam filters
9	To discuss the main findings of this study with other relevant studies and provide possible suggestions to combat spam in Saudi Arabia	• Revisiting the research questions • Discussion of major research findings • Research suggestions for the mitigation of email spam in Saudi Arabia
10	To conclude this research and explain its limitations and implications, and suggest future research work	• Revisiting the research aim and objectives • Discussion of major research findings and conclusion • Novelty of the research • Research limitation • Research implications • Recommendations for future research.

Chapter 1 has presented an overview of this research. This chapter began by specifying the background of the research problem. It has provided the scope of the research, described its aim and objectives, presented the research questions, outlined the research methodology, explained the contribution of the research and described the structure of the thesis. The next chapter will develop the theoretical concepts underpinning this research and identify the knowledge gap by critically reviewing the existing literature on the field of email spam, email users' awareness of it, the effects of email spam on the performance of email users and ISPs, and the efforts to combat it.

Chapter 2: Literature Review

The purpose of this chapter is to provide a comprehensive review of the existing literature in the field of email spam, the awareness of email users about it, its effects on the performance of email users and ISPs, and the efforts to combat it. Many empirical studies have been critically reviewed to develop the theoretical concepts underpinning this research.

This research aimed to investigate the nature of email spam in Saudi Arabia, the awareness of email users about it and the efforts to combat it; and to provide possible suggestions to mitigate it. The researcher reviewed the literature to identify the gap of knowledge about email spam in Saudi Arabia, and this helped in developing the research objectives and questions. A broad systematic review of studies that discussed email spam in different countries has been conducted by using relevant keywords and defining inclusion and exclusion criteria through databases accessible from the Flinders University library website. Previously published articles were collected and read, and then some new papers were found by using the reference lists of these published articles. The researcher extracted the main aspects of previous studies that have focused on email spam by categorising them into seven main aspects. The researcher concentrated on these seven main aspects in the literature review.

This chapter is organised as follows:

- Section 2.1: describes the search strategy used to find the relevant articles to identify the gap of knowledge.
- Section 2.2: reviews the literature on the definition of email spam.
- Section 2.3: reviews the literature about the awareness and education of email users about spam and anti-spam filters, and the efforts of organisations and governments in other countries in this matter.
- Section 2.4: reviews the literature on the nature of email spam. This includes literature that describes its volume, its languages, its types, and its sources or origins.
- Section 2.5: reviews the literature on how email users deal with email spam.

- Section 2.6: discusses the effects of email spam on the performance of email users and ISPs.
- Section 2.7: reviews legal, technical and other efforts used to combat email spam.
- Section 2.8: reviews the literature on the spammers' motivations for sending email spam, their methods used to collect email user addresses, and the tricks they use to bypass the anti-spam filters.
- Section 2.9: concludes this chapter, and introduces the knowledge gap researched in this thesis.

2.1 The Search Strategy to Find the Relevant Articles to Identify the Knowledge Gap

As described earlier, a broad systematic search of previous published articles was conducted to find the relevant articles for identifying the gap of knowledge. Different databases were used to conduct a systematic search, including ACM Digital Library, Google Scholar, IEEE Xplore and ScienceDirect. These databases were chosen because of previous similar systematic review studies as well as their accessibility via the Flinders University library website. The keywords used to access relevant published articles were based on those used in previous similar studies and also the aspects that have been highlighted in Table 2.1 (Email spam, Efforts, and Participants). In the search strategy, the selected articles were combined by using AND and OR to access more relevant articles. Table 2.1 shows the keywords used to find the relevant articles.

Table 2.1: Keywords used to find the relevant articles

Email Spam–related Keywords	Effort-related Keywords	Participant-related Keywords
Email	Effort	Participant
spam	Educational	Public
UCE	Knowledge	User
UBE	Awareness	Individual
Unsolicited	Attitude	Saudi
Bulk	Experience	Arab
Junk	Dealing	Business
Non-spam	Technical	Organisation
Language	Filter	Enterprise
Arabic	Method	company
English	anti-spam	ISP
Source	Measure	Internet

Email Spam–related Keywords	Effort-related Keywords	Participant-related Keywords
Business	Technique	Provider
Advertisements	Legal	Industry
Pornographic	Legislation	
Political	Law	
Forums	Government	
Malicious	ESP	
Phishing	Combat	
Effect		
Spammer		
Trick		

This systematic search considered some inclusion and exclusion criteria. Articles that were selected to be reviewed were published in the years from 1999 to 2011, written in English, and met keyword requirements. The final number of relevant articles was arrived at in three steps:

- The titles of all selected articles were reviewed, and some of them that were irrelevant to the study were omitted.
- The abstracts of articles that met the inclusion criteria were reviewed and some articles that were not related to this study were deleted.
- The full texts of relevant articles were printed and reviewed by the researcher. From the printed articles, other relevant articles that were cited and listed in the references of these articles were also reviewed. The researcher collected the titles of these articles and searched them to include them in the systematic review.

Finally, about 92 articles were reviewed to identify and explore the gap of the knowledge for this study.

2.2 The Definition of Email Spam

There are various definitions for email spam, also referred to as junk email (El-Halees 2009). These definitions distinguish email spam from non-spam, which is also known as legitimate or genuine email (Blanzieri & Bryl 2008). Previous studies have indicated that the most common definition of email spam is unsolicited bulk email (UBE) and unsolicited commercial email (UCE) (Fawcett 2003; Garcia, Hoepman & Nieuwenhuizen 2004; Sipior, Ward & Bonner 2004).

The acronym UBE means sending a large number of emails that are not requested by recipients (Androutsopoulos, Koutsias, et al. 2000; Cook et al. 2006; Damiani et al. 2004; El-Halees 2009; Hovold 2005; O'Brien & Vogel 2003; Polz & Gansterer 2009; Zaidan et al. 2011; Zhang, Zhu & Yao 2004; Zhuang et al. 2008). The term "Bulk" means sending emails in large quantities, and this term depends on the number of emails, regardless of content (Lueg 2005). On basis of this definition UBE includes all unwanted emails that are sent, whether they originate from commercial, political, religious, pornographic, phishing or fraud websites.

Previous studies have defined UCE as email that contains promotional advertisements sent by different businesses to a large number of recipients (Boykin & Roychowdhury 2004; Carreras & Marquez 2001; Cheng 2004; Sakkis et al. 2003). The term "commercial" is derived from the content of the email. It includes only commercial emails, such as all advertisement emails from businesses, and excludes non-commercial emails such as political, religious, pornographic, and fraud and malicious emails, which are covered by UBE. In contrast, studies such as Ahmed and Oppenheim (2006), Adam (2007), Polanski (2008), Fogel and Raghupathi (2013) and Arutyunov (2013) did not consider UCE to be a definition of email spam, and defined UCE as an easy and quick way to advertise products and services to customers.

The definitions of previous studies of email spam in different countries are summarised in Table 2.2.

Table 2.2: Definition of email spam by other studies in different countries

Author(s)(Year)	Country	Definition of Email Spam
Androutsopoulos et al. (2000) O'Brien & Vogel (2003) Damiani et al. (2004) Zhang, Zhu & Yao (2004) Hovold (2005) Chigona et al. (2005) Cook et al. (2006) El-Halees (2009) Polz & Gansterer (2009) Zaidan et al. (2011)	Greece Ireland USA China Sweden South Africa Australia Palestine Austria Malaysia	UBE
Carreras & Marquez (2001) Sakkis et al. (2003) Hermanson (2003) Boykin & Roychowdhury (2004)	Spain Greece USA USA	UCE

Author(s)(Year)	Country	Definition of Email Spam
Cheng (2004) Bujang & Hussin (2010)	UK Malaysia	
Sorkin (2001) Garcia, Hoepman & Nieuwenhuizen (2004) Pallas & Patrikakis (2005) Lam & Yeung (2007) Youn & McLeod (2007a)	USA USA Greece China USA	Both UBE) and UCE
Cormack & Kolcz (2009)	USA	Unwanted email that was sent indiscriminately, directly or indirectly, by a sender having no current relationship with recipient.
Zinman & Donath (2007)	USA	Email that users cannot easily stop receiving. They also preferred to depend on users' judgment to define email spam.
Hayati & Potdar (2008) Abdoh, Musa & Salman (2009)	Austria Sudan	Irrelevant and inappropriate email that is sent to numerous recipients.
Center for Democracy & Technology (2003)	USA	Junk mail to include jokes and chain letters from businesses, friends and family.
Schaub (2002)	Netherlands	The practice of sending unsolicited emails, most frequently of a commercial nature, in large numbers and repeatedly to individuals with whom the sender has had no previous contact, and whose email address was found in a public space on the Internet, such as news groups, mailing lists, directories or websites.
The Federal Trade Commission (2005)	USA	Any commercial electronic mail messages sent, often in bulk, to a consumer without the consumer's prior request or consent.
Lev & Goldin (2006)	USA	An email that is sent by unknown senders.
Al-A'ali (2007)	Bahrain	Email messages offering or attempting to sell a product or service or attempting to give information that the recipient never requested or has shown interest in, where such information could be offensive to the recipient's person or belief.
Yamakawa & Yoshiura (2010)	Japan	• An email that satisfies one of the following characteristics: • Is sent unilaterally in spite of receiver's intention • Is sent to the general public

Author(s)(Year)	Country	Definition of Email Spam
		- Aim is advertisement or publicity.
Ermakova (2010)	Russia	Email that contain useless information.

Table 2.2 lists definitions of email spam and shows that definitions vary from country to country and by researchers in the same country. In the USA, most of studies defined email spam as UBE and UCE, although some studies in the same country had specific definitions that were different from these common definitions. Different definitions of email spam have been used by studies in different countries in the European Union (EU). Countries such as Ireland, Sweden and Austria defined email spam as UBE, while UCE was used in the UK and Spain. Greece considered both UBE and UCE to be definitions of email spam. The Netherlands and Russia had various definitions of email spam that were varied from the definitions of other European countries. Email spam was defined in South Africa and Australia as UBE. Researchers in some Asian countries, such as China, Malaysia and Japan, have defined email spam: China and Malaysia as UBE and UCE; Japan as a specific definition, as seen in Table 2.2, which did not agree with those of other Asian countries.

From the Arab countries, such as Palestine, Sudan and Bahrain, a few studies have defined email spam. Table 2.2 reveals that researchers in Palestine defined email spam as UBE, which is the most common definition used in other countries, whereas researchers in Sudan and Bahrain had specific definitions that were not the same as the more international definitions. One study indicated that the definition of email spam varied from one country to another and suggested that this could be explained by the country's culture (Abdoh, Musa & Salman 2009). A study by Zinman and Donath (2007) preferred to rely on the user's personal judgement to define spam. However, no studies were found in the literature that indicated how Saudi Arabia defines email spam. Therefore, this study intends to investigate the definitions of email spam used by Saudi public users, businesses and ISPs. Understanding the Saudi society's definition of email spam could help the government and decision-makers to design strategies to combat spam in Saudi Arabia.

The definition of email spam used for the purpose of this research was:

> … an unsolicited, unwanted, bulk email that is sent indiscriminately, directly or indirectly, to a large number of recipients without their permission, and where there is no relationship between the recipients and the senders.

This definition has been used in previous studies such as Androutsopoulos et al. (2000), Cormack and Lynam (2005) and Cormack and Kolcz (2009). Specifically, the provenance of the definition is: bulk, unwanted and unsolicited emails (O'Brien & Vogel 2003) that are sent indiscriminately, directly or indirectly to a large number of recipients without their permission (Cormack & Kolcz 2009), and where there is no relationship between the recipients and the senders (Lev & Goldin 2006).

2.3 The Awareness and Education of Email Users about Spam and Anti-spam Filters

Increasing email users' knowledge about spam is one of the most effective ways to reduce it. Previous studies have suggested that the user and organisation awareness and methods of combating it could mitigate email spam, and they suggested that the information should be communicated to email users. Dantin and Paynter (2005) stated that "an important step towards minimising unsolicited and unwanted emails is raising the awareness of users". Lugaresi (2004) suggested that social awareness of spam, including its causes and appropriate ways to combat it, must be published and reinforced. Awareness could be achieved by educating users, businesses, information centres, associations and privacy advocates (Lugaresi 2004). D'Ambra (2007) stated that "education needs to play a larger role in the fight against spam as computer users either lack the understanding or are not interested in computer security". Refai and Nyanchama (2007) indicated that establishing awareness programs, which provide workshops, seminars and training about spam for employees and customers, could help to increase their awareness and combat the problem.

Šolić et al. (2011) stated that:

> As there are law regulations nowadays and technical solutions like spam filters on both users' and providers' side attention should be paid to the users' behaviour and their awareness of how to suppress spam.

One of the technology consultants at the Mirapoint Company (Everett 2004) recommended that users should be educated spam and staff should be encouraged to sign up to receive corporate email usage policies:

> Examples of corporate policies include writing agreed definitions of what constitutes spam and making it clear that staff should not delete anything that infringes this, but report it to a single point of contact in the company.

In their study of the spam phenomenon in Greece, Pallas and Patrikakis (2005) suggested that ISPs should offer awareness programs to inform email users about spam, effective methods against it, and how to deal with it. The authors reported that "ISPs should take actions by providing assistance to enforcement agencies along with undertaking of users' education about spam".

Awareness programs have been conducted by some governments, ISPs and businesses to educate email users about spam and the appropriate ways to combat it. Pfleeger and Bloom (2005) stated that "the US FTC organisation has made the public more aware of the spam". The member states of the EU have taken many actions to increase users' awareness of spam, such as conducting campaigns to make users aware of spam and the appropriate ways to deal with it (Pfleeger & Bloom 2005). Nine Danish ISPs designed a strategy to combat spam, which included establishing the "ISP Security Forum" organisation. One of the responsibilities of this organisation was to provide common guidelines for customers about spam and filters. Some countries have taken initiatives to educate users (consumers, companies, and public authorities) about spam and how they can avoid it. For example, the Confederation of Danish Industries and the Danish Consumer Ombudsman office produced reports about spam. The first report included anti-spam guidance for private individuals and companies. The second report included advising users about spam laws in the country (Frost & Udsen 2006). Jidiga and Sammulal (2013) stated that "the Indian government organizations like MCIT (Ministry of Communication and Information technology) setup separate divisions to conduct a security awareness programs to the people, employees, students about spam".

Some countries cooperated with each other to combat spam and increase user awareness. According to Moustakas, Ranganathan and Duquenoy (2005), some countries used international initiatives, for example, education, training and awareness of users and businesses. Examples included: the tripartite Memorandum of Understanding on spam enforcement cooperation (an agreement between the US, UK and Australia), the London Action Plan cooperation (an agreement between 15 countries), and the Organisation for Economic Cooperation and Development (OECD). More details about these three international initiatives are provided in this chapter Section 2.7.

Studies have been conducted in different countries to understand the awareness of email users about spam and anti-spam filters. A study conducted in Singapore showed that 42% of users did not know how to protect their computers from the problem. The results of this study suggested that public education was necessary for users and that education could be achieved through workshops and newsletters, and by ISPs advising email users to use anti-spam filters (Leng 2006). A study conducted in Bahrain revealed that most of the participants (74%) did not know about using anti-spam filters to combat spam, while only 26% knew about them (Al-A'ali 2007). Bujang and Hussin (2010) conducted a study in Malaysia to understand the awareness of Malaysian email users, and found that about 86.5% of the participants were aware of email spam, and 66.9% were aware of anti-spam filters.

It can be concluded that generating awareness about email spam is important way to combat it, and that educational efforts to educate users and organisations in the developed countries, such as the USA and some countries of the EU (e.g. UK and Denmark) were better than the efforts in other countries, especially the Arabic countries. With the exception of a few studies, such as that conducted in Bahrain, which found awareness of users in Bahrain to be lower than in countries such as Singapore and Malaysia, no previous studies were found to have conducted research into spam or to have investigated the educational efforts of Arabic countries to inform users about spam. Therefore, this study fills this gap by investigating the awareness of Saudi public users and businesses about email spam, anti-spam filters, and their awareness of the efforts to combat it in Saudi Arabia.

2.4 The Nature of Email Spam

Many studies have discussed different aspects of the nature of email spam, such as its volume, its languages, its types, and its sources or origins. This section reviews the literature for these aspects.

2.4.1 The Volume of Email Spam

This section reviews some studies and statistics provided by some researchers, businesses, ISPs and governments about the average number of email spam received by public users and businesses in different countries. A previous study by Siponen and Stucke (2006) of 500 US and Finland companies, found that the average number of email spam received by companies was 1,987,000 spam weekly. Another study showed that UK companies received an average of 101,500 email spam per week (Computer Fraud and security 2004). Researchers have mentioned different reasons for the large volume of email spam received by recipients. The first reason might be the lack of the awareness of email users about effective ways of dealing with it, which could result in the interaction with spam or responses to offers made by email spam (Barroso 2007; Simpson 2003). The second reason could be that the absence of anti-spam laws or legal efforts in some countries could encourage spammers to send more email spam (Lev & Goldin 2006). The third possible reason might be that spammers develop their methods to bypass filters, which could result in an increase in spam getting through to recipients (Hayati & Potdar 2009; Wang et al. 2007). Previous studies about the average number of email spam received by public users are listed in Table 2.3.

Table 2.3: The average number of email spam received by email users in different countries

Author(s)(Year)	Country	% Total Participants	Average Volume of Email Spam Received/Participant	Period
Gartner Group (1999)	USA	• 40% • 20% • 17% • 9% • 3% • 1% • 1%	• 1-5 • 6-10 • 11-20 • 21-35 • 36-50 • 51-100 • > 100	Weekly

Author(s)(Year)	Country	% Total Participants	Average Volume of Email Spam Received/Participant	Period
Hermanson (2003)	USA	• 31% • 31% • 29%	• ≥ 75%* • 25-74%* • 1-24% * *of all emails	Daily
Özgür, Güngör & Gürgen (2004)	Turkey	• 63% • 37%	• > 50 • > 1000	Weekly
Chigona et al. (2005)	South Africa	• all	• 15	Daily
Chuan et al. (2005)	China	• all	• 8	Weekly
Dong et al. (2006)	China	• all	• 19	Weekly
Grimes, Hough & Signorella (2007)	USA	• all	• 13	Daily
Al-A'ali (2007)	Bahrain	• 12% • 46% • 24% • 18%	• < 5 • 5-15 • 15-25 • > 25	Daily
Bujang & Hussin (2010)	Malaysia	• all	• 5	Daily

The data described in Table 2.3 reveals that the average number of email spam varies from country to country, and this average has increased with the time in some countries such as the USA and China. In the USA, the average (as reported by most of participants in the specified studies) has increased from 1-5 spam to 91 spam each week. The average number of email spam in China has risen from 8 spam to 19 spam weekly. This increased average number of email spam in these two countries, taking into account the lower average number of emails received by Chinese than American users, might be because of the lack of educational efforts undertaken to educate people about ways to deal with it (Dantin & Paynter 2005), or the lower technical measures used to combat it, which could lead to an increase the average number of spam emails (Cheng 2004). It can be seen from Table 2.3 that South Africa and Bahrain had the highest average number of email spam, while China had the lowest average number of email spam. This could be explained by better efforts (e.g. educational, legal and technical) being undertaken by the Chinese Government to combat spam were better than by the governments of other countries. However, no previous studies investigating the average number of email spam in Saudi Arabia were found in the literature. Therefore, this research covers that gap.

2.4.2 The Languages of Email Spam

Email spam is written in different languages, such as English and Arabic. According to Hayati and Potdar (2008), "spam is not restricted to the English language, it can also be seen in other languages like Arabic". Some spammers focus on the English language in email spam, because English is the language most used in the world (Altbach 2004; Huddleston & Pullum 2002; Kirkpatrick 2007), and can be understood by the most people. This could help spammers to reach more email users and reap more financial benefits (Cook et al. 2006; Rogers 2006). Some spammers, however, write email spam in their native language so that they can be understood by people who speak this language, who have the same nationality as the spammer, or who live in the spammer's region, such as Arabic countries (Lev & Goldin 2006).

Pfleeger and Bloom (2005) reported that 80% of the email spam received in the European Union was written in English, although the EU includes about 12 different official languages. A study by Pallas and Patrikakis (2005) to investigate the email spam in Greece has shown that 8% (10 emails) of 125 email spam received by Greek email users were Greek spam. The results showed that the average amount of Greek spam (8%) was smaller than the average amount of English spam received (92%). A study conducted by Dong et al. (2006) has revealed that the average amount of Chinese email spam received by recipients in China was 69%. A study by Symantec (2010) showed that the highest percentage of email spam received in Brazil was in Portuguese (33%), whereas the second most frequent email spam language was English (25.6%).

A survey conducted by Al-A'ali (2007) in Bahrain showed that most of the participants (64%) received English email spam, 18% received both Arabic and English email spam, and 18% received Arabic email spam. The finding of the Bahraini study about the language of email spam was supported by the El-Halees's study (2009), which indicated that "email users in the Arab world have received spam written mostly in Arabic, English or mixed Arabic and English". However, El-Halees's study did not mention which of the Arab countries have received Arabic, English or mixed Arabic and English email spam. Bujang and Hussin (2010) conducted a study of how Malaysian email users deal with email spam. Their results showed that English was the most popular language (90%) used in email spam, and

Malay was the second most frequently used language (82.9%) in spam in Malaysia.

It can be seen, then, that English was the most used language in email spam in different countries, including EU Members such as Greece, some Asian countries, such as China and Malaysia, and some Arab countries, such as Bahrain. This is supported by the Shrivastava and Bindu study (2012), which showed that the most popular language of email spam around the world was English. The literature revealed that English was followed in frequency by the native language of the country in which the email spam was received. However, no studies were found that investigate the languages of email spam received in Saudi Arabia; hence, this study fills that gap by investigating one aspect of the nature of email spam: the languages of email spam received by public users and businesses, and blocked by ISPs.

2.4.3 The Types of Email Spam

Although email spam is written in different languages in different countries, it seeks to achieve the same purposes (Lieven et al. 2007; O'Brien & Vogel 2003). Email spam can be divided into many types, according to the spammers' purposes. Sakkis et al. (2003) stated that "Unsolicited Commercial Email may advertise anything, from vacations to get-rich schemes". Many studies conducted by researchers, businesses, ISPs and governments have discussed the most common types of email spam in different countries. These studies are reviewed in Table 2.4.

Table 2.4: Types of email spam reported by other studies in different countries

Author(s)(Year)	Country	Types of Email SPAM
Cranor & LaMacchia (1998)	USA	Identified: • 35% money-making opportunities, including included pyramid-style schemes, multilevel marketing systems, investment opportunities • 11% adult entertainment, singles services, sexually oriented products and services • 10% direct marketing products and services • 9% informational and how-to guides • 7% advertisements for internet services, computer hardware and software, office products and services

Author(s)(Year)	Country	Types of Email SPAM
		• 3% non-commercial emails • 25% other products and services including phone services, vacation packages, nutritional supplements, weight loss products, credit cards, cable descramblers, online newsletters
Gartner Group (1999)	USA	Identified: • 37% “get rich quick” • 25% adult emails • 18% software offers • 6% website promotions • 5% investment emails • 2% health • 2% contests • 1% vacation • 4% other
Hinde (2002)	USA	Friendly spam defined as an email sent by people by the recipients know, (e.g..family, friends, and colleagues), e.g.: • chain letters • jokes • video clips
Hermanson (2003)	USA	Identified: • travel information • surveys • stocks or quotes • religious • political • offensive or adult • health products • durable goods • dating services • charities • banking • astrology • advanced sales notices
McAfee (2003)	USA	Identified: • 30% refinancing • 27% credit counselling • 27% sexual enhancement products.
The Coalition against Unsolicited Commercial Email (2004)	Canada	Identified: • chain letters • pyramid schemes (including multilevel marketing) • make-money-fast schemes

Author(s)(Year)	Country	Types of Email SPAM
		• phone sex lines • advertisements for pornographic websites
Phelps et al. (2004)	USA	Identified 16 main categories: • 48.8% jokes • 17.7% chain letters • 8.4% inspirational emails • 4.8% religious emails • 4.4% information emails • 3.5% warning emails • 2.8% naked pictures • 1.3% email digests • 1.2% free stuff • 1.2% were comments about a company • 1.1% games • 0.3% missing children • 0.3% company-originated emails • 0.2% political emails • 0.1% good deeds • 4.0% other types of spam. Subcategories: • Jokes emails → general, sexual, gender issues, work- or computer-related, current events, political, poems • Chain letters emails → general, religious, inspirational, luck, free stuff, money • Inspirational emails → thought for the day, “feel good” pictures • Information emails → current events, entertainment and events, helpful tips, recipes • Warning emails → computer viruses, crimes, product • Naked pictures → naked pictures, altered naked pictures
Hulten et al. (2004)	USA	Surveyed 100,000 volunteer Hotmail users, identified: • 30% domestic: e.g. financial services, insurance, government grant programs, items that are very expensive to ship internationally • 32% semi-domestic (Mexico, Canada) but still require shipping, e.g. Viagra and other medical products, college diplomas,

Author(s)(Year)	Country	Types of Email SPAM
		magazines • 38% international products or services, not physical shipping or domestic presence required, e.g. pornographic websites, software, scams
Dantin & Paynter (2005)	New Zealand	Identified: • 20-30% product offers • 20% adult emails • 20% healthcare products • <10% scams (fraud)
The China anti-spam market research (2005)	China	Identified: • 19% shopping online • 13% promoting IT products • 12% get-rich • 10% adult products • 9% vacation • 9% political • 8% business • 7% pornography and violence • 13% other
Chigona et al. (2005)	South Africa	Blocked by the ISPs: • 28% product • 17% adult • 15% financial • 9% scams • 7% health • 6% fraud email • 5% leisure • 4% internet • 4% political • 1% spiritual Received by public users: • > 60% pornographic or adult content • 55% hoaxes • 53% business investment schemes, e.g. get-rich-quick • 38% political speeches • 35% other
The Federal Trade Commission (2005)	USA	From 1,000 email spam, 90% of investment and business opportunities emails were fraudulent, e.g.: • bait-and switch • pyramid schemes • chain letters • credit repair scams

Author(s)(Year)	Country	Types of Email SPAM
		• bogus weight-loss programs
Lev & Goldin (2006)	Not mentioned	Most common types by country: • Russia: food, accessories, education, construction • China: fake invoices, designed to reduce the tax burdens of different businesses; anti-government • Germany: racist and white supremacist • Korea: financial or mortgage-related
Grimes, Hough & Signorella (2007)	USA	Categorised as: • financial • pornographic and other sexual • health • entertainment • computer hardware and software
Al-A'ali (2007)	Bahrain	Identified: • 73% marketing products • 21% fun • 6% sent by mistake
Hanke & Hauser (2008)	Austria	Identified new type: • stock email spam: unsolicited, includes information and recommendations about a specific stock
Yamakawa & Yoshiura (2010)	Japan	Most common types, by language, in Japan: • English: commercial advertisements • Japanese: sexual
Zaidan et al. (2011)	Malaysia	Identified: • commercial advertising • doubtful product • pornography • get rich quick scheme • viruses

It can be seen from Table 2.4 that the most common type of email spam received in different countries such as the USA, New Zealand, South Africa and Bahrain was business advertisements and marketing products. This indicates that the most common purpose of spammers in different countries was to make money (Blanzieri & Bryl 2008; Hayati & Potdar 2008). Some countries had specific types of email spam, which were related to political issues or other events in their countries, and these types could not be found in other countries. Examples include Russia (food and education), China (anti-government), Korea (mortgage-related) and Germany (racist and white supremacist). In Japan, the most common type of Japanese email spam was sexual, a type reported to be less prolific in Islamic or Arabic countries. According to Al-A'ali (2007) and Abdoh, Musa and Salman (2009), there is less pornographic or sexual email in Arabic-speaking countries than in other countries because the Islam religion prohibits pornography. Abdoh, Musa and Salman (2009) claimed that the types of email spam could differ from one country to another because of the motivations and cultures of spammers: some spammers send commercial emails, some send pornographic emails, and others send malicious programs. However, no studies were found that have investigated the types of email spam received by email users in Saudi Arabia. Therefore, this study fills that gap by investigating the types of email spam received by public users and businesses, and blocked by ISPs in Saudi Arabia.

2.4.4 The Sources (Origins) of Email Spam

There are many strategies for identifying the origin of email spam, such as investigating spammer IP, spammer domain location and email spam content (Lev & Goldin 2006); and analysing email headers and constructing honeypots (Boneh (2004). A honeypot is a system or a machine designed to collect email spam and to trap spammers (Pallas & Patrikakis 2005).

Studies have been conducted in different countries to investigate the origin of email spam. Hinde (2002) reported that most scam and fraud emails received in the USA were sent from Africa, especially Nigeria, and they cost victims billions of dollars. Pfleeger and Bloom (2005) found that 80% of email spam received in the EU originated from North America. According to ISPs, most email spam received in South Africa was sent from China, India, and North Korea (Chigona et al. 2005). A survey conducted by the Infocomm Development Authority (IDA) in Singapore revealed that 77% of email spam received in Singapore originated outside of Singapore (Leng 2006). Yamakawa and Yoshiura (2010) revealed that most Japanese email spam originated in China and Taiwan, while only 10% of Japanese spam originated in Japan. Further, a study conducted by the Computer Fraud and Security (2008) which showed that Asia was the top continent for spam generation in the world.

It would seem, then, that most of the email spam received in different countries originates in Asian countries. This might be because Asian countries may not have enacted anti-spam laws, and this would attract spammers from countries that do have them (Yamakawa & Yoshiura 2010). In the Arab world, Al-A’ali (2007) indicated that neither Saudi Arabia or any of the other Arabic countries have enacted laws to combat spam and spammers, which could increase the volume of email spam in that region and encourage spammers to send spam from Arabic countries to other countries. As no previous studies could be found in the literature that investigated the sources or origins of email spam received in Saudi Arabia, a study of this issue is needed. This study investigates the sources of email spam received in Saudi Arabia, as perceived by Saudi ISPs.

2.5 How Email Users Deal with Spam

This section reviews previous studies about the way in which email users deal with email spam. Examples include: reading it, deleting it, reporting it to the ISPs, blocking it with anti-spam filters, and interacting with it and responding to offers such as purchasing, selling and fun (Phelps et al. 2004).

A US survey of 1,018 participants aged 50 and older revealed that about 21% changed their email addresses to avoid receiving spam, 82% deleted it without opening it, and 54% used anti-spam filters to block it (Hermanson 2003). The Radicati Group revealed that 31% of the participants they surveyed had clicked on a link in an email spam (Rogers 2006). Chigona et al. (2005), who studied the perceptions of South African users about spam, reported that 4% of the participants opened email spam.

In New Zealand, Dantin and Paynter (2005) found that 32% of the respondents to their survey used anti-spam filters to block email spam. Grimes, Hough and Signorella (2007) found that 66% of the participants in their US study deleted it, 16.7% used filters to block it, 11.7% contacted their ISPs, and 0.5% contacted the government. A Bahraini study has shown that 82% of the participants read the header and deleted the email, 6% read the entire email and then deleted it, and 12% kept it (Al-A'ali 2007). Bujang and Hussin’s study (2010) of Malaysian email users revealed that 8.1% of the participants responded to email spam, 30.8% clicked on the link, 50.4% used anti-spam filters provided by their email service provider (ESP) and employer, 25.8% used their own filters to block spam, and 7.7% reported email spam.

As well, studies have investigated the reasons why email users respond to email spam. Hermanson (2003) found that 81% of the American participants purchased something and, of these, 16% said that they bought learning materials. The high proportion might be because they lack sufficient knowledge about its negative impact on their performances and computers, and effective ways to deal with it. Responding to spam can result in receiving more spam (Barroso 2007; Simpson 2003). About 4% of the participants in Hermanson's study said that email spam provided a way to find out about new products (Hermanson 2003). Over 10% of the participants in Rogers' (2006) study and 4% of 205 US participants in the study by Grimes, Hough and Signorella (2007) had purchased something from email spam.

It can be concluded that the way in which email users in different countries deal with spam varies from user to user. This can be explained by experience in using email, knowledge of spam and its effects, and knowledge of filters used to block it (Grimes, Hough & Signorella 2007). Responding to email spam can increase the volume of email spam by recipients and have negative impacts on their computers (Lambert 2003). Of interest to this study is the lack of previous studies reported in the literature researching how Saudi public users treat email spam, and the reasons why they respond to it. Hence, this study fills that gap.

2.6 The Effects of Spam on the Performance of Email Users and ISPs

Email spam has a huge effect on the performance of public users, employees, ISPs and businesses (Sorkin 2001). Hovold (2005) stated that "the vast volume of spam being sent wastes resources on the Internet, wastes time for users, and may expose children to unsuitable contents (e.g. pornography)". Spam is second on the list of problems faced by ISPs (Khorsi 2007). The following sections describe the effects of email spam on the performance of email users and ISPs.

2.6.1 Wasting Time and Reducing Productivity

Email users and administrators spend a lot of time in reading, deleting, filtering, blocking email spam, isolating spam from legitimate emails and fixing problems caused by it (Bujang & Hussin 2013; Pérez-Díaz et al. 2012). Employees also waste time checking spam folders to avoid losing important emails that are misclassified by anti-spam filters (Ridzuan, Potdar & Talevski 2010), which reduces the productivity of email users and employees (Moustakas, Ranganathan & Duquenoy 2005; Zhang, Zhu & Yao 2004). The average time that employees have reported sorting out spam problems each day has been reported as 10 minutes in the US (Hinde 2002), 21 minutes in South Africa (Chigona et al. 2005), and 13 minutes the USA and Finland (Siponen & Stucke 2006), and Caliendo et al. (2008) reported that employees spend about 1,200 minutes each year identifying and deleting email spam.

Cook et al. (2006) point out that deleting spam manually from a user's inbox wastes time, which costs employers money in lost productivity. Everett (2004) estimated that the global cost of spam for companies around the world was about $14 per user per month, and Hinde (2002), in the EU, estimated that the cost of spam was US$8 billion a year worldwide. It has been estimated that the cost of email spam for US companies is $10 billion in lost productivity a year (Cook et al. 2006), for Singapore consumers the Singapore Infocomm Development Authority (IDA) reckons the cost at S$23 million in lost productivity a year (Leng 2006). SpamCon Inc. has estimated that, in loss of productivity and resources, fixing spam related problems, and technical support, the cost of one email spam is from $1 to $2, and that this cost can reach millions of dollars a day depending on the number of spam emails sent and received (Atkins 2003 cited in Khorsi 2007). In Japan, the Gross Domestic Product (GDP) loss due to processing email spam was estimated at 500 billion yen a year (Takemura & Ebara 2008).

Pfleeger and Bloom (2005) looked at the burden on ISPs and their staff. The ISPs frequently upgrade their servers, software and hardware to block email spam. The staff read and handled customers' feedback and complaints about email spam, and regularly maintained anti-spam filters software or hardware (Pfleeger & Bloom 2005), which takes time and reduces ISP productivity. The America OnLine (AOL) ISP, which is also an ESP (Goodman & Rounthwaite 2004), claimed that more than

1 billion email spam sent by two spammers prompted 8 million customer complaints (Pfleeger & Bloom 2005). Brod (2004) claimed that the time lost in dealing with email spam problems in the US was 40 minutes weekly. In terms of money spent, the US Federal Trade Commission (US FTC) forum reported that American ISPs spent billions of dollars to stop spam (Allman 2003), and in the EU ISPs paid about 10 billion euro a year to combat spam (Garcia, Hoepman & Nieuwenhuizen 2004).

2.6.2 Offensive Content

The content of email spam differs from one country to another with the motivations and cultures of spammers. Some spammers send commercial emails, some send pornographic emails, and others send malicious programs. Abdoh, Musa and Salman (2009) reported the use of pornographic content, which conflicted with the Arabic culture, to advertise adult products and sexual websites. Offensive spam can also include emails with false claims.

2.6.2.1 False claims

Some studies have discussed false claims in email spam. A study conducted by the FTC in 2003 revealed that common offers included “work at home plans”, “pyramid schemes”, and “get rich quick schemes”. The results revealed that 90% of spam claims for businesses and investments were false (Federal Trade Commission 2003) – a problem for users who respond to the emails. Chigona et al.’s (2005) study of South African users stated that that 51% of the participants said that not only was pornographic spam considered offensive, but also hoaxes and commercial schemes (including get-rich-quick).

2.6.2.2 Pornography

Pornography spam is a significant type of offensive content email spam (Moustakas, Ranganathan & Duquenoy 2005). It can include images, videos and links to external websites, and is a major concern when children have access to them (Al-A'ali 2007). Spammers sometimes send pornographic emails with false statements in the subject line of the email so the recipients do not know what the content of email is about before they open it (Hamel 2004; Simon 2004). An FTC study revealed that 40% of pornographic emails have false statements in the subject line. For example, some spammers send an email with the subject “Re”, when the content of the email is pornographic (Federal Trade Commission 2003).

Pornographic spam can be harmful when read by children (Zhang, Zhu & Yao 2004), as it often includes pornographic advertisements and links to websites that are unsuitable for minors (Androutsopoulos, Koutsias, et al. 2000). Parents are concerned about their children accessing pornographic spam (O'Brien & Vogel 2003). About half (49%) of the participants in Al-A'ali's (2007) Bahrain study were extremely disturbed by pornographic spam and 83% of parents who participated in the study were worried that pornographic spam emails would have an effect on their children.

2.6.3 Consumption of Internet Resources

The large volume of email spam can cause problems with network traffic, bandwidth, memory, and storage space (Androutsopoulos, Koutsias, et al. 2000; Sorkin 2001). This requires organisations and ISPs to pay more money to buy extra bandwidth, capacity for email systems, anti-spam hardware and software, and servers (Chigona et al. 2005; Garcia, Hoepman & Nieuwenhuizen 2004; O'Brien & Vogel 2003).

The US FTC forum reported that ISPs spent billions of dollars to combat spam, without counting the time wasted by the individual recipients (Allman 2003). In the European Union, the ISPs paid about 10 billion euro a year to combat spam (Garcia, Hoepman & Nieuwenhuizen 2004). Cournane and Hunt (2004) stated that the ISPs buy extra bandwidth to provide Internet service to subscribers, and if the large volume of email spam consumes the bandwidth, ISPs have to decide whether to provide a slower Internet service to subscribers or pay more money to increase the bandwidth, and increase charges to subscribers due to the large bandwidth usage. Androutsopoulos et al. (2000) indicated that the email spam consumes Internet, computing and network resources, such as bandwidth, and causes delays for Internet users even if the users do not receive the spam. According to Cook et al. (2006), "running any sort of spam filter on a mail server steals processing time from the server's major purpose: delivering email". O'Brien and Vogel (2003) stated that email spam could be harmful to the capacity, and could slow down the speed of servers and services.

2.6.4 Infection of Computers and Systems by Malicious Programs

Email spam can be a way to transfer malicious programs, such as viruses, worms, trojans, spyware, fraud and phishing, to users' computers and organisations' systems (Cournane & Hunt 2004; Moustakas, Ranganathan & Duquenoy 2005; Pallas & Patrikakis 2005). In this way, email spam can be a threat to computer and network security (Cournane & Hunt 2004; Dantin & Paynter 2005; Hershkop & Stolfo 2004; Moustakas, Ranganathan & Duquenoy 2005; Pallas & Patrikakis 2005; Pfleeger & Bloom 2005; Sorkin 2001). According to Polz and Gansterer (2009):

> Unsolicited email messages evolved from a mere annoyance to a threat for the user. Phishing messages are sent to spy on private data and viruses can be spread via email.

Some malicious programs aim to steal users' identities and confidential information, and some aim to crash users' software and hardware. The US law enforcement officials, federal agencies, and experts at McAfee Corporation state that email spam is a way to steal the identities and money of unsuspecting consumers. The Internet Fraud Complaint Centre (IFCC) estimated that the cost to consumers of online fraud was $17.8 million in 2001, and the estimated number of Americans who are victims to identity theft is 500,000-700,000 each year (Hinde 2002).

Hermanson (2003) stated that email spam can be associated with fraud identity, bank account numbers, passwords, and other important data, affecting customers and businesses. A study revealed that the cost of fraud emails for consumers in the USA was $700 million in 2001 (Federal Bureau of Investigation 2002 cited in Hermanson 2003). In 2002, 36,802 complaints about email fraud were received by the National Consumer League Internet Fraud Watch in the USA (National Consumer League 2003 cited in Hermanson 2003). According to Pfleeger and Bloom (2005), "Trojan horses have turned innocent victims into sources of spam or unleashed malicious software that uses the victim's email account to launch large quantities of spam". In a Chigona et al. study (2005), 56% of the participants in South Africa said that they received viruses from email spam and that it invaded their privacy.

2.6.5 Other Effects

This section describes other effects of email spam, such as causing loss of confidence in using email, annoyance and cost, loss of ISPs' and businesses' reputation, and loss of important emails.

2.6.5.1 Loss of confidence in using email

Being the victim of email spam can lead to loss of confidence in using email, and reducing the usage of email. A study in the USA revealed that 52% of American email users have less trust in email and 25% have reduced their use of email due to the ever-increasing volume of spam (Fallows 2003 cited in Boykin & Roychowdhury 2004). Another study conducted in South Africa found that 59% of the participants complained that spam ruined the reputation and the effectiveness of email (Chigona et al. 2005).

2.6.5.2 Annoyance and cost

Email spam can be annoying and costly for users, businesses and ISPs (Garcia, Hoepman & Nieuwenhuizen 2004). Claiming annoyance value of email spam in various studies were: in South Africa, 91% of a study's participants (Chigona et al. 2005); in a New Zealand study, 64% of participants (Dantin & Paynter 2005); in the USA, 40% of participants, although most of the participants were annoyed at receiving adult and offensive emails, dating services emails, and astrology emails (Hermanson 2003). It annoys most users and slows down the speed of dial-up connections (Sakkis et al. 2003) and costs users money for connections (Androutsopoulos, Koutsias, et al. 2000; Khong 2001).

A study conducted by the European Community revealed that the cost of spam to Internet users was 30 euros a year (Garcia, Hoepman & Nieuwenhuizen 2004). For ISPs, email spam has several financial impacts. ISPs pay money for their infrastructure (e.g. hardware and software development), and for personnel (e.g. customer support personnel and system administrators) (Moustakas, Ranganathan & Duquenoy 2005).

2.6.5.3 Loss of reputation for ISPs and businesses

The reputations of ISPs, ESPs and businesses can be affected when customers receive a large volume of email spam (Moustakas, Ranganathan & Duquenoy 2005), if customers think that the network administrators have sanctioned it. A study conducted in the US revealed that 53% of customers switched their ISPs due to receiving a large volume of UCE (Gartner Group 1999).

2.6.5.4 Loss of important emails

Email spam sometimes causes the loss of important emails due to misclassifying by anti-spam filters. Anti-spam filters have blocked emails sent from “the University of Sussex” because its domain name includes the word “sex”, which is one of the most common spam keywords, and emails that contain information about flight plans sent to the users by some travel agencies, such as Orbitz (Pfleeger & Bloom 2005).

Zhang, Zhu and Yao (2004) reported users email inboxes filling with spam, which can lead to loss of legitimate emails, including important ones, and crashing email servers. Some ESPs (e.g. Hotmail and Yahoo) specify a quota limit (megabytes) for an inbox. When these quotas are exceeded, legitimate emails are rejected by the servers and users lose important emails due to consumption of the inbox quota by junk emails (Cournane & Hunt 2004). A report published by the PEW Internet and American Life Project revealed that 30% of users were concerned that anti-spam filters may block their incoming important emails, and 25% were concerned that their email may not reach others’ inboxes due to anti-spam filters (Fallows 2003 cited in Cook et al. 2006).

It is clear that email spam has different impacts on the performance of public users, businesses and ISPs in different countries such as the USA, the EU member states and South Africa. These impacts reduce the productivity of employees and then affect the economy in these countries (Moustakas, Ranganathan & Duquenoy 2005). To reduce these effects, some countries have introduced legal and technical measures to combat spam (Chigona et al. 2005; Lugaresi 2004), and these efforts will be discussed in the following section. Such action has reduced the volume of email spam and its effects in these countries (Cheng 2004; Lev & Goldin 2006).

In the literature, no research or evidence of previous studies were found that investigated the effects of email spam on the performance of public email users, businesses and ISPs in Saudi Arabia. Therefore, this research fills this gap. Understanding the effects of email spam on the performance of email users, organisations and ISPs, and the size of the issue in Saudi Arabia can help the Saudi Government and other decision-makers to design strategies and policies to combat spam and its effects at an early stage.

2.7 Efforts to Combat Email Spam

There are at least two ways available to combat spam: legal and technical. They are not, however, completely successful, because of a number of issues and difficulties (Seigneur et al. 2004). The effect of legal measures in one country is mitigated when other countries do not enact laws to combat the problem. This can result in more email spam being sent from countries that do not legislate against spam to other countries. This makes the applying of spam laws very complex (Khong 2004). Also, there are three particular issues that affect the application of legal action: evidence, deterrence, and cross-border jurisdiction (Khong 2004). Moustakas, Ranganathan and Duquenoy (2005) have argued that legislation can be effective if the penalties against spam are defined, and if they are applied in court when victims complain.

When developing technical approaches to blocking spam, difficulties arise with specific features of the languages of the email, as each language has specific properties that are different from those of other languages. This can reduce the performance of anti-spam filters for different languages. Users' interpretation of keywords and phrases in spam is another difficulty. Some users take certain keywords and phrases as indicators that the email is spam, when others consider the same keywords and phrases to be legitimate. This can complicate the development of anti-spam filters (Lev & Goldin 2006). Legal, technical and other efforts to combat spam are reviewed in the following sections.

2.7.1 Legal Efforts

Spam is becoming an international problem and has caused many issues for different countries (Cook et al. 2006). In response, many countries have applied laws against spam to reduce its impact. Some countries, such as the USA, EU countries, Australia and some Asian countries, have enacted laws to combat it. This section reviews the legal efforts of such countries to combat spam.

2.7.1.1 USA

In the USA, there are two levels of laws: federal and state. Federal spam laws enacted on 16 December 2003 were the first US attempt to combat spam by legislation. These laws are regulated by the FTC (Rogers 2006; Sorkin 2009).

Some states in the US have also enacted special laws to combat spam (Sorkin 2009).

The first laws were legislated in Nevada in 1997, giving recipients the opportunity to be able to opt out of receiving spam (Pfleeger & Bloom 2005; Rogers 2006). Laws for Washington State prevent sending spam to the state's residents. California State required that email spam be identified with "ADV" in the subject line to allow anti-spam filters to detect incoming email spam. Kansas State gave recipients the right to sue the senders of unlawful email spam (Fogo 2000).

Virginia enacted legislation to combat spam that included criminal penalties for fraudulent and high-volume spamming. Examples of these penalties are imprisonment for one to five years and forfeiture of computer equipment for spammers who send more than 10,000 spam in 24 hours or 100,000 spam in a 30-day period (Butler 2003). Virginia State law also attempted to ban misleading subject lines and forged email headers. Grimes (2004) stated that:

> … the Virginia law enacted under that state's Computer Crimes Act addresses the use of misleading subject lines, forged email headers, and criminal trespass when a spammer illegally uses a computer to send out email messages and help disguise the origin of the email.

2.7.1.2 The EU

The EU member states have issued many directives regarding users' privacy (Hinde 2003; Khong 2001; Lugaresi 2004). The first directive concerned privacy protection: each user needs to authorise a company to use personal data such as email. The second directive was about customer protection and long-distance contracts: companies must get permission from a user before they can advertise their services and products to them via the Internet. The third directive was on telecommunications privacy protection, which:

> … outlaws all automatic systems to call a user and says that all advertising expenses must be paid by the company and not the user (faxes and emails are instead paid by the user) (Sorkin 2009).

The fourth directive concerned electronic commerce (Moustakas, Ranganathan & Duquenoy 2005; Rogers 2006; Schaub 2002). Each member state of the European

Union has implemented these directives based on its national legislation. Examples are the United Kingdom (UK) (Cheng 2004), Denmark (Frost & Udsen 2006), Austria, Finland, and Italy (Khong 2001). In Denmark, two mailboxes were created by the Danish Consumer Ombudsman office to receive complaints about spam. The first mailbox was to receive complaints about Danish spam and the second was for international spam. The office received 300-400 complaints monthly about Danish spam and 30,000-40,000 complaints about international spam. The staff of the Danish Consumer Ombudsman take appropriate action (Frost & Udsen 2006).

2.7.1.3 Australia

In Australia Spam laws became effective on 11 April 2004. The Australian Communications and Media Authority (ACMA) enforces these laws, which are responsible for providing information about spam to customers and businesses. These laws have provided definitions of spam, its main types, and the legal procedures to combat spam (Australian Communications & Media Authority 2006; Cheng 2004).

2.7.1.4 Some Asian Countries

Examples of Asian countries that enacted laws against spam are Japan and Singapore. In 2002, Japan enacted a law on the regulation of transmission of specified electronic mail. This Act, under the jurisdiction of the Ministry of Internal Affairs and Communications, explained the meaning of electronic mail and legislation about email issues (Ministry of Internal Affairs and Communication 2007; Moustakas, Ranganathan & Duquenoy 2005). The legal efforts to combat spam in Japan reduced its volume compared to other countries around the world. According to Lev and Goldin (2006), “the spam to legitimate email ratio in Japan is much lower than average due to the strict attitude towards law enforcement”.

Singapore enacted the Spam Control Act in 2007 to combat unsolicited bulk commercial communications email. It provided definitions, methods of collecting email addresses, and the legal procedures to combat spam (Attorney General's Chamber 2007).

2.7.1.5 South Africa

In South Africa, the problem of spam was included in the 45th section of the

Electronic Communications and Transactions (ECT) Act in 2002. This Act made spam illegal in South Africa; spammers could face fines or 12 months imprisonment. However, “the Act is fraught with problems. It fails to give a clear definition as to what spam is and it actually even restricts the filtering of spam by ISPs” (Bolin 2005 cited in Chigona et al. 2005). Most ISPs in South Africa have reported that legal efforts to combat spam are not effective because, to avoid the local laws, the spammers move their activities to countries with no laws against it (Chigona et al. 2005).

In spite of the enactment of anti-spam laws in different countries and their effectiveness in reducing email spam in these countries, e.g. Japan (Lev & Goldin 2006) and the USA (Xu 2010; Yamakawa & Yoshiura 2010), there have been many issues in applying anti-spam laws (Chigona et al. 2005; Khong 2004). One of these issues is that when spam is sent from outside of countries that have applied laws prohibiting it, the laws are ineffective, and spammers are encouraged to continue sending email spam from countries that do not implement laws to combat spam (Khong 2004). As a result, some countries are cooperating with each other (Leng 2006); the tripartite Memorandum of Understanding on spam enforcement cooperation, the London Action Plan cooperation, and the Organisation for Economic Cooperation and Development (OECD) (Moustakas, Ranganathan & Duquenoy 2005).

The tripartite Memorandum of Understanding on spam enforcement cooperation is “an agreement between the UK, US, and Australia in combating spam”. The tasks of the agreement include the collaboration of authorities in the three countries to investigate spammers in those countries, and joint training programs to combat spam (Department of Trade and Industry 2004 cited in Moustakas, Ranganathan & Duquenoy 2005). The London Action Plan is an agreement between 19 bodies from 15 countries to combat spam and its problems. It involves communication and collaboration between agencies in developing effective legislation and techniques against spam, educating people and businesses about spam, and effective ways to support government agencies in combating spam (Office of Fair Trading 2004 cited in Moustakas, Ranganathan & Duquenoy 2005). The OECD created a task force to follow up the efforts of governments, businesses and civil society in combating email spam. The OECD aims are:

> … coordinating international policy responses in the fight against spam, encouraging best practices in industry and business, promoting enhanced technical measures to combat spam along with improved awareness and understanding among consumers, as well as facilitating cross-border law enforcement (OECD 2004 cited in Moustakas, Ranganathan & Duquenoy 2005).

Educating public email users, businesses, ISPs and other organisations about the legal efforts to fight spam can effectively reduce its effects on their performances (Lugaresi 2004). In the literature, the awareness of governments' legal efforts (e.g. Malaysia) has been investigated. A study conducted by Bujang and Hussin (2010) on email spam in Malaysia revealed that only 14.6% of the participants were aware of legal efforts and services enacted by the Ministry of Science, Technology and Innovation. This may not be sufficient to mitigate the volume of email spam and its effects in Malaysia, as most email users had not been informed about these efforts by the Malaysian Government, and did not know the appropriate legal procedures to follow when receiving it. In Saudi Arabia, no studies have been found that investigate the awareness of Saudi society about the government's efforts to combat spam. This study fills this gap by investigating the awareness of public users, businesses and ISPs in Saudi Arabia of government efforts to combat spam including the legal efforts, if any.

2.7.2 Technical Efforts

Technical efforts are another way to combat spam. Many ISPs have applied technology, such as effective anti-spam filters (Lam & Yeung 2007). Organisations using anti-spam filters can save millions of dollars. Osterman Research Inc. (2008) estimated that the cost of email spam to a company with 1,200 employees could be US$2.4 million, but by using anti-spam filters, they could save US$1.2 million.

Anti-spam filters can be software or hardware, and have been designed using different methods, such as content or reputation based methods. Studies have indicated that these filters are effective in detecting email spam. According to Sorkin (2001), "filtering by ISPs and third-party proxy filtering services like Brightmail can be more effective than end user filtering, requiring less effort and expertise on the part of the users". The ISPs can block known spammer's addresses and their origins, and can collaborate with other ISPs to identify spammer; and third party proxy filtering services can filter out spam. This stops email spam before it reaches users' inboxes (Sorkin 2001).

A technology consultant at one of the companies that sells email security products (Mirapoint), recommended several ways to protect networks from security attacks. One way was for network managers to use an email firewall with anti-spam software to monitor and clean machines, and update the software regularly. Another way was to implement intrusion detection software to prevent spammers' activities from taking place within the firewall (Everett 2004). Khong (2001) stated that two levels of anti-spam filtering are needed to combat email spam effectively: by users and by ESPs. This section reviews the literature on technical efforts of ISPs and ESPs to combat email spam by using a variety of filters.

Some email client software, such as Eudora and Microsoft Outlook, include filtering services that can identify spam and delete it automatically (Sorkin 2001). Email spam can be identified by the email header, email content, spammers' blacklists, and email spam archives. Some ISPs have installed email "smarthost" servers for their customers. The customers use an email client to transfer outgoing emails to the smarthost. The smarthost servers then arrange emails for delivery to remote sites. Some ISPs redirect all outgoing port 25 traffic by Simple Mail Transfer Protocol (SMTP) to the smarthost, and make its use compulsory (Clayton 2004). ESPs such as Hotmail have enabled users to classify email addresses to whitelists or blacklists (Hershkop & Stolfo 2004). Most ISPs in North America have used commercial software to block email spam. The most common filter used in the literature was Brightmail, which is a filter produced by the Norton Corporation. The effectiveness of Brightmail in blocking email spam was high, blocking 95% of email spam (Gartner Group 1999 cited in Chigona et al. 2005).

Lam and Yeung (2007) reported that many ISPs around the world have implemented anti-spam filters at the email server level, the most common filter being Naive Bayes (NB). South African ISPs have used many open source filters to block email spam before it reaches their SMTP servers. Examples of these filters were Postfix, Sender

Policy Framework (SPF), and SpamAssassin. With the open source filters, the ISPs also used different filtering techniques, such as Bayesian filters, distributed blacklists, heuristic engines, and statistical classification filters, to reduce the chances of spam penetrating SMTP servers. Most ISPs agreed that the Bayesian filters were more effective than other filters in detecting email spam (Chigona et al. 2005). Most ISPs in Greece use anti-spam filters such as DNSBLs, heuristic techniques, and custom techniques to block email spam. The ISPs were satisfied with the filters that they used, but had concerns about the effectiveness of these filters in classifying email as spam or legitimate (Pallas & Patrikakis 2005).

As discussed in the previous paragraphs, various anti-spam filters have been used by different countries, such as South Africa and Greece, and some were effective in detecting email spam while some were not. Studies by Wang et al. (2007) and Hayati and Potdar (2009) claimed that the effectiveness of anti-spam filters might be reduced, because spammers continuously develop their methods and tricks to bypass these filters. The use of effective anti-spam filters can save companies millions of dollars. A study by Osterman Research Inc. (2008) indicated that the cost of email spam onto a company with 1,200 employees could be $2.4 million, but by using anti-spam filters, they can save $1.2 million. Ridzuan, Potdar and Talevski (2010) argued that renewing the licence or updating anti-spam filters can also cost businesses much money, but this cost is still lower than the cost of email spam to loss of productivity.

Other researchers such as Çıltık and Güngör (2008) and El-Halees (2009) have claimed that the effectiveness of filters in detecting email spam differed from one language to another, with higher effectiveness for English than for other languages. Subramaniam, Jalab and Taqa (2010) claims that “anti-spam methods used for English language spam detection may not produce higher performances given the nature of different human languages”. Anti-spam filters have been shown to be more effective at detecting English spam than Arabic spam (El-Halees 2009); Turkish (Çıltık & Güngör 2008); and Vietnamese (Nguyen, Tran & Nguyen 2008). However, no previous studies could be found in the literature that investigated the effectiveness of anti-spam filters used by the Saudi ISPs to detect Arabic and English email spam, which is one of the aims of this research study.

2.7.3 Other Efforts

Some researchers suggested that an important solution to email spam could be a combination of effective measures, such as technical, legal, international collaboration, and educational. Cheng (2004) suggested that an important solution was "a combination of self-help preventive measures such as anti-spam filtering tools, robust regulation, international cooperation, and education and awareness of users". Frost and Udsen (2006) suggested the use of a combination of legislation, technological improvements such as using advanced filters, user and company education and self-regulation by businesses and ISPs. This section reviews other efforts of businesses, ISPs and some government sectors to combat spam.

2.7.3.1 Applying clear policies against email spam

Some ISPs and businesses have implemented strong standards and policies for employees and customers to control the use of email in the organisation. These policies could contribute to reducing the volume and effects of email spam. Sorkin (2001) described the application by some ISPs and organisations of clear policies forbidding the use of their facilities to send email spam, and have blacklisted and boycotted spammers and spam-friendly providers. Some companies have developed standards and policies to combat spam. The ePrivace Group has developed the Trusted Email Open Standard (TEOS) to reduce the volume of spam (Pfleeger & Bloom 2005).

Industry groups representing marketers and ISPs have combated spam by applying self-regulatory policies. The policy implemented by the Direct Marketing Association (DMA) specified that members of DMA are prohibited from sending email spam to email addresses that appear in the DMA database (Leng 2006; Sorkin 2001). Sunner (2005), studying 182 IT security professionals in the UK, revealed that 51% had formal policies relating to security attacks.

2.7.3.2 Creating specific research groups, scientific forums, or work teams to combat spam

Establishing specific research groups and forums, and work teams with specific responsibility to combat spam, is an important way to reduce its effects. The aim is to discuss and develop effective ways to combat email spam in legal, technical and other ways. Von Solms (2005) claims it is important for the organisations to establish business units or create teams to manage network security, including spam. These units or teams apply and update internet security software or hardware to block security attacks, and design security policies for the organisation (von Solms 2005). Ridzuan, Potdar and Talevski (2010) stated that companies and ISPs need to spend money to recruit employees to deal with spam problems, and to provide the required training for those employees. As mentioned by previous studies, there are many things those employees can do to combat spam. According to Alongi (2004), "ISPs hire employees to screen spam, install filtering programs, terminate spammer accounts, and file lawsuits". Alepin (2004) reported that ISPs hire personnel to solve problems caused by email spam, provide technical support for customers and handle users' complaints about it.

In 2003, the Anti-Spam Research Group (ASRG) was formed to combat spam and to reduce its effects (Allman 2003). Operating under the Internet Research Task Force (IRTF), the ASRG looked at spam problems that could be solved by technical solutions. It was tasked with developing anti-spam tools, techniques for preventing spam, and administrative tools; evaluating frameworks and measurements; and investigating effective technical solutions (Internet Research Task Force 2013).

In 2006, nine Danish ISPs, which account for about 98% of the internet users in Denmark, created an organisation to combat security attacks. This organisation, called "the ISP Security Forum", aimed to achieve the following tasks: provide a central spam filter for customers; and take actions against spammers who send spam from their internet connections (Frost & Udsen 2006). Unix to Unix Network (UUNET), which is located in the US and one of the largest ISPs in the world, created a special group of six employees with a budget of one million dollars and with a specific responsibility to combat spam (Khorsi 2007). Another study by Johnson and Koch (2006) revealed that about 12% of the budget of the IT department was spent on network security in the USA.

In 2005, the European Network and Information Security Agency (ENISA) was established in Greece to achieve a high level of information security within EU institutions and member states. According to Rossow (2007), "the ENISA seeks to develop a culture of network and information security for the benefit of citizens, consumers as well as business and public sector organisations in the European

Union". In (2013), Arutyunov reported that an American company allocated one full-time IT person for every 690 employees to fix problems related to spam.

2.7.3.3 Cooperation of ISPs with other sectors (public or private) to combat spam

Cooperation between the ISPs, ESPs, users, business and the government is important to trace the sources of email spam and then combat it legally or technically. According to Leng (2006), the collaboration of ISPs with network service providers is necessary to trace the origin of email spam. Butler (2003) reports that AOL, Microsoft and Yahoo collaborated actively with US law enforcement agencies to combat spam. They have developed a mechanism that includes preserving evidence relating to spammers' activities and coordinating enforcement efforts with industry, including referral of spammers to the police or government agencies (Butler 2003).

At the London Internet Exchange (LINX) forum, about 150 ISPs tackled spammers who hosted their websites on reputable ISPs but sent spam from other networks. LINX recommended shutting down websites that sell spamming accessorise, such as stolen email addresses. Malcolm Hutty, a LINX regulation officer, said that LINX was the best current practice to stop spam. He said that the number of open relay mail servers that sent spam was about 20% of the UK mail servers in 1999 and this number decreased to less than 1% in 2003. He also said that LINX was responsible for reducing the volume of spam sent from the UK to less than 1% ('ISPs get tougher on spam' 2004).

Users can reduce the volume of email spam and its effects by cooperating with ISPs, for example, by paying for spam filtering services. A study conducted by the Gartner Group (1999) revealed that 24% of the participants in the USA were willing to pay money to ISPs that provide a spam filtering service. Of those the participants, 70% would pay $1 or more per month for the ISPs to filter spam (Gartner Group 1999). Another way for users to cooperate with the ISPs to combat email spam is by paying additional money for exceeding an email limit in a certain period: "ISPs could play a role in curbing spamming by limiting the number of emails a person can send over a certain period or charge the sender for exceeding the quota" (Leng 2006).

This section reviewed and discussed many efforts conducted by governments, ISPs and businesses in different countries to combat email spam. This discussion leads to the following questions:

- Are public users and businesses aware of the government and ISPs efforts to combat spam in Saudi Arabia?
- What do public users and businesses in Saudi Arabia perceive these efforts to be?

The literature search found no evidence of previous studies that answer these questions for Saudi Arabia. This study answers them.

2.8 Spammers and Email Spam

Email is a useful tool that enables people to communicate both text and multimedia messages to each other. According to Phelps et al. (2004), “more than 90% of the internet users use email, and 50% of the online population is using email on average each day”. However, some people, spammers, exploit email for their own purposes (Pathak, Hu & Mao 2008). Kumar et al. (2014) defined spammers as “senders who send fake mails and try to spam the client’s mailbox in order to get some sort of information”. Madleňák (2006) defined spammers as people use email addresses that have been collected without the consent of the owners of these addresses. This section describes the motivations behind spammers and the methods they use to collect recipient email addresses, and reviews previous studies about tricks they used bypass anti-spam filters.

2.8.1 Spammer Motivations

Why do spammers continue to send email spam to users despite a number of methods used to fight their activities? The easy answer to this question is: to achieve huge benefits in a short time at low cost (Carreras & Marquez 2001; Hayati & Potdar 2008; Lieven et al. 2007; O'Brien & Vogel 2003); they can send one email to thousands of people in a few minutes (Cook et al. 2006). In 2003, the *New York Times* interviewed “one of the most prolific senders of junk email messages in the world”. He reported that he had over 150 million email addresses from over 24 countries, and can send email spam to 70 million users per day, making about $500 from each one million emails sent (Rogers 2006). The major motivations for email spam are described below.

2.8.1.1 Service and product advertisements

Spammers collect email addresses in different ways, such as from forums groups, buying from individuals or collecting addresses by automated software. Their purpose is to advertise their commercial products, such as medical, software, and hardware, and services, such as educational consultations (Blanzieri & Bryl 2008; Cook et al. 2006; Dantin & Paynter 2005; Hayati & Potdar 2008).

2.8.1.2 Stealing confidential email user information

Spammers accumulate email addresses so that they can send spam to them (Pfleeger & Bloom 2005). They can gain access to users' computers and steal users' information by including a link or attachment in the email for recipients to click on (Blanzieri & Bryl 2008; Hayati & Potdar 2008). Email spam is not only used for marketing products; it can also be used to steal email user identities via phishing and fraud (Hershkop & Stolfo 2004). Symantec has revealed that the volume of phishing emails increased 44% from the first half of 2005 to the second half (Lam & Yeung 2007). Spammers send spam containing malicious programs to recipients, such as CryptoLocker, in 2013. CryptoLocker is a malicious program that attacks computers, encrypts victims' files (e.g. documents, videos and images), and then asks victims to pay US$300 within 48 or 72 hours in order to receive a decryption key and retrieve their data (RIT 2013; Sophos 2013). Users who are not aware of the effects of the malicious programs download them to their computers, potentially losing important information such as credit card details, passwords and email addresses (Kumar 2009).

2.8.1.3 Crashing computers and email servers

Spammers also send malicious programs, links or attachments that include viruses, trojans, or worms to the recipients (Pallas & Patrikakis 2005). When users download a malicious program or an attachment, or click on a link, malicious actions occur, such as pop-up advertisements, opening websites, and running and then crashing users' computers and email servers (Blanzieri & Bryl 2008; Dantin & Paynter 2005; Hayati & Potdar 2008).

The different methods that spammers use to collect large numbers of email addresses and tricks to bypass anti-spam filters are described in the following sections.

2.8.2 Methods Used by Spammers to Collect Email User Addresses

Spammers seek to send spam to users and businesses easily and quickly, so they need to collect a huge number of email addresses. To achieve this, spammers have used many methods, such as email harvesting, direct spamming, anonymous operations, zombie networks, and Border Gateway Protocol (BGP) spectrum agility.

2.8.2.1 Email harvesting

Spammers collect valid email addresses using automated browsing software called crawlers (Cournane & Hunt 2004). Crawler software sends a hypertext transfer protocol (HTTP) request to find web pages and documents. After retrieving a HTTP response from the web server, spammers send content and links to email addresses, which are used to build lists and create databases for potential users, and uses links to other web pages to continue the crawlers' process. This method is very important for spammers because they can build lists of victims before sending them spam email (Andreolini et al. 2005).

2.8.2.2 Direct spamming

Spammers can buy connectivity from 'spam-friendly ISPs' that do not care about spamming activity. Sometimes spammers who purchase connectivity are forced to change their ISPs when they send spam from ISPs that do not accept their activity (Ramachandran & Feamster 2006). They can also purchase email addresses from individuals and organisations (Cook et al. 2006).

2.8.2.3 Anonymous operations (open relays or proxies)

Spammers can sometimes hide their traces by using one or more open proxies (Boneh 2004). The open relay or proxy is an SMTP server that allows connection between the user and server without the need of authentication (Ramachandran & Feamster 2006). So spammers establish a Transmission Control Protocol (TCP) connection in the first open proxy (Garcia, Hoepman & Nieuwenhuizen 2004) and use this connection to create a new proxy connection with another open proxy. As a result, chains of proxy connections are established and spammers use these chains to forward emails to users. By using the open proxy chains, spammers are difficult to trace (Andreolini et al. 2005).

2.8.2.4 Zombie networks (bot networks)

A zombie is a computer that is infected by viruses, worms or trojan horses and can be controlled and used by remote entities to achieve special motivations (Ramachandran & Feamster 2006; Xie et al. 2008). A bot network was defined as thousands of machines that are used to run malicious programs (Boneh 2004). According to Cook et al. (2006), "a large amount of these computers, usually called a network or army can be co-opted to send spam emails, requiring little of the spammer's own computing power and network bandwidth". This is a popular method because it protects spammers' identity (Paulson 2004).

2.8.2.5 BGP spectrum agility

Border Gateway Protocol spectrum agility is a new cloaking mechanism (Kosik, Ostrihon & Rajabiun 2009). Ramachandran and Feamster (2006) state that "spammers briefly announce (often hijacked) IP address space from which they send spam and the routes to that IP address space once the spam has been sent". In addition, spammers can use spectrum agility to complement spamming by other methods (Ramachandran & Feamster 2006).

Although spammers have used the methods described above to collect email addresses, different filters have been developed to combat their activities. Consequently, spammers use tricks to bypass these filters (Wittel & Wu 2004). These are reviewed next.

2.8.3 Tricks Used by Spammer to Bypass Anti-Spam Filters

Spammers use a variety of tricks in the header and body of email spam to achieve their objectives. Some researchers (Nielson, Aycock & de Castro 2008; Zuo et al. 2009) have claimed these tricks are used to bypass the anti-spam filters. Other studies (Attar, Rad & Atani 2013; Dhinakaran, Jae Kwang & Nagamalai 2009) suggest that the tricks are used to lure the recipients to open and read the email. This section describes some common tricks used by spammers to achieve their purposes, and reviews studies that have investigated these tricks in different countries and languages.

2.8.3.1 Using attractive words or false statement in the subject line of email spam

To lure the recipients to open emails or to make them think that it is important and they should read it, spammers use attractive words, phrases or false statements in the subject line (Attar, Rad & Atani 2013; Dhinakaran, Jae Kwang & Nagamalai 2009). Examples of attractive words or phrases observed in the subject lines of email spam have been described in the literature; for example:

- “Account confirmation”, “message from the bank”, “security warning”, and “update details” (Dhinakaran, Jae Kwang & Nagamalai 2009)
- Fake or real news events, inexpensive products, or easy ways to make money (Smith 2008)
- “Sex”, “for sale”, “get rich” and “best deal” (Wang & Chen 2007)
- “Hi”, “Hello”, “Was this from you?”, “Alert”, or “Thank you” (Wei et al. 2008)
- “Re” – implies the spammer is answering an email from the recipient (Chen, Zhan & Li 2010).

Many of these keywords and phrases have also been observed in the header and content of email spam in different languages, such as English and Arabic. In English, for example, recurring words are “Viagra”, “Sex”, “Pizza”, “refinance”, and “Mortgage”(Lev & Goldin 2006); “Viagra”, “Sex”, “Buy Now”, “You’ve Won”, and “Free” (Cook et al. 2006). Keywords in both English and Arabic spam include “ التحق, Join”, “الآن, Now”, and “اضغط, Click” (Goweder, Rashed & Alhamammi 2008). Wahsheh, Alsmadi and Al-Kabi (2012) used Google’s search-based keyword (SBK) tool to extract the following top 10 keywords used in Arabic and English spam: “ألعاب, Games”, “صور, photos”, “أغاني, songs”, “فيس بوك, Facebook”, “يوتيوب, YouTube”, “جامعة, University”, “دردشة أو شات, Chat”, “منتديات, Forums”, “طرب, Tarab”, and “بلياردو, billiards”.

A false statement in the subject line (also called a misleading subject line) is another trick that spammers use to bypass anti-spam filters. It was defined by Hamel (2004) and Simon (2004) as a subject line that does not indicate the content of the email. For misleading subject lines, spammers added, for example, greetings or thank words or phrases in the subject lines, while the content included phishing attachments or product advertisements. This can make it difficult for the recipients to determine the content of the email before they open it (Chigona et al. 2005).

The meanings of keywords and phrases were similar in both Arabic and English email spam, and they aimed to achieve the same spammer purpose, such as business and entertainment advertisements. O'Brien and Vogel (2003) and Lieven et al. (2007) claimed that email spam is written in different languages in different countries, but it seeks to achieve the same purposes. This section reviewed keywords and phrases observed in Arabic and English because Arabic is the official language in Saudi Arabia (Chejne 2009), English is the most used language in the world (Altbach 2004; Huddleston & Pullum 2002; Kirkpatrick 2007), and the researcher was able to understand both languages. Christina, Karpagavalli and Suganya (2010) has indicated that "using combinations of keywords is a good solution to enhance filtering efficiency". Knowledge of the keywords and phrases used in Arabic and English email spam could lead to the development of more effective anti-spam filters.

2.8.3.2 Using different formats for the content of email spam

Another trick used by spammers is to use different formats when creating the content of email spam, such as text embedded in an image (also called image spam). Image spam began in 2004 (Kelly 2007), its volume reaching 1% of all email spam around world in late 2005 (Soranamageswari & Meena 2010). This volume grew to be 55% of all emails spam in 2010 (Attar, Rad & Atani 2013). Image spam was defined as a type of email spam in which the content of the email appeared as an image instead of text in the body of the message (Soranamageswari & Meena 2010; Xu, Wang & Shao 2009).

Studies have indicated that the reason for using image spam was to bypass the anti-spam filters, especially text- based filters. In (2013), Attar, Rad and Atani described it as:

> … a new threat which is the most sophisticated kind of spam emails up to now, because it makes the message interesting for the user and hard to detect by text based anti-spam filters.

Image spam was developed for circumventing anti-spam filters that classify spam based on texts included in the body of messages (Nielson, Aycock & de Castro 2008; Zuo et al. 2009). Gargiulo and Sansone (2008) described it as a new trick that can be attractive to users and remain undetected by text-based filters.

2.8.3.3 Adding links or attachments to the content of email spam

Another trick that spammers may have developed in an attempt to evade text-based anti-spam filters is to add links and attachments to the content of the email. These links direct users to webpages that promote products or commercial services, rather than including commercial advertisements for products as text in the body of emails, which is easy for text-based anti-spam filters to detect (Attar, Rad & Atani 2013). Email spam can include different forms of links, such as a uniform resource locator (URL), a clickable link to social websites such as Facebook and YouTube, or a clickable link to spammers' targets, such as fake bank web pages, counterfeit business websites and forged unsubscribe links. Kumar (2009) stated that spammers currently use social network websites such as Facebook to trick users and their friends in order to get their personal information. With that personal information (e.g. email addresses), they spam these email addresses, possibly spread malicious programs such as malware and worms to their computers. Smith (2008) stated that fake YouTube links have been used to download malware onto users' computers when they click them. Email spam has included spoofed links that open fake webpages of banks or popular businesses with the aim of stealing important user information, such as credit card details (Barroso 2007; Leavitt 2005).

A forged or false unsubscribe link is also an example of links included in the body of email spam. The unsubscribe link is an option that enables users to remove their email addresses from mailing lists that they have subscribed to (Allman 2003; Malcolm 2004; Vaile 2004). Spammers have exploited the unsubscribe link by adding a so-called false or spoofed unsubscribe link into the message body (McCusker 2004). One possible reason for spammers to do this is that clicking onto the false unsubscribe link could be an indicator that the email address is valid, which can lead to sending more email spam (Chigona et al. 2005; Lambert 2003; Simpson 2003).

A spoofed unsubscribe link can be a way to add the victims' addresses to spammers' lists. According to Allman (2003), "the unsubscribe link removes you from the list in question, but it also adds your address to another list". Websites that send email spam could use the unsubscribe information to annoy the recipients by distributing email addresses to other spammers, resulting, of course, in receiving more spam (Andaker et al. 2006). Spammers have added false unsubscribe links to open advertisements for some businesses and products (Andaker et al. 2006; Lambert 2003), and others have added a deceptive or inoperative unsubscribe link in spam emails to evade the strict spam laws of countries such as the US and South Africa.

Spammers have also used different types of attachments, such as images and pdf files to advertise products or services (Dhinakaran, Lee & Nagamalai 2007a), because text in the body of the email (the traditional way of spamming) can be blocked by text-based anti-spam filters (Attar, Rad & Atani 2013). Dhinakaran, Lee and Nagamalai (2007b) described the use in spamming in this way of sophisticated tools that had not previously been used without attachments. The sophisticated software can hide the sender's identity, select text messages randomly, identify open relay machines, have mass mailing capability and define the spamming time and duration.

Malicious attachments can be used to achieve nasty and mischievous objectives, and can be key way for spammers to infect users' computers with viruses and malware (Cournane & Hunt 2004; Dantin & Paynter 2005; Hershkop & Stolfo 2004; Moustakas, Ranganathan & Duquenoy 2005; Pallas & Patrikakis 2005; Pfleeger & Bloom 2005; Sorkin 2001). One type of dangerous attachment included in email spam is executable (exe) files, which are mostly used to transfer malicious programs such as worms, viruses and trojans (Nagamalai, D, Dhinakaran, C & Lee, J-K 2010).

2.8.3.4 Using fake or obfuscated email addresses to hide spammers' identities

Using fake or obfuscated email addresses is another favourite ploy of spammers that allows spammers to hide their identities, bypass anti-spam filters and deceive the recipients (Hayati & Potdar 2009). There are many methods for generating fake email addresses. The first of these methods is by using spam software to generate addresses with similar but different formats to that used by the genuine company (Dhinakaran, Jae Kwang & Nagamalai 2009). Examples of this kind of software are Phasma Email Spoofer, Bulk Mailer, Aneima 2.0, Avalanche 3.5, and Euthanasia (Dhinakaran, Lee & Nagamalai 2007a).

Spoofed IP addresses are also popular (Nagamalai, D, Dhinakaran, BC & Lee, JK

2010), and are used to generate spam and denial of service (DoS) attacks without concern about revealing the spammers' identities (Hu & Mao (2007). Li and Hsieh (2006) claimed that "the spammer can use unused IP addresses on the same local area network (LAN) to spoof its source IP address". A study conducted by Krishnamurthy (2006) revealed that 20% of the IP addresses blocked by anti-spam filters were spoofed.

Open proxy servers can be used by spammers to send email spam without revealing their identities. The open relay or proxy is an SMTP server that allows connection between the user and server without the need of authentication (Ramachandran & Feamster 2006). With this method, when email spam is forwarded from the proxy to the recipient the email spam contains the proxy address, not the spammers' address (DeBarr & Wechsler 2010; Levy 2004; Xu et al. 2009). A study by Boneh (2004) revealed that more than 60% of all email spam was sent through open proxy servers. Hoanca (2006) claimed that most of the open relays or proxies were in the US, with a small number of them in China, Korea and other countries. To detect open proxy servers, tools such as Send-Safe have been used by spammers to search for open proxies on the internet (Gansterer et al. 2005). Another method that spammers use is to renting botnets. Using this method, spammers send email spam from multiple computers to avoid spammers blacklist updates and to hide their identities (Eggendorfer 2008). Boneh (2004) indicated that some spammers used false names and untraceable payment methods to buy ISP roaming access. This can hide their identities while they conduct their activities.

The previous paragraphs described and discussed some tricks used by spammers to achieve their purposes. The literature reveals studies that have been conducted to investigate these tricks in different languages and countries. The following paragraphs review these studies.

The US Federal Trade Commission (2003) analysed a collection of 1,000 email spam that was forwarded to the commission by customers, investigate the deceptions used in the header of English email spam, such as misleading subject lines and the 'from' line. The results indicated that 40% of the subjects of spam emails sent to American recipients did not indicate the content of email (false statement).

Dhinakaran, Lee and Nagamalai (2007b) set up a spam trap and analysed the content of 400,000 email spam collected from worldwide spam traffic over a period of 14 months from January 2006 to February 2007. A spam trap (also called a spam honeypot) is a decoy email address that is used for the purpose of collecting email spam (Boneh 2004; Pallas & Patrikakis 2005). The authors claimed that the investigation of characteristics of email spam could help to better understand the features of spam and spammers technology. The results demonstrated that more than 50% of the email spam collection included attachments, the attachments consisted of images and executable files, and most of the images were in the format of .gif and .jpeg files.

Dhinakaran, Jae Kwang and Nagamalai (2009) analysed a collection of 700,000 email spam that have been collected from worldwide spam traffic in a period of 13 months from February 2008 to February 2009. This study aimed to investigate the malicious content of the email spam, such as types of phishing. They also investigated tricks to obfuscate spammers' email addresses. This study found identified two types of phishing attack. The first is to target end users by convincing them to click a link to divulge sensitive information to the phishers' controlled machines. The second is through viruses, malware and trojans, which infect the users' computers, taking them to fraudulent websites to divulge their identity. This study observed many characteristics of the fake or obfuscated email addresses:

- The length of the spammers' email addresses was ranged from 21 characters to 28 characters.
- There were three parts before the "@" symbol: (Word1)(numericvalue)(word1)@forged domain.com.
- Word 1 included a sensitive word, e.g. customer service, support, operator, service-number, operator-id, client service, ref, reference number, customers.
- The length of the second part (numericvalue) of the address ranged from 5 to 12 characters.
- The length of third part was two characters.

Yamakawa and Yoshiura (2010) analysed the headers and bodies of a collection of 134,793 Japanese and English spam mails sent to four email addresses over the period 1 April 2001 to 17 September 2008. The study aimed to investigate attractive words in the subject line, the format (appeared as image or text) and the content (including links and attachments). The researchers found that most words used in English spam were related to commercial advertisements, whereas the sexually loaded words were used most in Japanese email spam. Different types of attachments were included in Japanese email spam, such as pdf files, compressed files, Microsoft Word documents, Microsoft PowerPoint slides, Microsoft Excel sheets, and binary data. Of the attachments contained in Japanese spam, most were pdf files.

Ermakova (2010) analysed and examined a collection of 4,000 English, French, Russian and Italian email spam (about 1,000 spam for each language) received in Russia. This study found some different tricks used in the header and body of email spam in the different languages, particularly Russian. These tricks included substitution of letters with digits, substitution of Cyrillic symbols with a similar Latin letter, and the use of unnecessary symbols and blanks.

This section described the reasons why spammers send email spam, the methods they use to collect recipient email addresses and the tricks they use in the header and body of email spam in different languages and countries. These tricks can affect the effectiveness of anti-spam filters and have been used to bypass anti-spam filters (Nielson, Aycock & de Castro 2008; Zuo et al. 2009). Researchers such as Çıltık and Güngör (2008), Nguyen, Tran and Nguyen (2008) and El-Halees (2009) have found that the effectiveness of anti-spam filters is higher in detecting English email spam than non-English spam. This discussion leads to ask the following questions:

- Do these tricks that are observed in English email spam also appear in Arabic email spam?
- How do the tricks that are used in Arabic email spam differ from tricks used in English email spam?

The literature was reviewed to find studies about tricks used in Arabic email spam, but found none, indicating the need for investigation of this issue. Therefore, this study begins to address this gap by analysing a collection of Arabic and English email spam received in Saudi Arabia. It aims to identify the tricks used by spammers in the header and body of Arabic and English email spam, and how they differ for each language.

2.9 Conclusions

In this chapter, relevant literature was reviewed to establish the theoretical basis for this research. The literature review revealed that, to date, there has been no research investigating:

- the nature of email spam in Saudi Arabia
- the awareness of email users about it and efforts to combat it
- how public users, businesses and ISPs deal with it
- its effects on their performances.

This study will help to fill this gap, and provide some suggestions that could help in mitigating it in Saudi Arabia.

This chapter was divided into eight sections. Section 2.1 described the search strategy used to find the relevant articles for identifying the gap of knowledge. A broad systematic review was conducted using different databases (e.g. IEEE Xplore and ScienceDirect), as they have been used in previous similar studies and are also accessible via the Flinders University library website. Various keywords were used (e.g. “email spam”, “efforts” and participant-related words), as these have also been used in previous studies to access relevant published articles. Inclusion and exclusion criteria were considered in the systematic search. Articles that were selected to be reviewed were published in the years from 1999 to 2011, written in English, and met keyword requirements. As a result, about 92 articles were reviewed to identify the gap of the knowledge for this study.

Section 2.2 described the definitions of email spam produced by other research and studies in different countries. The literature revealed no evidence of previous studies for the definition of email spam in Saudi Arabia. Hence, this research will fill this gap by investigating the definition of email spam by public users, businesses and ISPs.

Section 2.3 of this chapter reviewed previous studies that investigated the awareness and education of users about email spam and anti-spam filters, and the efforts of organisations and governments in some countries in this regard. There was no evidence of previous studies that have investigated the awareness of public users and businesses about email spam and anti-spam filters, and the efforts conducted in Saudi Arabia to inform email users about it. Therefore, this study will fill this gap.

This led to the development of the following question:

> Are public users and businesses aware of email spam and anti-spam filters, what are the sources of their knowledge and how do they define email spam?

Section 2.4 of this chapter reviewed previous studies about the nature of email spam. This section described the volumes of spam received by email users, its languages, its types, and its origins or sources. The literature review found no evidence of previous studies that have investigated the nature of email spam received in Saudi Arabia, and where email spam was sent from. Therefore, this study will seek to fill this gap by asking the following questions:

> What is the volume of email spam received by public users and businesses and blocked by ISPs in Saudi Arabia; in which languages does it occur; and what are the sources or origins of Arabic and English email spam?
>
> What are the differences between Arabic and English email spam?

Section 2.5 of this chapter reviewed many studies about ways in how email users deal with email spam. The literature revealed no previous studies that have investigated how public users, businesses and ISPs in Saudi Arabia deal with email spam. Therefore, this study endeavours to cover this gap of knowledge by asking the following question:

> How do public users, businesses and ISPs deal with email spam?

Section 2.6 of this chapter reviewed the effects of email spam on the performance of users, businesses and ISPs. The literature review found no previous studies that have investigated the effects of email spam on the performance of email users and ISPs in Saudi Arabia. Therefore, this study seeks to fill this gap by considering the following research question:

> What are the effects of email spam on the performance of public users, businesses and ISPs?

Section 2.7 of this chapter reviewed the efforts to combat email spam. These efforts were of legal, technical and other kinds. The literature review found no evidence of the previous studies into the awareness of Saudi society about the efforts of Saudi Arabia to combat email spam. Therefore, this research seeks to fill this gap by investigating the awareness of public users and businesses about the efforts of government and ISPs to combat email spam in Saudi Arabia.

This section also reviewed the anti-spam filters used by the ISPs in different countries, and the effectiveness of these filters in detecting email spam. However, no previous studies were found that have investigated the anti-spam filters used in Saudi Arabia to combat email spam, and their effectiveness in detecting Arabic and English email spam. Therefore, this research fills the gap by investigating the anti-spam filters used by Saudi ISPs, and their effectiveness in detecting Arabic and English email spam as perceived by Saudi ISPs.

To address the gaps identified in this section, the following questions were adopted:

> Are public users and businesses aware of government and ISPs efforts to combat spam in Saudi Arabia?
>
> What anti-spam filters are used by Saudi ISPs to block email spam, and how effective are they in detecting Arabic and English email spam?

Section 2.8 reviewed the literature on spammers’ reasons for sending email spam and the methods they use to collect recipients’ email addresses. Studies have identified a variety of different tricks used in the header and body of email spam to bypass the anti-spam filter and to lure recipients. These included using attractive words or false statements in the subject line, using different formats in writing the content (text or image spam), adding links or attachments into the content, and using fake or obfuscated email addresses. Studies have investigated the tricks used in different countries and languages. However, no previous studies could be found that investigated these tricks in Arabic email spam and the difference between Arabic and English email spam. Therefore, this research will attempt to close this gap by exploring the following questions:

> What is the extent of the following spammers’ tricks used in the headers and bodies of Arabic and English email spam, respectively:
>
> - attractive words or false statements in the subject line
> - texts or texts embedded in images in the content
> - malicious links and attachments, by type
> - fake or obfuscated email addresses.

Chapter 3: Research Methodology

This study aims to understand the nature of email spam in Saudi Arabia, to investigate the awareness of public users and businesses about it and the efforts to combat it, and to provide possible suggestions to mitigate it. The literature review provided background information of recent studies in the field of email spam, the attitudes and experiences of email users with it, and efforts to combat it; however, there was a lack of studies of email spam issues in Saudi Arabia. A knowledge gap was identified and research questions were developed to cover this gap. The purpose of this chapter is to develop a methodology for answering the research questions. This chapter is organised as follows:

- Section 3.1: revisits the research aim, objectives and questions.
- Section 3.2: describes the research approach followed to achieve the research objectives.
- Section 3.3: presents a detailed description of research design, such as methods used to select participants, sample size, and inclusion and exclusion criteria followed in choosing the participants.
- Section 3.4: provides a description of the questionnaire instrument used to collect data, including justification for its use, how the questionnaire items developed, and its validity.
- Section 3.5: describes the purpose and procedures for conducting a pilot study.
- Section 3.6: presents a detailed description of procedures followed to collect data from the participants.
- Section 3.7: describes the independent and dependent variables considered in this study.
- Section 3.8: describes the statistical approaches used to analyse data.
- Section 3.9: provides information about ethical considerations.
- Section 3.10: describes the methodology followed in the analysis of the email spam corpora received from the participants.

- Section 3.11: concludes this chapter.

3.1 Research Aim, Objectives and Questions

The main aim of this research was to investigate the nature of email spam in Saudi Arabia, email users' awareness of it, and the efforts to combat it; and to provide suggestions to mitigate it. In order to meet the aim of the research, a number of objectives were addressed. These objectives are:

- To investigate the awareness of public users and businesses about email spam, anti-spam filters and the efforts to combat it in Saudi Arabia.
- To investigate the nature of email spam (volume, languages and types) received by public users and businesses, and blocked by ISPs.
- To investigate the differences between Arabic and English email spam.
- To investigate how public users, businesses and ISPs deal with email spam.
- To investigate the effects of email spam on the performance of public users, businesses and ISPs.
- To investigate the anti-spam filters used by Saudi ISPs, and their evaluation of the effectiveness of these filters in detecting Arabic and English email spam.
- To investigate the differences between spammers' tricks used in Arabic and English email spam to bypass anti-spam filters.

On the basis of the literature review findings and to achieve the research objectives describe above, the following research questions were developed:

Awareness of, filters for, and efforts to combat email spam

Q1: Are public users and businesses aware of email spam and anti-spam filters, what are the sources of their knowledge and how do they define email spam?

Q2: Are public users and businesses aware of government and ISPs efforts to combat spam in Saudi Arabia?

The nature of email spam

Q3: What is the volume of email spam received by public users and businesses and blocked by ISPs in Saudi Arabia; in which languages does it occur; and what are the sources or origins of Arabic and English email spam?

Q4: What are the differences between Arabic and English email spam?

Dealing with email spam

Q5: How do public users, businesses and ISPs deal with email spam?

The effects of email spam

Q6: What are the effects of email spam on the performance of public users, businesses and ISPs?

Anti-spam filters and their effectiveness in detecting Arabic and English spam

Q7: What anti-spam filters are used by Saudi ISPs to block email spam, and how effective are they in detecting Arabic and English email spam?

Spammers' tricks used in the headers and bodies of Arabic and English email spam

Q8: What is the extent of the following spammers' tricks used in the headers and bodies of Arabic and English email spam, respectively:

- attractive words or false statements in the subject line
- texts or texts embedded in images in the content
- malicious links and attachments, by type
- fake or obfuscated email addresses.

3.2 Research Philosophy

Research philosophy is defined as the development of the research background and knowledge. Research philosophy can be defined with the help of a research paradigm (Saunders & Thornhill 2004). The research paradigm is an important part of the research methodology, as it guides the way the research is conducted (Gliner & Morgan 2000): the researcher's choice of tools, instruments, participants and methods used in the research (Denzin & Lincoln 2000). In the realm of the social science research, there are two major types of research paradigms: quantitative and qualitative (Ponterotto 2005). Quantitative research attempts to understand phenomena by collecting numerical data and using statistical methods to analyse these data (Aliaga & Gunderson 2000). According to Punch (2013) quantitative methods "conceptualises in terms of variables; measures these variables; and studies relationships between these variables". The second type of research paradigm, qualitative, has been defined by Denzin and Lincoln (2009) as "method that can help in understanding how an intervention is experienced, while providing insight into factors which might hinder successful implementation". The quantitative approach is empirical research in which the data are in the form of numbers, the qualitative approach is empirical research in which data are not in the form of numbers (Punch 2013).

In the domain of information security, including email spam, many researchers use quantitative method to answer the research questions. This study also adopted the quantitative approach, of collecting and analysing data in order to determine the relationship between independent variables and dependent variables in the Saudi population. The use of quantitative study enabled the researcher to investigate a larger sample than is possible using qualitative methods. In this way, the sample is large enough to be representative of the whole population being researched, so that the results can be generalised to the entire population (Blessing & Chakrabarti 2009; Krishnaswamy, Sivakumar & Mathirajan 2009).

3.3 Research Design

Research design is an important process in building the structure of research before data collection and analysis. According to de Vaus (2001), "the function of a research design is to ensure that the evidence obtained enables us to answer the initial question as unambiguously as possible". In this study, in order to answer the research questions, quantitative, descriptive and cross-sectional data were collected from the participants during the period February to July 2011. Research design involves a series of decision-making choices regarding the sample type, methods used to collect data, and the measurement and analysis of variables (Cavana, Delahaye & Sekeran 2001). The research activities for this study are described and discussed in the following sections.

3.3.1 Sampling Method and Sample Size

There are many methods to create a population sample for a research study. A population sample is defined as a group of people in a study who represent the total population investigated by the research study. Including the total population in the study is expensive and impractical, so a sample is used to represent it (Thompson 1999). Sampling methods can be categorised into two main types: probability samples and nonprobability samples. The idea of the probability samples is to take a random selection, where the sample represents the target population (Kitchenham & Pfleeger 2002; Teddlie & Yu 2007). Teddlie and Tashakkori (2003) have defined probability samples as:

> … selecting a relatively large number of units from a population, or from specific subgroups (strata) of a population, in a random manner where the probability of inclusion for every member of the population is determinable.

Examples of these samples include simple random sampling, stratified random sampling, systematic sampling and cluster sampling (Teddlie & Yu 2007). Nonprobability samples, on the other hand, are used in cases where the researcher cannot select the kinds of probability used in social surveys. They allow the researchers to involve a larger population without the requirements of random selection (Tansey 2007). Kitchenham and Pfleeger (2002) defined nonprobability samples as “samples are created when respondents are chosen because they are easily accessible or the researchers have some justification for believing that they are representative of the population”. Availability sampling, quota sampling, purposive sampling and snowball sampling are examples of nonprobability samples (Feild et al. 2006).

A major issue when determining sampling methods is achieving an appropriate sample size. Kitchenham and Pfleeger (2002) give two reasons why sample size is important. The first reason is that a small sample size can lead to getting results that are not significant statistically. The second reason is that a small sample size may not allow the researcher to compare different subsets of the population.

This section describes the sampling method used to select samples for this research, and the sample size of the three groups of participants: public users, businesses and ISPs.

3.3.1.1 Public users group

The sampling method used to select public user participants was multiple cluster random sampling, a complex type of cluster random sampling method. Cluster random sampling is a probability method that is used to divide the population into separate groups, called clusters (Kitchenham & Pfleeger 2002; Onwuegbuzie & Collins 2007). According to Hoshaw-Woodard (2001), "in a cluster sample, the population is divided into non-overlapping subpopulations usually based on geographic or political boundaries". Multiple cluster random sampling subdivides larger clusters into small clusters for the purpose of survey (Teddlie & Tashakkori 2003). In Saudi Arabia, there are five regions: eastern, western, central, southern and northern. A total of 1,500 participants from different sectors (e.g. universities and schools) in different cities (e.g. Riyadh and Jeddah) in the five regions were selected randomly for this study. The regional distribution of the 1,500 participants was:

- 300 participants from two cities (Dammam and Alahsa) in the eastern region
- 300 participants from the western region (Jeddah)
- 400 participants from the central region (Riyadh)
- 250 participants from the southern region (Abha)
- 250 participants from the northern region (Hail).

Figure 3.1 shows the total number of samples selected from each region in Saudi Arabia.

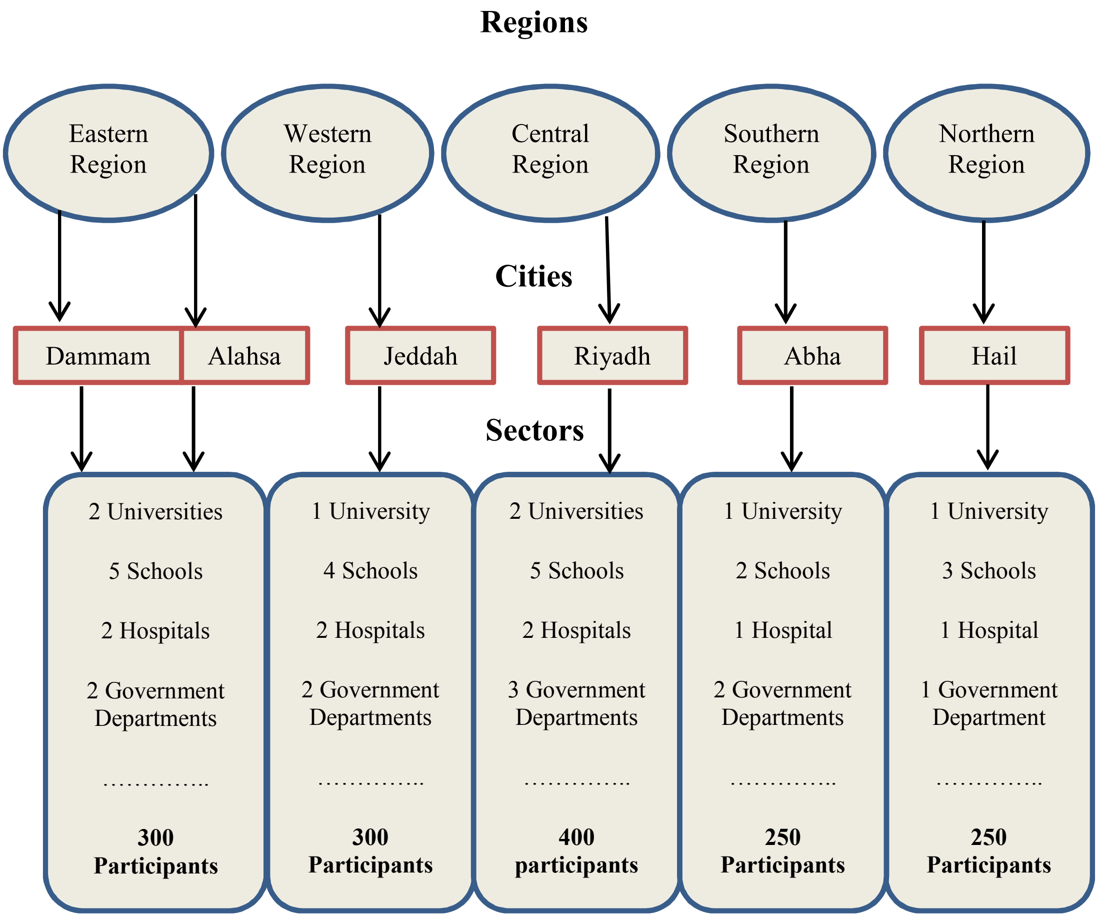


Figure 3.1: The total number of sample size of each region in different places of Saudi Arabia

The number of samples was sufficient for a survey as the same sample size has been used in previous email spam studies in different countries such as the USA, Bahrain, Malaysia and Singapore, some of which have a population similar to or greater than that of Saudi Arabia (Al-A'ali 2007; Bujang & Hussin 2010; Grimes, Hough & Signorella 2007; Hermanson 2003; Leng 2006; Yeh & Chang 2007).

As shown in Figure 3.1, public users were selected from different settings. The data were collected from participants in universities, schools, hospitals and government departments in the five regions of Saudi Arabia. Previous studies have used these sectors to conduct their studies of email spam. The participants of a study conducted by Al-A'ali (2007) about email spam issue in Bahrain were selected from colleges, universities, schools, hospitals and public organisations (Al-A'ali 2007). The participants of a study of email spam conducted in the USA were recruited from academic settings and from non-academic settings such as senior citizen, political organisations, social gathering and workplace settings (Grimes, Hough & Signorella

2007). Bujang and Hussin (2010) conducted a study about email spam in Malaysia, for which the researchers recruited participants from schools, higher education institutions, and different organisations (Bujang & Hussin 2010).

3.3.1.2 Businesses group

Similarly, a multiple cluster random sampling method was adopted to select participants for the business questionnaire. A total of 300 businesses were selected randomly from the eastern (Dammam), western (Jeddah) and central (Riyadh) regions. One hundred businesses were selected randomly from each surveyed city in the three regions. No businesses were selected from the southern and northern regions. These businesses had asked for permission from their head offices to participate in this study, but it was refused, so they requested that the survey be conducted at their head offices in Riyadh, Dammam and Jeddah. The same sample size has been used in past studies of email spam in business users (Siponen & Stucke 2006; Viudes 2011). The distribution of the 92 completed questionnaires for businesses in the three regions was:

- 28 from the eastern region
- 21 from the western region
- 43 from the central region.

3.3.1.3 ISPs group

The availability sampling method was used to select ISP participants. An availability sampling (also known as convenience sampling) is a type of nonprobability sampling method that contains participants who are known to the researcher, or convenient or available to the researcher (Özdemir, St. Louis & Topbaş 2011), and is used to select participants on the basis of accessibility. Teddlie and Yu (2007) defined convenience sampling as "a method that involves drawing samples that are both easily accessible and willing to participate in a study". According to the Communication and Information Technology Commission (2012), 27 ISPs were licensed by the CITC to provide the Internet service in Saudi Arabia. These were located in the eastern (Dammam), western (Jeddah) and central (Riyadh) regions in Saudi Arabia. The 27 ISPs were divided into the three regions as follows:

- 5 ISPs in the eastern region

- 6 ISPs in the western region
- 16 ISPs in the central region.

All 27 ISPs were surveyed for this research.

3.3.2 Inclusion and Exclusion Criteria

This section describes the criteria for inclusion and exclusion of the participants in the research.

This research included participants who were:

- resident (lived or were located) in the eastern, western, central, southern and northern regions
- male and female
- 15 years (high school students) and older
- students and employees
- known email users
- employees and customers of private organisations and government departments
- interested in participating in this study.

In this study, anybody who was not interested in participating or who did not use email was excluded.

3.4 Research Instruments

As this study applied a quantitative approach to answering the research questions, a questionnaire was considered to be the most appropriate method to collect data (Creswell 2013) and is one of the most commonly instruments used in quantitative research (Saunders et al. 2011). The targeted samples of this study were public users, businesses and ISPs in the eastern, western, central, southern and northern regions in Saudi Arabia. In the field of email spam, previous studies have also used the questionnaire to collect data from participants (Al-A'ali 2007; Bujang & Hussin 2010; Chigona et al. 2005; Dantin & Paynter 2005; Grimes, Hough & Signorella 2007; Hermanson 2003; Leng 2006; Pallas & Patrikakis 2005).

3.4.1 Questionnaire Instruments Development

On the basis of literature review findings and also the questionnaires which have been used in previous studies, three different self-administrated questionnaires were designed for three study groups (public users, businesses and ISPs). Some questions included in all three questionnaires were adopted from previous studies, and some questions were developed specifically for this research. In developing new survey questions, this study followed the eight-step approach recommended by Worthington and Whittaker (2006):

> [D]etermine clearly what you want to measure, generate an item pool, determine the format of the measure, have experts review the initial item pool, consider inclusion of validation items, administer items to a development sample, evaluate the items, and optimize scale length.

A public user questionnaire with 33 questions, a business questionnaire with 29 questions and an ISP questionnaire with 33 questions, was developed using this method.

The public user and business questionnaires were divided into three parts: (1) demographic information, (2) email spam, awareness about it, its effects and dealing with it, and (3) the efforts to combat email spam in Saudi Arabia and awareness about them. The ISP questionnaire included the same three parts as public user and business questionnaires, with an additional part about the anti-spam filters used to block email spam, and their effectiveness in detecting Arabic and English email spam. Some questions from the three parts of the ISP questionnaire were also used in the questionnaires from one or both of the public users and business groups (see Appendices B, C, D, E, F and G). The following sub-sections describe the questions for each part in the three different groups of questionnaires.

3.4.1.1 Part 1: Demographic Information

In this part of the public user questionnaire, specific questions were asked about demographic characteristics, whereas most questions in the business and ISP questionnaires were common to the two groups.

3.4.1.1.1 Public user demographic information

To identify the target population and describe the participants, this part of the questionnaire included eight questions (1-8) about aspects of the participant demographic profile: gender, age, nationality, education level, study discipline, work status and work position. They also helped to the researcher understand the participants' perceptions and experiences with email spam, based on demographic factors. Previous studies such as Chigona et al. (2005), Leng (2006), Johnson and Koch (2006), Al-A'ali (2007), Grimes, Hough and Signorella (2007), Bujang and Hussin (2010) and Mohamed (2011) have examined demographic factors such as region, gender, age, nationality, education level, and work status to understand perceptions and attitudes about email spam. Two of the demographic factors (study discipline and work position) have not been used in previous research. The researcher used these two factors to investigate Saudi email users' perceptions of spam.

3.4.1.1.2 Business and ISP demographic information

This Part comprised 10 questions (1-10) in the business questionnaire and nine questions (1-9) in the ISPs questionnaire. The demographic factors (1-5) in the business questionnaire and (1-4) in the ISP questionnaire, which include the establishment year of the organisation, the organisation size, number of employees, number of customers, and organisation sector (the organisation sector is developed specifically for businesses), were to help understand the perceptions and dealing of businesses and ISPs with email spam. These factors have been used in previous studies (Bernik 2013; Chigona et al. 2005; Ramady & Sohail 2010; Yeh & Chang 2007) to understand experiences of businesses and ISPs and how they deal with security issues such as email spam.

Questions 6 to 10 (business questionnaire) and questions five to nine (ISP questionnaire) were general questions about network security in the organisation. These questions asked businesses and ISPs if they established a business unit or team to manage the network security of the organisation, the network security responsibilities of this unit, the number of employees in this unit, and the number of employees in this unit with a specific responsibility to combat spam and their tasks regarding email spam problems. These questions have not been used in previous studies and they were developed for this study to cover the gap of knowledge.

3.4.1.2 Part 2: Questions about email spam, Awareness about it, its Effects, and Dealing with it

Part 2 of the questionnaire was concerned with information about email spam as perceived by public users, businesses and ISPs; their awareness about it, its effects on their performance and how they deal with it. Some questions in this part were common to all three groups of questionnaires, some common to two of them, and some were specific to each group. Part 2 comprises 19 questions (9-27) in the public user questionnaire, 12 questions (11-22) in the business questionnaire, and 12 questions (10-21) in the ISP questionnaire.

At the beginning of this part of the questionnaire, public users, businesses and ISPs were asked to define email spam in their own words (Q.9 "public users", Q.11 "businesses" and Q.10 "ISPs") in order to understand the users' understandings of the definition of email spam. Then the study defined email spam as:

> … an unsolicited, unwanted, commercial or non-commercial email that is sent indiscriminately, directly or indirectly, to a large number of recipients without their permission, and there is no relationship between the recipients and sender.

This definition was provided in the questionnaire as a reference point for the remainder of the questions. To prevent introducing a strong bias, care was taken to ensure that the respondents did not see the supplied definition until after they had supplied their own definition. The variety of responses to the question of what is spam is evidence that this approach was successful. This question was used in a previous study by Pallas and Patrikakis (2005) to investigate the definition of email spam based on participant understanding.

3.4.1.2.1 The awareness of public users and businesses about email spam

Questions 10 and 11 (public user questionnaire), and questions 12 and 13 (business questionnaire) were about the awareness of public users and businesses about email spam and the source of their knowledge about it. These questions were used in previous studies (Al-A'ali 2007; Bujang & Hussin 2010) to investigate the awareness of email users about spam.

3.4.1.2.2 The nature of email spam as perceived by public users, businesses and ISPs

Questions (12-17) in the public user questionnaire, questions (14-19) in the business questionnaire, and questions (11-19) in the ISP questionnaire were about the nature of email spam, such as the number of email spam, the last time public users and businesses received spam, the languages of email spam, types of Arabic and English spam, sources of Arabic and English spam, and keywords and phrases used in spam. These questions have been applied in numerous previous studies (Al-A'ali 2007; Bujang & Hussin 2010; Chigona et al. 2005; Grimes, Hough & Signorella 2007; Leng 2006; Pallas & Patrikakis 2005) to understand the email spam characteristics in different countries such as Bahrain, Malaysia, Singapore, South Africa, USA and Greece.

3.4.1.2.3 Email account providers used by public users, their experiences in using it, and dealing with it

In the public user questionnaire, questions 18 and 19 were about users' email account providers (e.g. Hotmail, Yahoo and Gmail) and their experiences in using it. These questions were adopted from previous studies such as Chigona et al. (2005), Grimes, Hough and Signorella (2007) and Bujang and Hussin (2010). Questions 20 to 22 were about how public users deal with email spam (e.g. read it, delete it, and contact ISP about it); response to spam; and the benefits they derived from spam (i.e. positive impact), such as learning, purchasing and selling, and fun. These questions was used by Hermanson (2003), Chigona et al. (2005), Al-A'ali (2007), Grimes, Hough and Signorella (2007) and Bujang and Hussin (2010) in their studies of how email users dealt with spam.

3.4.1.2.4 Effects of email spam on the performance of public users, businesses and ISPs

Questions 23 and 24 in the public user questionnaire, and question 20 in the business and ISP questionnaires asked the participants about the effects of email spam on their performance. These questions were adopted from other studies such as Chigona et al. (2005), Leng (2006), Grimes, Hough and Signorella (2007), Al-A'ali (2007) and Bujang and Hussin (2010) to investigate the effects of spam on the performance of email users.

3.4.1.2.5 The awareness of public users about anti-spam filters

Questions 25 and 26 in the public user questionnaire were about the awareness of public users of anti-spam filters and the source of their knowledge about these filters. These questions were used by other researchers such as Leng (2006), Al-A'ali (2007) and Bujang and Hussin (2010) to understand email users' awareness of various anti-spam filters.

3.4.1.2.6 The time spent by ISPs to fix email spam problems

Question 21 in the ISP questionnaire asked the participants about the time they spent fixing problems relating to email spam. This question was used by Leng (2006) to investigate the time lost in fixing email spam issues.

3.4.1.2.7 Anti-spam filters used by public users and businesses

Question 27 in the public users' questionnaire, and questions 21 and 22 in business questionnaire, asked the participants if they used anti-spam filters and if so, what was the effectiveness of these filters in detecting Arabic and English email spam. These questions have been used by other researchers such as Pallas and Patrikakis (2005) and Bujang and Hussin (2010) to understand email users' awareness of anti-spam filters and their evaluation of the effectiveness of filters that they used in detecting spam.

3.4.1.3 Part 3: The Anti-spam Filters Used by Saudi ISPs, and their Effectiveness in Detecting Arabic and English Spam

This part was technical in nature and it was developed specifically for the ISP participants. It sought information about the anti-spam filters used by Saudi ISPs to block email spam, and their effectiveness in detecting Arabic and English spam. At the beginning of Part 3, the two major techniques (content and origin-based techniques) used in developing anti-spam filters were defined to help the ISPs answer the questions. Part 3 comprised six questions (22-27) that sought information about types of content and origin-based filters, their effectiveness in detecting Arabic and English email spam, and whether or not the ISPs updated these filters regularly. These questions were developed by previous studies (Chigona et al. 2005; Pallas & Patrikakis 2005; Rossow 2007) to investigate the anti-spam filters used by ISPs and their effectiveness in detecting email spam.

3.4.1.4 Part 3 (Public Users and Businesses) and Part 4 (ISPs): Efforts to Combat Spam in Saudi Arabia and Awareness of them

This part of the questionnaires investigated the efforts conducted in Saudi Arabia to combat email spam, the awareness about them and any possible suggestions the participants think they could help in combating email spam in Saudi Arabia. This part included some questions that were common to all three groups of questionnaires, common to two of them, or specific to each group. This part comprised six questions (28-33) in the public user questionnaire, seven questions in the business questionnaire (23-29), and six questions (28-33) in the ISP questionnaire.

3.4.1.4.1 The awareness of public users, businesses and ISPs about government efforts to combat spam

Questions 28 and 29 (public users), questions 23 and 24 (businesses), and question 28 (ISPs) investigated efforts of the government to combat spam, as perceived by public users, businesses and ISPs; and the awareness of public users and businesses about them. These questions were used in a previous study conducted by Bujang and Hussin (2010) to investigate email spam in Malaysia, the awareness of email users about it, and the possible efforts to combat it.

3.4.1.4.2 The awareness of public users and businesses about ISPs efforts to combat spam

Questions 30 and 31 (public users), and questions 25 and 26 (businesses) seek information about the ISPs' efforts to combat email spam, as perceived by public users and businesses. These questions have not been used in previous studies and they were developed for this study to cover the gap of knowledge.

3.4.1.4.3 The educational efforts of businesses and ISPs to educate their employees and customers about email spam

Question 27 (businesses) and questions 29 to 31 (ISPs) investigated the educational efforts conducted by businesses and ISPs to inform their customers and employees about email spam and effective methods to combat it. These questions were not used in previous studies and they were developed to fill the gap of knowledge.

3.4.1.4.4 Suggestions to combat spam in Saudi Arabia, provided by public users, businesses and ISPs

Questions 32 and 33 (public users), questions 28 and 29 (businesses), and questions

32 and 33 (ISPs) asked the participants to provide appropriate technical, legal or other suggestions (based on their opinions) to combat email spam in Saudi Arabia. These questions were used in previous studies by Hermanson (2003), Chigona et al. (2005) and Bujang and Hussin (2010), in which the participants were asked about their beliefs or opinions about possible solutions to combat spam.

3.4.2 Validity of the Questionnaires

Several components of the three questionnaires (public users, businesses and ISPs) have been used many times in other questionnaires in previous studies in different countries (Al-A'ali 2007; Bujang & Hussin 2010; Chigona et al. 2005; Grimes, Hough & Signorella 2007; Leng 2006; Pallas & Patrikakis 2005).

In this study, the researcher used two types of validity tests: face validity and content validity. Face validity ensures that a test is going to measure what it is supposed to measure (Collis & Hussey 2003). Content validity is the extent to which specific items represent the content domain (Waltz, Strickland & Lenz 2010).

All three questionnaires were developed in English and their components were checked by the researcher's supervisors at Flinders University in Australia and by two faculty members in Saudi Arabia (experts in the field of information security). All three questionnaires were translated into Arabic by the researcher. The Arabic translation of questionnaires was reviewed by an academic faculty member at Flinders University who speaks both Arabic and English, and also by three Saudi academic members. Then the items of three groups of questionnaires and their translation (English into Arabic) were modified and refined according to the feedback and comments provided (Appendices B, C, D, E, F and G).

3.5 Pilot Study

The pilot study is used to identify any potential problems in following the research procedures (Bell 2010; van Teijlingen & Hundley 2002). According to Roberts-Holmes (2011), “the pilot study can alert the researcher to any potential future difficulties and the research can be appropriately amended”. The purpose of the pilot study is to improve the questionnaire and to make sure that participants will not find problems when answering questions (Saunders et al. 2011). The pilot survey ensures that the instrument is understood by the participants and that there are no problems with the wording of the survey (Cavana, Delahaye & Sekeran 2001).

In this study, the original English and Arabic versions of the questionnaires were distributed to a few participants who had characteristics similar to that of the study sample, to check the questionnaire language, the readability, the level of understanding and the time it would take to complete the questionnaire. The participants were informed that these questionnaires were for pilot purposes. The period of pilot questionnaires distribution was from 22 December 2010 to 30 January 2011.

The pilot questionnaire for public users was distributed to 40 public users for comments about the structure of the questionnaire, its questions and the time taken to complete it (20 Arabic questionnaires and 20 English questionnaires). About 29 public users returned the questionnaire, with their comments. As a result, some of the questions that proved to be vague, were revised. The time taken to complete the entire questionnaire ranged from 24 minutes (minimum) to 60 minutes (maximum); an average of 31 minutes each.

The pilot business questionnaire was distributed to 10 businesses (five Arabic and five English) and was returned by all of the businesses with comments, resulting in the revision of some of the questions. Businesses estimated the time taken to complete the entire questionnaire as ranging from 15 minutes (minimum) to 40 minutes (maximum), an average of 26 minutes each.

Three ISPs were selected for the pilot study (two questionnaires in Arabic and one in English). Comments and feedback were provided by all three and some of questions were modified as a result. The ISPs estimated that they took between 20 and 50 minutes to complete the questionnaire, an average of 30 minutes each. Small changes were made to the wording and structure of the questionnaires.

In the pilot study, draft versions of all three questionnaires were discussed with two academic supervisors and one statistical consultant at Flinders University (Australia), and with two faculty members (experts in the field of information security) in Saudi Arabia. Based on their face and content validity, some questions were modified or deleted from the final versions of questionnaires (Appendices B, C, D, E, F and G).

3.6 Procedures for Data Collection

This section describes the procedures for collecting data from participants in the three groups: public users, businesses and ISPs. Data collection began 12 February 2011 and finished on 30 July 2011 (took about 6 months).

3.6.1 Data Collection of the Public User Questionnaire

The period of data collection of the public user questionnaires was two months and 18 days, starting 12 February 2011 and finishing 30 April 2011. The researcher administered the survey in the eastern (Dammam and Alahsa) and central (Riyadh) regions, while three male faculty members, one from each region, administered the surveys for the researcher in the western (Jeddah), southern (Abha) and northern (Hail) regions. Because of cultural issues that segregates males and females in education and work in the Saudi society (Al-Saggaf & Williamson 2004; Alhazmi & Nyland 2010), five female faculty members, one from one university in each region, helped the researcher to collect data from female participants in schools, universities, hospitals and government departments. Because the male and female faculty members who helped the researcher in data collection worked in the same area as the researcher's field of study and had the same level of knowledge, the researcher had no difficulty explaining the purpose of the research and the data collection procedures to them. The researcher explained the process of data collection and with the need for confidentiality of the participants' data.

At the beginning of data collection, the researcher selected universities, schools, hospitals and government departments randomly in the eastern, western, central, southern and northern regions. The government electronic websites of the Ministry of Higher Education, Ministry of Education, Ministry of Health, and the Saudi Government website helped the researcher to select universities, schools, hospitals and government departments in the five regions randomly. In the eastern and central regions, the researcher explained the purpose of the research to potential participants who met the inclusion criteria and invited them to participate. Once they were approved to participate in the research, the researcher selected a random sample of participants from each sector (university, school, hospital, and government department), and provided a hard copy of the questionnaire to the participants with a letter of introduction signed by the research supervisor (Appendices B and C).

The public user questionnaire and a letter of introduction were written in both Arabic and English to enable participants to choose the language they felt most comfortable with. The researcher asked the participants to provide their mobile phone numbers and to hand the completed questionnaires to nominated persons in each sector to receive questionnaires. The researcher gave the participants adequate time (two months and 18 days) to complete the questionnaire and did not pressure them to complete it. Three reminders were sent to the participants to remind them to complete the questionnaire: four weeks after the first visit, three weeks after the first reminder, and three weeks after the second reminder. Participants who did not respond to the last reminder were eliminated from the participation in this research. The academic faculty members (males and females) who supported the researcher in collecting data in the western, southern and northern regions followed the same procedure. The researcher contacted them weekly to check on the progress and to discuss issues they found during the data collection. When data collection was complete, every faculty member put the completed questionnaires in a folder that was coded by the name of the region they collected from and sent the responses the researcher by express mail. This code helped in identifying completed questionnaires for each region.

3.6.2 Data Collection of the Business Questionnaire

The business questionnaire was distributed and collected from the participants over the period from 1 May 2011 and to 30 June 2011. The researcher collected data from businesses located in the eastern and central regions, while an academic faculty member who worked in one of the universities in the western region (Jeddah) collected data from businesses located in the western region. Details about the data collection procedures and the confidentiality of participants' data were explained to the faculty member, and he asked to send the collected completed questionnaires to the researcher by express mail.

Before collecting data from businesses, the researcher randomly selected a sample of businesses in the three regions using the electronic website of the Ministry of Commerce and Industry. The researcher visited businesses in the central and eastern regions, explained the purpose of the research and asked them if they would be like to participate in the survey. Those who approved were given a copy of the questionnaire with a letter of introduction signed by the research supervisor (Appendices D and E).

Both the Arabic and English versions of the business questionnaire were provided and the participants were given the choice of versions. Some participants' businesses asked the researcher to send them the questionnaire by email, and when completed they returned them by the same way. An adequate time was given for businesses to complete the questionnaire, and three reminders were sent to the participants (one reminder every two weeks). The participants, who did not respond after the third reminder, were eliminated from the participation in this study.

3.6.3 Data Collection of the ISP Questionnaire

The data collection for the ISP questionnaire was conducted from 1 July 2011 to 30 July 2011. All ISPs in Saudi Arabia were located in the central, eastern and western regions. There were no ISPs in the southern and northern regions (CITC 2012). The researcher collected the data from ISPs in the eastern and central, while an academic faculty member who worked in one of universities located in the western region (Jeddah) collected data from ISPs in that region. The data collection procedures and the confidentiality of participants' data were explained to the faculty member, and he asked to send the researcher the completed questionnaires collected from the participants by express mail.

All 27 ISPs in the three regions were asked if they would be volunteers in this research. The ISPs considered and all approved the request, after which the questionnaire, with a letter of introduction signed by the research supervisor (Appendices F and G), was distributed. The participants were given the choice of completing the Arabic or English version of the questionnaire. Some ISPs asked the researcher to send the questionnaire to them by email, and when they had completed it they returned them by the same method. An adequate time was given for ISPs to complete the questionnaire, and three reminders sent to the participants, one reminder every one week. Participants who had not responded after the third reminder were eliminated from the participation in this research.

3.7 Variables

In this study, the independent variables were: region, gender, age, nationality, education level, study discipline, work status, work position (public user questionnaire), establishment year of the organisation, organisation size (business and ISP questionnaires), and business sector (business questionnaire).

The dependent variables of this study were awareness about email spam, anti-spam filters and efforts to combat it; the definition, volume, languages of email spam; types and sources of Arabic and English email spam; the email account provider used; dealing with email spam; effects of email spam; anti-spam filters; and their effectiveness in detecting Arabic and English email spam.

3.8 Data Analysis

Different statistical tests were used to analyse the data in this study. Chi-square test (X^2) was used to test the categorical data between independent variables. For small cell sizes (value of any cell is less than 5), Fisher Exact test was used to compare the categorical data between variables. To compare the mean of variables between two unrelated groups (two independent variables), the independent-samples t-test was used. The paired sample t-test was employed to compare the means between two related groups (two dependent variables). One-way ANOVA was used to test if there is a difference in the means of variables between three or more groups. The 95% confidence interval (CI) was calculated to measure the reliability of an estimate of population parameter.

The data for the three groups of questionnaires were cleaned and coded for data entry (see Appendix K). Data analysis was conducted primarily by using the SPSS software (version 18) for Windows. A p-value less than 0.05 ($p<0.05$) was considered statistically significant.

3.9 Ethical Considerations

All ethical requirements were followed by the researcher for the different stages of this research. Consideration of ethical issues is an important step in conducting studies related to human subjects (Babbie 2012). In this study, the guidelines of the Social and Behavioural Research Ethics Committee (SBREC) at Flinders University were followed. A form was signed by the researcher and supervisor, and the questionnaires were submitted to the committee. Letters of introduction for the public user, business and ISP questionnaires (in both Arabic and English) were attached with the form and the questionnaires and sent to the committee. A letter of introduction included: the researcher's name; the school's name; the title and purpose of the research; the time expected to complete the questionnaire; a brief statement that the information provided by participants would be kept confidential and they were free to discontinue their participation at any time or to decline to answer particular questions; the supervisor's name and signature; and the contact details for more information about the research (Appendices B, C, D, E, F and G). This research was approved by the committee (A copy of the letter of Ethical Approval is attached in the Appendix A).

Permission to conduct public user questionnaire was obtained from the deans of research in the universities, the headmasters of schools, the directors of hospitals, and the directors of government departments in the five regions of Saudi Arabia. The directors of businesses and ISPs, which were located in the eastern, western and central regions, were contacted by email to obtain their permission to conduct the questionnaire in their organisations. The participants (public users, businesses and ISPs) were asked if they would be volunteers in this research. They were informed that none of them would be individually identifiable, the information provided would be kept confidential and they were free to discontinue their participation at any time or to decline to answer particular questions.

3.10 Methodology Followed in the Analysis of Arabic, English and Mixed Email Spam Corpora

One of objectives of this research was to investigate the differences between the spammers' tricks used in Arabic and English emails spam to bypass anti-spam filters. The literature review found many studies that have analysed the headers and bodies of email spam corpora received in different countries, in different languages, and which the kinds of tricks spammer employed to deceive the filters. Those tricks included the use of attractive words or false statements in the subject line, different formats in writing the content (e.g. text, text embedded in an image), adding malicious links or attachments into the content of email spam, and using fake or obfuscated email addresses to hide identities. However, no previous studies could be found that investigated the tricks used in the headers and bodies of Arabic email spam and how they are different from those used in English email spam. To address this gap, the following research questions were developed:

Q8: What is the extent of the following spammers' tricks used in the headers and bodies of Arabic and English email spam, respectively:

- **attractive words or false statements in the subject line**

This question was used in previous studies, such as that by the Federal Trade Commission (2003), to investigate false statements in the subject line of English email spam corpora received in the USA, and a study by Yamakawa and Yoshiura (2010) of a collection of English and Japanese email spam received in Japan.

- **texts or texts embedded in images in the content**

This question has been used by other researchers such as Dhinakaran, Lee and Nagamalai (2007a) and Yamakawa and Yoshiura (2010) to investigate the format of the content of email spam.

- **malicious links and attachments, by type**

This question has been used in previous studies such as Dhinakaran, Lee and Nagamalai (2007a), Cova, Kruegel and Vigna (2008) and Yamakawa and Yoshiura (2010).

- **fake or obfuscated email addresses.**

This question was used in studies conducted by Dhinakaran, Lee and Nagamalai (2007a), Ermakova (2010), and Yamakawa and Yoshiura (2010).

3.10.1 Analysis Criteria

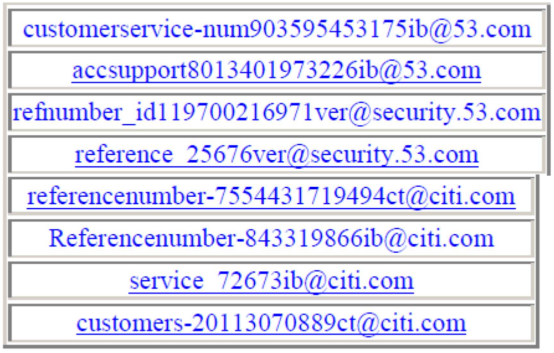
Different criteria have been used in the analysis of headers and bodies of Arabic and English email spam to investigate spammers' tricks. For tricks used in the subject line of email spam, spammers add attractive words and false statements to lure the reader. Studies reported in the literature (Chen, Zhan & Li 2010; Dhinakaran, Jae Kwang & Nagamalai 2009; Smith 2008; Wang & Chen 2007; Wei et al. 2008) have identified various types of attractive words in the subject line of English email spam, such as business advertisements, pornography and phishing. Examples include: “Alert”, “For sale”, “Sex”, “Re”, “Hi”, and “Update details”. In Arabic email spam, attractive words such as “Games”, “Chat”, “Forums”, “Photos”, and “Music” were identified by Goweder et al. (2008), and Wahsheh, Alsmadi and Al-Kabi (2012), among others. Therefore, any email spam that contained any of the attractive words used in these previous studies was classified as email spam with attractive words in the subject line.

False statements were defined by previous studies such as Hamel (2004) and Simon (2004) as subject lines that do not indicate the content of the email (or misleading subject line). Based on this definition, Arabic and English email spam were classified as email spam with false statement in the subject line. This process was achieved by reading the subject line of email spam and then comparing it with the content of email to see if the subject indicated its content or was used to trick the recipients.

Regarding fake email addresses, many characteristics have been identified in the fake or obfuscated email addresses of spam. According to Dhinakaran, Jae Kwang and Nagamalai (2009):

> … the length of the sender account is always more than 21 characters and up to 28 characters. It has three parts before the “@” symbol. The format of the senders mail account is: (Word1)(numericvalue)(word1)@forged domain.com.

This criterion has been used in this study to classify Arabic and English email spam as emails sent from fake or obfuscated email addresses. Examples of fake or obfuscated email addresses used by spammers can be seen in Figure 3.2.



customerservice-num903595453175ib@53.com

accsupport8013401973226ib@53.com

refnumber_id119700216971ver@security.53.com

reference_25676ver@security.53.com

referencenumber-7554431719494ct@citi.com

Referencenumber-843319866ib@citi.com

service_72673ib@citi.com

customers-20113070889ct@citi.com

Figure 3.2: Spammer' IDs syntax (Dhinakaran, Jae Kwang & Nagamalai 2009)

For tricks in the body of email spam, spammers use image spam (texts embedded in images) to escape text-based filters (Attar, Rad & Atani 2013). "Image spam" was defined by other researchers such as Xu et al. (2009) and Soranamageswari and Meena (2010) as the text content of the email appeared as an image in the body of message. On the basis of this definition, any Arabic or English email spam that contained an image spam in the body of message was considered as text embedded in image.

Spammers also add links and attachments to email spam. Studies such as Dhinakaran, Lee and Nagamalai (2007a) and Yamakawa and Yoshiura (2010) identified various types of attachments included in email spam – images in different formats such as Joint Photographic Experts Group (jpeg), Graphics Interchange Format (gif), Portable Document Format (pdf) files, executable (exe) files and text (txt) files. These formats were used in this study to classify types of attachments observed in Arabic and English email spam received in Saudi Arabia.

Another way spammers encourage users to interact with their emails is to include links in the content. Dhinakaran, Lee and Nagamalai (2007a) and others classified these clickable links to a website as Uniform Resource Locators (URLs). This category was also used in this study to classify Arabic and English email spam. Any Arabic and English email content that included a URL or clickable link was classified as an email with links.

Links and attachments can be malicious. Malicious content is included in email spam as attachments includes viruses, worms or trojans; and links that redirect the user to spammers' or phishers' websites or run malicious codes (Dhinakaran, Lee &

Nagamalai 2007a; Nagamalai, D, Dhinakaran, C & Lee, J-K 2010). This definition was used to classify malicious emails in Arabic and English email spam.

As demonstrated by previous studies such as Allman (2003), Smith (2008), Leavitt (2005), Barroso (2007) and Attar, Rad and Atani (2013), most malicious links included in email spam were fake bank webpages and forged unsubscribe links. Fake bank web pages can steal important information, such as credit card details (Barroso 2007). Forged unsubscribe links can redirect email users to spammers' websites (Dhinakaran, Lee & Nagamalai 2007a) or could run malicious programs in the computer (John et al. 2009). Therefore, all these criteria were considered in the classification of Arabic and English email spam into these two types of malicious links.

3.10.2 Procedures for Collection and Analysis of Arabic, English and Mixed Email Spam

During the collection of public user, business and ISP questionnaires, the researcher explained to participants the purpose of collecting Arabic and English email spam and asked if they would be able to assist in this process. The purpose and consent request in Arabic and English were also explained at the end of each questionnaire (Appendices B, C, D, E, F and G respectively). If the participants reported receiving Arabic and English email spam (i.e. email spam that bypassed or was not detected by anti-spam filters) and agreed to assist in this process, they were asked to forward it to a specific email address created for the purpose of this research. These spam corpora were collected in the period from 16 February 2011 to 10 June 10 2013. A collection of 1,937 email spam was received from the participants, who comprised public users, businesses and ISPs. Because this research focused on the investigation of Arabic and English emails spam only, about 160 email spams were excluded because they were written in an unknown language. Some of the spams forwarded were repetitive (total of 507 email spam), so they were also excluded. The remaining 1,270 email spams were analysed to achieve the purposes of this research. The 1,270 email spam included: 1,035 Arabic spam, 179 English spam, and 56 mixed language (Arabic and English) spam.

The analysis of email spam corpora (1,270 spam) was conducted manually, using a checklist. This checklist included 1,270 rows, which represent the email spam ID (from 1 to 1,270); and 9 columns, which represent the language of the email spam and the spammers' tricks. Column 1 listed the language of the email spam using one of three values: 1 (Arabic), 2 (English) and 3 (mixed).

Columns 2 to 9 listed the various tricks identified. Column 2, using attractive words and false statements in the subject line of email spam, used two values: 1 (attractive words) and 2 (false statements). Column 3, using different formats in writing content of email spam, used two values: 1 (text) and 2 (text embedded in an image). Column 4 identified whether links or attachments were added to the content of the email spam, using two values: 1 (links) and 2 (attachments). Column 5 identified the types of the attachments, using four values: 1 (images), 2 (PDF files), 3 (text files) and 4 (executable files). Column 6 designated whether or not the attachments were malicious, using two values: 1 (yes) and 2 (no). Column 7 identified if malicious links were used, using two values: 1 (yes) and 2 (no). Column 8 identified types of malicious links, using two values: 1 (fake bank's website link) and 2 (forged unsubscribe link). Column 9, hiding or obfuscating email addresses/identity, used two values: 1 (yes) and 2 (no). After completing the checklist, the values of each trick for different languages (Arabic, English and mixed) were calculated and analysed using SPSS.

A special computer was used solely to analyse the collected spam, so that malicious content could not cause problems with other programs or documents on the computer. The Internet security software, Kaspersky 2013, was used during the analysis to protect from any potential malicious links or attachments in Arabic and English spam.

When Arabic, English and mixed email spam corpora were classified on the basis of the analysis criteria described above, the statistical chi-square test (X^2) was used to test the categorical data between variables, to detect significant differences between the spammers' tricks used in Arabic, English and mixed email spam. The analysis was conducted by using SPSS software (version 18) for Windows. A p-value that was less than 0.05 ($p<0.05$) considered statistically significant.

3.10.3 Ethical Considerations

Ethical requirements were considered in the collection of Arabic, English and mixed email spam from the participants (public users, businesses and ISPs). The collection of email spam corpora was approved by the Social and Behavioural Research Ethics Committee (SBREC) at Flinders University (Appendix A) and consent was obtained from the public user, business and ISP participants. The participants were asked if they would be willing to assist in this research by forwarding Arabic and English email spam that they received to a specific email address created especially for this study. The approval request (in both Arabic and English) appeared on the last page of the questionnaires for public users, businesses and ISPs (Appendices B, C, D, E, F and G respectively).

3.11 Conclusions

This chapter described the methodology employed to answer the research questions. It began by revisiting the aim, objectives and research questions, and describing the research philosophy and sampling methods used to select the participants. The inclusion and exclusion criteria were specified and the research instrument was explained. The validity of the questionnaires used to collect data was discussed and the procedures for the pilot study and data collection were explained. Variables, data analysis and management, and ethical considerations were discussed, as well as the methodology followed in the analysis of Arabic, English and mixed email spam corpora.

The next three chapters (4, 5 and 6) will provide the results of the questionnaires administered to the three groups of participants (public users, businesses and ISPs), to address the research objectives and questions.

Chapter 4: Public Email Users' Experiences with Email Spam in Saudi Arabia

In this study, public users included email users from universities, schools, hospitals and government departments in Saudi Arabia. This chapter presents the results of the survey of public users about their perception about email spam, their awareness of anti-spam filters, and their awareness of government and ISPs efforts to combat spam. This chapter also describes how public users dealt with email spam and its effects on their performance. The results, based on some demographic factors, are analysed and discussed.

This chapter is divided into the following sections:

- Section 4.1: presents the results of the public users' questionnaire.
- Section 4.2: discusses the results of the public users' questionnaire.
- Section 4.3: describes the conclusions of this chapter.

4.1 Results

This section describes the demographic characteristics of Saudi Arabian public email users, their awareness about email spam, anti-spam filters, and the efforts to combat it. This section presents the results for the nature of Arabic and English email spam as perceived by public users. It also describes how public users dealt with email spam and its effects on their performance.

This section also analyses and presents the results based on demographic factors such as region, gender, age, nationality, education level, study discipline, work status, and work position.

Different statistical tests were used to analyse the data, including chi-square test (X^2), independent samples t-test, paired sample t-test and one-way ANOVA test. A p-value less than 0.05 ($p<0.05$) was considered statistically significant. More details about these tests are described in section 3.8.

4.1.1 Participants' Demographic Information

The distribution of public email users on the basis of region, gender, age, nationality, education level, study discipline, work status, and work position is shown in Table 4.1.

Overall, 1,020 participants from five regions of Saudi Arabia participated in this study as public email users. Public email users from the central region had the highest percentage of participation (34.5%), and participants from the northern region had the lowest frequency of participation (12.7%).

About 60% of the participants were male, while 40.4% were female. Most of the participants (45.4%) were aged from 15-25 years and only 4.4% were aged 46 and older.

Of the participants, 83.8% were Saudi nationals and 16.2% were non-Saudis. The results showed that about 58% of the participants had completed a bachelor degree, whereas 4.8% had completed a diploma degree.

Most of the educated participants (28.7%) had studied computer science and information technology, a few (9.1%) had studied social sciences and physical and biological sciences. About 55.4% of the public email users were employed, most of them in positions within educational institutions (26.9%), and 44.6% were students.

Table 4.1: Percentages of distribution of public users in Saudi Arabia based on their demographic information

General Information	Frequency	Percentage (%)
Region		
Eastern	203	19.9
Western	201	19.7
Central	352	34.5
Southern	134	13.1
Northern	130	12.7
Gender		
Male	608	59.6
Female	412	40.4
Age		
15-25	463	45.4
26-35	358	35.1
36-45	154	15.1
46+	45	4.4
Nationality		
Saudi	855	83.8
Non-Saudi	165	16.2
Education Level		
High School	145	14.2

General Information	Frequency	Percentage (%)
Diploma	49	4.8
Bachelor	588	57.6
Master	144	14.1
PhD	94	9.2
Study Discipline		
Education and Teaching	159	15.6
Computer Science and Information Technology	293	28.7
Social Sciences	93	9.1
Physical and Biological Sciences	93	9.1
Health Sciences and Medicine	88	8.6
Other[1]	149	14.6
Work Status		
Student	455	44.6
Employed	565	55.4
Work Position		
Educational	274	26.9
Medical	58	5.7
Technical	91	8.9
Management	97	9.5
Other[2]	45	4.4

4.1.2 The Awareness of Public Users about Email Spam, Anti-spam filters, and the Efforts to Combat it in Saudi Arabia

Table 4.2 summarises the awareness of public email users about email spam, anti-spam filters, and the efforts to combat it. Approximately two-thirds of the participants (62%, 95%CI[3]: 59%-64.9%) were aware of email spam before participating in the survey and one-third of the public email users (37.9%, 95%CI: 35%-40.9%) were aware of anti-spam filters.

The participants had become aware of email spam and anti-spam filters through a number of sources. The most common channel was the Internet and forums (38.8%, 95%CI: 35.9%-41.8%), although a few of the participants believed that there was a government effort (2.4%, 95%CI: 1.6%-3.4%) to inform them.

Categorising public users' definitions of email spam revealed a variety of descriptions. Most public users (33.9%, 95%CI: 29.5%-38.4%) defined email spam as "email that was sent randomly and contained malicious programs such as viruses",

[1] Other disciplines included: accounting, psychology, Islamic studies, mathematics, marketing agriculture, law, and commerce.

[2] The other employment positions were in academia, administration, banking, transport, research, information systems analysis, engineering and laboratory work.

[3] 95% Confidence Interval

while the lowest percentage of participants (3.9%, 95%CI: 2.4%-6.1%) defined email spam as "annoying email that was not related to recipients' work".

Nearly a quarter of the participants (24.4%, 95%CI: 21.9%-27.1%) were aware of government efforts to combat email spam. The highest percentage of public users (14.9%, 95%CI: 10.9%-19.7%) thought that most of the government's efforts to combat spam were technical, conducted by the Communication and Information Technology Commission (CITC) and King Abdulaziz City for Science and Technology (KACST) sectors, which are responsible for information technology and communication in Saudi Arabia. A few participants (13.6%, 95%CI: 11.6%-15.8%) were also aware of the ISPs' efforts to combat email spam. Public users believed the ISPs' efforts to combat email spam to be were mainly by the use of anti-spam filters (15.1%, 95%CI: 9.9%-21.8%).

Table 4.2: Percentages of distribution of the awareness of public email users about email spam, anti-spam filters, and efforts to combat it in Saudi Arabia

Question	Frequency	Percentage (%)	95% CI
Awareness of email spam			
Yes	632	62	59-64.9
No	388	38	35.1-41
Knowledge source about email spam			
ISPs	52	5.1	3.9-6.6
Internet and forums	396	38.8	35.9-41.8
Broadcast media, e.g. TV	88	8.6	7-10.5
Government	24	2.4	1.6-3.4
School or university education	241	23.6	21.1-26.3
Definition of email spam			
UBE	98	22.7	19-26.9
Sent by unknown senders without recipient's permission	103	23.9	20.1-28.1
Sent randomly, contain malicious programs, e.g. viruses	146	33.9	29.5-38.4
UCE	67	15.5	12.4-19.2
Annoying email unrelated to recipients' work	17	3.9	2.4-6.1
Awareness about anti-spam filters			
Yes	387	37.9	35-40.9
No	633	62.1	59.1-65
Knowledge source for anti-spam filters			
ISPs	25	2.5	1.6-3.5
Internet and forums	254	24.9	22.3-27.6
Broadcast media, e.g. TV	22	2.2	1.4-3.2
Government	14	1.4	0.8-2.2
School or university education	154	15.1	13-17.4
Awareness of government efforts to			

Question	Frequency	Percentage (%)	95% CI
combat spam			
Yes	249	24.4	21.9-27.1
No	771	75.6	72.9-78.1
Government efforts to combat spam			
Technical efforts by CITC and KACST	37	14.9	10.9-19.7
Awareness efforts by CITC	13	5.2	3-8.5
Receiving ISPs' reports about spam issues	19	7.6	4.8-11.4
Awareness of ISPs efforts to combat spam			
Yes	139	13.6	11.6-15.8
No	881	86.4	84.2-88.4
ISPs' efforts to combat spam			
Using anti-spam filters	21	15.1	9.9-21.8
Providing awareness information	6	4.3	1.8-8.7
Reporting spam-related issues to CITC	4	2.9	1-6.7

4.1.2.1 Public users' awareness of email spam, anti-spam filters, and the efforts to combat it, by geographic region

As shown in Table 4.3, there were significant differences in the awareness of public users from the different regions about email spam, anti-spam filters and the efforts to combat it.

Public users in the central region (71.9%) were more aware of email spam than those in other regions (p<0.001). The western region had the highest percentage of public users (52.7%) who knew about email spam from the Internet and forums than did other regions (p<0.001), as well as a higher percentage of public users (14.9%) who were informed about email spam by broadcast media than other regions (p=0.001). The central region, on the other hand, had more public users (29.5%) who were informed about email spam by school or university education than did other regions (p=0.006).

In all of the regions, most public users defined email spam as "email sent randomly and containing malicious programs such as viruses". This definition was used mostly in the eastern and western regions (43.4% and 43% respectively), and least in the southern region (21.7%, p=0.016).

Public users in the eastern and western regions (38.4% and 38.3% respectively) were more aware of anti-spam filters than users in other regions (p=0.01). Public users in the western region (30.3%) were informed about anti-spam filters through the Internet and forums more than public users in other regions (p=0.017), while users in the central region (20.5%) had a knowledge about anti-spam filters by school or university education more than users in other regions (p=0.005).

The results revealed that users in the central region (30.1%) were more aware of the government efforts to combat email spam than users in other regions (p=0.04).

Table 4.3: Percentages of distribution of the awareness of public users about email spam, anti-spam filters, and efforts to combat it based on the geographic region

	Region					P*
Question	**E**	**W**	**C**	**S**	**N**	
	n=203	**n=201**	**n=352**	**n=134**	**n=130**	
Awareness of email spam (%YES)	57.1	69.7	71.9	56	36.9	**<0.001**
Knowledge source for email spam						
ISPs	4.9	5	4.5	7.5	4.6	0.763
Internet and forums	38.4	52.7	42.6	28.4	18.5	**<0.001**
Broadcast media, e.g. TV	5.9	14.9	9.7	6	3.1	**0.001**
Government	3.4	0	2.8	2.2	3.1	0.160
School or university education	21.7	20.4	29.5	24.6	14.6	**0.006**
Definition of email spam						
UBE	23.7	16.8	25.8	28.3	19.1	**0.016**
Sent by unknown senders without recipients' permission	18.4	25.2	23.2	39.1	17	
Sent randomly, contain malicious programs, e.g. viruses	43.4	43	27.1	21.7	31.9	
UCE	9.2	13.1	20	8.7	23.4	
Annoying email unrelated to recipients' work	5.3	1.9	3.9	2.2	8.5	
Awareness of anti-spam filters (%YES)	38.4	38.3	34.8	31.3	27.7	**0.01**
Knowledge source for anti-spam filters						
ISPs	1.5	3	2.6	3	2.3	0.873
Internet and forums	25.6	30.3	27	16.4	18.5	**0.017**
Broadcast media, e.g. TV	2.5	1	3.4	1.5	0.8	0.243
Government	2.5	0	1.4	0	3.1	0.055
School or university education	12.8	10.4	20.5	16.4	10	**0.005**
Awareness of government efforts to combat spam (%YES)	20.2	22.4	30.1	20.1	23.1	**0.04**

Question	**Region**					**P***
	E **n=203**	**W** **n=201**	**C** **n=352**	**S** **n=134**	**N** **n=130**	
Awareness of ISP efforts to combat spam (%YES)	11.3	14.9	15.9	12.7	10	0.366

*P values are based on chi-square test between public users in different regions; P values < 0.05 were considered statistically significant.
Abbreviations: E = Eastern, W = Western, C = Central, S = Southern, and N = Northern.

4.1.2.2 The awareness of public users about email spam, anti-spam filters, and the efforts to combat it, based on gender

As shown in Table 4.4, there were significant differences in the awareness of email spam, anti-spam filters and efforts to combat it among males and females. Males were more aware of email spam than females (66.8% vs 54.9%, p<0.001). Most of the male participants (41.9%) mentioned that they learnt about email spam from the Internet and forums, and this percentage was higher than that observed in females (34.2%, p=0.013).

Significantly more females than males defined email spam as "email was sent from unknown senders and without recipients' permission to receive it" (33.7% vs 21.1%, p=0.034). The reverse was true for (17.3% vs 9.5%, p=0.034) definitions of email spam as UCE.

Significantly more males were aware of anti-spam filters than females (41.6% vs 32.5%, p=0.003). Significantly more males than females learnt about anti-spam filters through ISPs, broadcast media, and school or university education (ISPs: 3.6% vs 0.7%, p=0.003; broadcast media: 3.3% vs 0.5%, p=0.002; school or university education: 18.9% vs 9.5%, p<0.001).

Table 4.4: Percentages of distribution of the awareness of public users about email spam, anti-spam filters, and efforts to combat it based on gender

Question	**Gender**		**P***
	Male **n=608**	**Female** **n=412**	
Awareness of email spam (%YES)	66.8	54.9	**<0.001**
Knowledge source for email spam			
ISPs	7.7	1.2	**<0.001**
Internet and forums	41.9	34.2	**0.013**
Broadcast media, e.g. TV	11.2	4.9	**<0.001**
Government	3.3	1	**0.012**
School or university education	28.3	16.7	**<0.001**

Question	Gender		P*
	Male n=608	**Female n=412**	
Definition of email spam			
UBE	22.9	22.1	**0.034**
Sent by unknown senders without recipients' permission	21.1	33.7	
Sent randomly, contain malicious programs, e.g. viruses	33.9	33.7	
UCE	17.3	9.5	
Annoying email unrelated to recipients' work	4.8	1.1	
Awareness of anti-spam filters (%YES)	41.6	32.5	**0.003**
Knowledge source for anti-spam filters			
ISPs	3.6	0.7	**0.003**
Internet and forums	26.2	23.1	0.262
Broadcast media, e.g. TV	3.3	0.5	**0.002**
Government	1.6	1	0.364
School or university education	18.9	9.5	**<0.001**
Awareness of government efforts to combat spam (%YES)	23.4	26	0.340
Awareness of ISP efforts to combat spam (%YES)	13.5	13.8	0.874

*P values are based on chi-square test between male and female users; P values <0.05 were considered statistically significant.

4.1.2.3 The awareness of public users about email spam, anti-spam filters, and the efforts to combat it, based on age groups

The data in Table 4.5 reveals that most of participants in each age group were aware of email spam, but there were no significant differences between age groups in their awareness of email spam and anti-spam filters. Public users aged 26-35 were more aware of email spam from school and university education than public users in other age groups (31%, p<0.001).

There were significant differences in the way the various age groups defined email spam. Most of the younger users (40.7%) defined it as "email that was sent randomly and contain malicious programs such as viruses", whereas most of the older users (23.1%) defined spam as annoying email that was unrelated to the recipients' work (p=0.001).

The older public users (aged 46 and older) were more aware of government (40%, p<0.001) and ISPs efforts (28.9%, p<0.001) to combat email spam than other age groups.

Table 4.5: Percentage of distribution of the awareness of public users about email spam, anti-spam filters, and efforts to combat it based on age group

Question	Age Groups				P*
	15-25 n=463	26-35 n=358	36-45 n=154	46+ n=45	
Awareness about email spam (%YES)	58.5	62.6	68.8	68.9	0.095
Knowledge source about email spam					
ISPs	4.1	5.6	6.5	6.7	0.578
Internet and forums	41	36.6	35.7	44.4	0.402
Broadcast media, e.g. TV	10.4	6.4	9.7	4.4	0.156
Government	1.7	2.8	3.2	2.2	0.653
School or university education	16.6	31	29.9	15.6	**<0.001**
Definition of email spam					
UBE	18.6	25.9	29.2	23.1	**0.001**
Sent by unknown senders, without recipients' permission	20.1	25.9	33.3	23.1	
Sent randomly, contain malicious programs, e.g. viruses	40.7	30.1	22.9	15.4	
UCE	17.6	15.7	6.3	15.4	
Annoying email unrelated to recipients' work	2.9	2.4	8.3	23.1	
Awareness about anti-spam filters (%YES)	34.3	39.9	42.9	42.2	0.165
Knowledge source for anti-spam filters					
ISPs	1.9	2.2	1.9	11.1	**0.002**
Internet and forums	25.1	23.7	26	28.9	0.864
Broadcast media, e.g. TV	1.5	3.1	1.9	2.2	0.500
Government	1.1	1.4	1.9	2.2	0.823
School or university education	8.4	20.7	22.1	15.6	**<0.001**
Awareness of government efforts to combat spam (%YES)	20.5	22.9	35.1	40	**<0.001**
Awareness of ISPs efforts to combat spam (%YES)	11.2	11.2	22.1	28.9	**<0.001**

*P values are based on chi-square test between public users in different age groups; P values < 0.05 were considered statistically significant.

4.1.2.4 The awareness of public users about email spam, anti-spam filters, and the efforts to combat it, based on nationality

Table 4.6 shows that there were significant differences between Saudis and non-Saudis in their awareness of email spam and anti-spam filters. Non-Saudis were more aware of email spam than Saudis (70.3% vs 60.4%, p=0.016) and also more aware of anti-spam filters than Saudis (49.7% vs 35.7%, p=0.001).

Significantly more non-Saudis than Saudis learnt about email spam and anti-spam filters from school or university education (35.8% vs 21.3%, and 27.3% vs 12.7% respectively, p<0.001). They were also more aware of the government and ISPs efforts to combat email spam than Saudis (40% vs 21.4%, and 24.8% vs 11.5%, p<0.001).

Table 4.6: Percentages of distribution of the awareness of public users about email spam, anti-spam filters, and efforts to combat it based on nationality

Question	Nationality		P*
	Saudi n=855	Non-Saudi n=165	
Awareness of email spam (%YES)	60.4	70.3	**0.016**
Knowledge source for email spam			
ISPs	4.8	6.7	0.317
Internet and forums	40.1	32.1	0.054
Broadcast media, TV	9.1	6.1	0.200
Government	2.7	0.6	0.106
School or university education	21.3	35.8	**<0.001**
Definition of email spam			
UBE	22.8	22.4	0.969
Sent by unknown senders without recipients' permission	23.9	24.1	
Sent randomly, contain malicious programs, e.g. viruses	34.3	31	
UCE	15.3	17.2	
Annoying email unrelated to recipients' work	3.8	5.2	
Awareness of anti-spam filters (%YES)	35.7	49.7	**0.001**
Knowledge source for anti-spam filters			
ISPs	2.5	2.4	0.981
Internet and forums	24.4	27.3	0.442
Broadcast media, e.g. TV	2	3	0.399
Government	1.5	0.6	0.355
School or university education	12.7	27.3	**<0.001**
Awareness of government efforts to combat spam (%YES)	21.4	40	**<0.001**
Awareness of ISP efforts to combat spam (%YES)	11.5	24.8	**<0.001**

*P values are based on chi-square test between Saudi and non-Saudi users; P values <0.05 were considered statistically significant.

4.1.2.5 The awareness of public users about email spam, anti-spam filters, and the efforts to combat it based on education level

Table 4.7 shows that there were significant differences among users with different education levels in their awareness about email spam, anti-spam filters and efforts to combat it. Users with a PhD were more aware of email spam than users with other education degrees (77.7%, p<0.001), and users with a diploma degree were more aware of anti-spam filters than those with other education degrees (55.1%, p=0.001).

There were significant differences in the definitions given to email spam by users with different levels of education. More PhD users (39.5%) defined spam as UBE than did users with other degrees (p=0.034).

Users who had completed diplomas and PhDs were more aware of ISPs efforts to combat email spam than users with other degrees (20.4% and 20.2% respectively, p=0.017).

Table 4.7: Percentages of distribution of the awareness of public users about email spam, anti-spam filters, and efforts to combat it based on education level

Question	**Education Level**					**P***
	HS **n=145**	**D** **n=49**	**B** **n=588**	**M** **n=144**	**PhD** **n=94**	
Awareness of email spam (%YES)	49	67.3	60.5	68.8	77.7	**<0.001**
Knowledge source for email spam						
ISPs	4.8	6.1	3.9	5.6	11.7	**0.035**
Internet and forums	36.6	36.7	39.3	35.4	45.7	0.548
Broadcast media, e.g.TV	9	6.1	9.4	6.3	8.5	0.765
Government	0	4.1	2.2	4.9	2.1	0.086
School or university education	5.5	38.8	24	35.4	23.4	**<0.001**
Definition of email spam						
UBE	13.2	23.5	20.1	29.2	39.5	**0.034**
Sent by unknown senders without recipients' permission	28.3	23.5	21.8	25	28.9	
Sent randomly, contain malicious programs, e.g. viruses	39.6	32.4	38.9	22.2	18.4	
UCE	17	17.6	16.7	15.3	5.3	
Annoying email unrelated to recipients' work	1.9	2.9	2.6	8.3	7.9	
Awareness of anti-spam filters (%YES)	25.5	55.1	37.9	43.8	39.4	**0.001**
Knowledge source for anti-spam filters						
ISPs	0	8.2	2	3.5	4.3	**0.013**
Internet and forums	20	30.6	26.2	20.1	28.7	0.221
Broadcast media, e.g. TV	1.4	4.1	2.2	1.4	3.2	0.706
Government	0	0	1.4	3.5	1.1	0.114
School or university education	0.7	30.6	14.8	24.3	17	**<0.001**
Awareness of government efforts to combat spam (%YES)	24.8	30.6	21.6	30.6	28.7	0.114
Awareness of ISP efforts to combat spam (%YES)	10.3	20.4	11.6	18.8	20.2	**0.017**

*P values are based on chi-square test between public users with different levels of education; P values <0.05 were considered statistically significant.

Abbreviations: HS = High School, D = Diploma, B = Bachelor, and M = Master.

4.1.2.6 The awareness of public users about email spam, anti-spam filters, and the efforts to combat it, based on study discipline

Table 4.8 has shown that there were significant differences between users in different study disciplines in their awareness about email spam, anti-spam filters and efforts to combat it. Users who studied computer science and information technology were more aware of email spam and anti-spam filters (85% and 63.5% respectively) than those who studied in other areas (p<0.001).

Their most common source of knowledge about email spam and anti-spam filters was through school and university education. The percentage of users, who knew about email spam and anti-spam filters through school and university education, was larger in the area of computer science and information technology than all other study disciplines (61.8% and 43% respectively, p<0.001).

Users studied social sciences were more aware of government (38.7%, p=0.003) and ISPs (20.4%, p=0.036) efforts to combat email spam than users studied other areas.

Table 4.8: Percentages of distribution of the awareness of public users about email spam, anti-spam filters, and efforts to combat it, based on study discipline

Question	Study Discipline						P*
	E&T **n=159**	**CS&IT** **n=293**	**SS** **n=93**	**P&BS** **n=93**	**HS&M** **n=88**	**Other** **n=149**	
Awareness of email spam (%YES)	49.1	85	45.2	58.1	54.5	60.4	**<0.001**
Knowledge source for email spam							
ISPs	3.1	5.8	3.2	5.4	1.1	9.4	0.058
Internet and forums	32.1	47.8	33.3	41.9	31.8	36.2	**0.005**
Broadcast media, e.g. TV	6.9	13.3	2.2	4.3	2.3	11.4	**0.001**
Government	3.1	3.1	1.1	2.2	1.1	4	0.688
School or university education	5.7	61.8	1.1	9.7	5.7	18.8	**<0.001**
Definition of email spam							
UBE	20.4	25.5	31.8	29	14.8	22.8	0.291
Sent by unknown senders without recipients' permission	27.8	18.8	9.1	19.4	48.1	26.6	
Sent randomly, contain malicious programs, e.g. viruses	35.2	33.3	31.8	41.9	14.8	34.2	
UCE	14.8	17	18.2	9.7	18.5	12.7	
Annoying email un related to recipients' work	1.9	5.5	9.1	0	3.7	3.8	
Awareness of anti-spam filters (%YES)	25.8	63.5	23.7	28	28.4	33.6	**<0.001**
Knowledge source for anti-spam filters							

Question	**Study Discipline**						**P***
	E&T	**CS&IT**	**SS**	**P&BS**	**HS&M**	**Other**	
	n=159	**n=293**	**n=93**	**n=93**	**n=88**	**n=149**	
ISPs	1.3	3.4	1.1	2.2	1.1	6	0.094
Internet and forums	20.1	34.1	20.4	21.5	20.5	24.2	**0.004**
Broadcast media, e.g. TV	1.3	5.1	0	0	0	2	**0.003**
Government	3.8	2	0	0	1.1	0.7	0.105
School or university education	2.5	43	3.2	6.5	0	9.4	**<0.001**
Awareness of government efforts to combat spam (%YES)	16.4	23.9	38.7	19.4	26.1	26.8	**0.003**
Awareness of ISPs efforts to combat spam (%YES)	8.2	14.7	20.4	9.7	13.6	18.8	**0.036**

*P values are based on chi-square test between public users in different study disciplines; P values <0.05 were considered statistically significant.

Abbreviations: E&T = Education and Teaching, CS&IT = Computer Science and Information Technology, SS = Social Sciences, P&BS = Physical and Biological Sciences, HS&M = Health Sciences and Medicine.

4.1.2.7 The awareness of public users about email spam, anti-spam filters, and the efforts to combat it, based on work status

As seen in Table 4.9, there were significant differences between employees and students in their awareness of email spam, anti-spam filters and efforts to combat it. Employees were more aware of email spam and anti-spam filter than students (email spam: 67.1% vs 55.6%, anti-spam filters: 42.8% vs 31.9%, respectively, p<0.001).

More employees knew about email spam and anti-spam filters from school and university education than did students (30.1% vs 15.6% for awareness of email spam; 20.7% vs 8.1% for awareness about anti-spam filters, respectively, p<0.001).

There were significant differences between employees and students in how they defined email spam. More employees than students defined it as UBE (27.2% vs 17.2%, p=0.02), whereas the reverse was true for those who defined email spam as “email was sent randomly and contain malicious programs such as viruses” (39.6% vs 29.3%, p=0.02). More employees than students considered any email that was not related to recipients’ work to be email spam (5.4% vs 2.1%, p=0.02).

Employees were more aware than students of government and ISPs’ efforts to combat email spam (28.8% vs 18.9%, respectively, p<0.001 and 16.3% vs 10.3%, respectively, p=0.006).

Table 4.9: Percentages of distribution of the awareness of public users about email spam, anti-spam filters, and efforts to combat it based on work status

Question	Work Status		P*
	Student n=455	Employee n=565	
Awareness of email spam (%YES)	55.6	67.1	**<0.001**
Knowledge source for email spam			
ISPs	3.7	6.2	0.076
Internet and forums	37.8	39.6	0.548
Broadcast media, e.g. TV	8.8	8.5	0.867
Government	1.3	3.2	0.051
School or university education	15.6	30.1	**<0.001**
Definition of email spam			
UBE	17.2	27.2	**0.02**
Sent by unknown senders without recipients' permission	24	23.8	
Sent randomly, contain malicious programs, e.g. viruses	39.6	29.3	
UCE	17.2	14.2	
Annoying email unrelated to recipients' work	2.1	5.4	
Awareness of anti-spam filters (%YES)	31.9	42.8	**<0.001**
Knowledge source for anti-spam filters			
ISPs	1.3	3.4	**0.036**
Internet and forums	23.3	26.2	0.287
Broadcast media, e.g. TV	1.3	2.8	0.098
Government	1.1	1.6	0.500
School or university education	8.1	20.7	**<0.001**
Awareness of government efforts to combat spam (%YES)	18.9	28.8	**<0.001**
Awareness of ISPs efforts to combat spam (%YES)	10.3	16.3	**0.006**

*P values are based on chi-square test between student and employee users; P values <0.05 were considered statistically significant.

4.1.2.8 The awareness of public users about email spam, anti-spam filters, and the efforts to combat it based on work position

As shown in Table 4.10, significant differences were found between users in different work positions in their awareness of email spam, anti-spam filters and efforts to combat it. Those who worked in technical positions were more aware of email spam and anti-spam filters than users who worked in other work positions (87.9% and 73.6% respectively, p<0.001).

The most common source of knowledge of users who worked in technical positions about both email spam and anti-spam filters was through school and university education (75.8% and 60.4% respectively, p<0.001). Users who worked in technical positions were also more aware of the government efforts to combat email spam than users in other work positions (41.8%, p=0.003).

Table 4.10: Percentages of distribution of the awareness of public users in different work positions about email spam, anti-spam filters, and efforts to combat it

Question	Work Position					P*
	EP n=274	MP n=58	TP n=91	MTP n=97	Other n=45	
Awareness of email spam (%YES)	63.1	55.2	87.9	66	70.6	**<0.001**
Knowledge source about email spam						
ISPs	6.2	3.4	7.7	4.1	11.8	0.453
Internet and forums	41.6	34.5	40.7	39.2	38.2	0.892
Broadcast media, e.g. TV	7.3	0	12.1	9.3	20.6	**0.008**
Government	2.6	1.7	4.4	1	8.8	0.159
School or university education	21.9	6.9	75.8	22.7	32.4	**<0.001**
Definition of email spam						
UBE	25.7	23.1	23.2	34.9	42.1	0.174
Sent by unknown senders without recipients' permission	22.8	53.8	16.1	30.2	10.5	
Sent randomly, contains malicious programs, e.g. viruses	30.7	7.7	33.9	25.6	31.6	
UCE	12.9	15.4	19.6	7	15.8	
Annoying email unrelated to recipients' work	7.9	0	7.1	2.3	0	
Awareness of anti-spam filters (%YES)	37.6	25.9	73.6	43.3	26.5	**<0.001**
Knowledge source about anti-spam filters						
ISPs	2.9	0	7.7	1	5.9	**0.04**
Internet and forums	28.5	22.4	26.4	26.8	11.8	0.304
Broadcast media, e.g. TV	1.8	1.7	7.7	1	2.9	**0.031**
Government	1.1	1.7	3.3	1	2.9	0.616
School or university education	13.1	5.2	60.4	13.4	14.7	**<0.001**
Awareness of government efforts to combat spam (%YES)	27.4	37.9	41.8	21.9	14.7	**0.003**
Awareness of ISP efforts to combat spam (%YES)	15.3	17.2	24.2	14.4	11.8	0.286

*P values are based on chi-square test between public users in different work positions; P values < 0.05 were considered statistically significant.

Abbreviations: EP = Educational Positions, MP = Medical Positions, TP = Technical Positions, MTP = Management Positions.

4.1.3 The Nature of Email Spam as Perceived by Public Users

The results for public users' experiences with their email account provider, and the average number of email spam received are shown in Table 4.11.

Most participants (68.1%, 95%CI: 65.2%-70.9%) used Hotmail. Those who used "other" types of emails, such as email provided by work or university, were the smallest group (0.7%, 95%CI: 0.3%-1.3%). About three-quarters of the participants (71.4%, 95%CI: 68.5%-74.1%) had used email for less than 8 years.

Approximately three-quarters of the participants (73.1%, 95%CI: 70.4%-75.8%) said that they received spam, and nearly half of the participants (46.4%, 95%CI: 42.6%-49.7%) reported receiving more than 25 emails spam weekly. Most of the email spam (59%, 95%CI: 57%-61%) received by public users in Saudi Arabia were written in English, followed by Arabic (34.4%, 95%CI: 32.4%-36.3%).

Table 4.11: Percentage distribution of email account providers used, number and languages of email spam received by public users

Question	Frequency	Percentage (%)	95% CI
Email account provider			
Hotmail	695	68.1	65.2-70.9
Yahoo	175	17.2	14.9-19.6
Gmail	143	14	12-16.3
Other	7	0.7	0.3-1.3
Experience using email			
< 8 years	728	71.4	68.5-74.1
≥ 8 years	292	28.6	25.9-31.5
Receive email spam			
Yes	746	73.1	70.4-75.8
No	274	26.9	24.2-29.6
Average number of spam / week			
< 5	54	7.2	5.5-9.3
5-15	270	36.4	32.8-39.7
16-25	74	10	7.9-12.2
> 25	344	46.4	42.6-49.7
Languages of email spam			
English	707	59	57-61
Arabic	681	34.4	32.4-36.3
Unrecognised languages	188	4	3.3-4.7
Other languages	108	2.5	1.9-3

There were significant differences between Arabic and English email spam received (see Table 4.12). Much more Arabic language spam was connected to forums (35.2% vs 3.3%, p<0.001). The percentage of religious and political emails was also higher in Arabic email spam (4.8% vs 2.6%, p<0.001).

On the other hand, there were significantly more pornographic emails in English spam than in Arabic (24% vs 9.8%, p<0.001). There were also more phishing and fraud emails in English than in Arabic (27.4% vs 6.2%, p<0.001).

Table 4.12: Percentages of types of Arabic and English email spam received by public users

Types of email spam	Arabic (%)	English (%)	P*
Business	31	29.4	0.07
Religious and Political	4.8	2.6	**<0.001**
Pornographic	9.8	24.7	**<0.001**
Forums	35.2	3.3	**<0.001**
Products and Services	11.6	11.9	0.785
Phishing and Fraud	6.2	27.4	**<0.001**
Other	1.3	0.8	0.208

*P values are based on paired-samples t-test between types of Arabic and English email spam; P values <0.05 were considered statistically significant.

4.1.3.1 The nature of email spam as perceived by public users in different regions

As shown in Table 4.13, there were significant differences between public users' experience of email service in the different regions of Saudi Arabia. Hotmail was used more in the western region (78.6%) than elsewhere, and Yahoo was used more in the northern region (24.6%) than elsewhere (p=0.01).

Public users in the eastern region had more experience in using email than those in other regions. The percentage of public users with eight years or more experience with email, was highest in the eastern region (36.4%, p=0.001).

Those in the central region received the highest average number of email spam. The central region had significantly more users who received more than 25 spam weekly (51.4%) than in other regions (p=0.033).

Table 4.13: Percentages of email account providers used and the number of email spam received by public users in different regions

Question	Region					P*
	E n=203	W n=201	C n=352	S n=134	N n=130	
Email account provider						
Hotmail	69.5	78.6	65.6	66.4	58.5	**0.01**
Yahoo	17.7	10.9	15.9	21.6	24.6	
Gmail	11.8	10	17.3	11.9	16.9	
Other	1	0.5	1.1	0	0	

Question	Region					P*
	E n=203	W n=201	C n=352	S n=134	N n=130	
Experience using email						
< 8 years	63.6	72.9	79.6	71.6	77.7	**0.001**
≥ 8 years	36.4	27.1	20.4	28.4	22.3	
Average number of spam/ week						
<5	9.2	13.3	3.5	9.1	2.4	**0.033**
5-15	36.9	32.7	36.6	35.5	42.9	
16-25	8.5	11.3	8.6	12.7	10.7	
>25	45.5	42.7	51.4	42.7	44	

*P values are based on chi-square test between public users in different regions; P values <0.05 were considered statistically significant.

Abbreviations: E = Eastern, W = Western, C = Central, S = Southern, N = Northern.

There were significant differences between users in different regions in the languages of email spam received, as revealed in Table 4.14. The percentage of English email spam was higher in the central region than in other regions (60.7%, p=0.002), and the percentage of Arabic email spam was larger in the western region than in other regions (43.3%, p<0.001). The southern region received more spam in languages such as Chinese and Turkish (4.2%, p=0.042).

There were significant differences between the regions in the types of Arabic email spam received. There were more Arabic language phishing and fraud emails in the western region than in other regions (11.3%, p<0.001), but more emails related to forums in the northern region than in other regions (42%, p=0.003).

English spam also varied in the types of between users in different regions. The western region received more emails in English than did other regions that were related to forums (5.7%, p<0.001) and products and services (16.7%, p=0.005). Phishing and fraud emails were received more in the southern and eastern regions than in other regions (31.8% and 31.1% respectively, p=0.005).

Table 4.14: Percentages of languages and types of Arabic and English email spam received by public users in different regions

Question	Region					P*
	E n=203	W n=201	C n=352	S n=134	N n=130	
Languages of email spam						
English	59.7	51	60.7	26.6	26.3	**0.002**
Arabic	33.5	43.3	33.3	29.5	29.2	**<0.001**
Unrecognised languages	3.7	3	4	5	5.2	0.365

Question	Region					P*
	E n=203	W n=201	C n=352	S n=134	N n=130	
Other languages	3	2.5	2	4.2	1.1	**0.042**
Types of Arabic email spam						
Business	30.8	28.3	32.3	34.2	30.3	0.306
Religious and political	5.8	5	4.8	4.5	4.8	0.874
Pornographic	9	11.3	10.4	5.7	8.8	0.122
Forums	35.4	29.4	35.8	38.5	42	**0.003**
Products and services	13.3	12.9	10.3	11.4	8.7	0.099
Phishing and fraud	4.2	11.3	4.9	4.8	4.8	**<0.001**
Other	1.4	1.7	1.4	0.7	0.5	0.790
Types of English email spam						
Business	26.7	28.1	30.8	29.6	32.1	0.325
Religious and political	2.5	3.1	1.9	2.7	1.8	0.391
Pornographic	26.4	22.4	23.6	23.4	25.5	0.526
Forums	2.6	5.7	2.4	3.1	2	**<0.001**
Products and services	9.6	16.7	12.6	8.9	9.9	**0.005**
Phishing and fraud	31.1	22	28.6	31.8	27.4	**0.005**
Other	1	1.8	0	0.3	1.8	0.157

*P values are based on ANOVA test between public users in different regions; P values <0.05 were considered statistically significant.

Abbreviations: E = Eastern, W = Western, C = Central, S = Southern, and N = Northern.

4.1.3.2 The nature of email spam as perceived by males and females

As shown in Table 4.15, males had more experience in using email than females. The percentage of males with an experience of eight years and more in using email was higher than the percentage of females (33.7% vs 22.3%, p<0.001).

There was a significant difference between males and females in the average number of email spam received weekly. The percentage of males receiving more than 25 spam weekly was higher than the percentage of females (58% vs 29.5, p<0.001).

Table 4.15: Percentages of email account providers used, and the number of email spam received by male and female users

Question	Gender		P*
	Male n=608	Female n=412	
Email account provider			
Hotmail	69.7	65.8	0.077
Yahoo	15	20.4	
Gmail	14.3	13.6	
Other	1	0.2	
Experience using email			
<8 years	67.3	77.7	**<0.001**

Question	**Gender**		**P***
	Male n=608	**Female n=412**	
≥8 years	33.7	22.3	
Average number of spam / week			
<5	10	3.3	**<0.001**
5-15	22.5	56.6	
16-25	9.5	10.6	
>25	58	29.5	

*P values are based on chi-square test between male and female users; P values <0.05 were considered statistically significant.

Table 4.16 shows that females received more English email spam than did males (68% vs 53%, p<0.001), and males received more Arabic email spam than that females (39.7% vs 26.5%, p<0.001).

There were significant differences types of Arabic and English spam received by males and females. Males received more pornographic emails in Arabic than females (13.4% vs 3.9%, p<0.001), and females received more emails related to forums than males (43.9% vs 29.5%, p<0.001).

Males received more religious and political emails in Arabic than females (6% vs 3.6%, p<0.001), but males received more products and services emails in English than females (14.7% vs 8.4%, p<0.001).

Table 4.16: Percentages of languages and types of Arabic and English email spam received by male and female users

Questions	**Gender**		**P***
	Male n=608	**Female n=412**	
Languages of email spam			
English	53	68	**<0.001**
Arabic	39.7	26.5	**<0.001**
Unrecognised languages	4.2	3.7	0.426
Other languages	3.1	1.7	**0.01**
Types of Arabic email spam			
Business	30.3	32.7	0.142
Religious and political	6	3.6	**<0.001**
Pornographic	13.4	3.9	**<0.001**
Forums	29.5	43.9	**<0.001**
Products and services	11.6	11.2	0.761
Phishing and fraud	7.7	3.6	**<0.001**
Other	1.5	1	0.457

Questions	Gender		P*
	Male n=608	Female n=412	
Types of English email spam			
Business	30.3	28.3	0.194
Religious and political	2.7	2	0.145
Pornographic	22.6	26.1	**0.015**
Forums	4.3	1.5	**<0.001**
Products and services	14.7	8.4	**<0.001**
Phishing and fraud	24.5	33	**<0.001**
Other	0.85	0.60	0.655

*P values are based on independent-samples t-test between male and female users; P values <0.05 were considered statistically significant.

4.1.3.3 The nature of email spam as perceived by public users in different age groups

There were also differences for age groups. Table 4.17 shows that Hotmail was used mostly by younger users (80.6%), and Gmail was used mostly by older users (26.7%, p<0.001).

Older users had more experience in using email than other age groups, shown by a higher percentage of users aged 46 and over (55.6%) who had used email for eight years or more, than other age groups (p<0.001). The percentage of users aged 26-35, who received more than 25 spam weekly, was higher than the percentages of other age groups (49.6%, p=0.002).

Table 4.17: Percentages of email account providers used, and the number of email spam received by users in different age groups

Question	Age Groups				P*
	15-25 n=463	26-35 n=358	36-45 n=154	46+ n=45	
Email account provider					
Hotmail	80.6	61.7	50.6	51.1	**<0.001**
Yahoo	9.3	20.7	31.8	20	
Gmail	9.7	17.6	14.9	26.7	
Other	0.4	0	2.6	2.2	
Experience using email					
<8 years	86.8	62.8	53.2	44.4	**<0.001**
≥8 years	13.2	37.2	46.8	55.6	
Average number of spam / week					
<5	11.8	4.4	2.5	3.3	**0.002**
5-15	34.2	33.8	44.9	50	
16-25	8.4	12.1	8.5	13.3	
>25	45.7	49.6	44.1	33.3	

*P values are based on chi-square test between public users in different age groups; P values <0.05 were considered statistically significant.

Table 4.18 showed that users aged 36-45 received more English email spam than did other age groups (69.9%, p<0.001), while users aged 15-25 received more Arabic email spam the other age groups (41.4%, p<0.001).

The types of Arabic and English email spam received by public users in different age groups differed significantly. Users aged 15-25 received more pornographic emails in Arabic than other age groups (12.6%, p<0.001) and users aged 26-35 received more emails related to forums in Arabic than other age groups (39.3%, p=0.002).

Users aged 15-25 also received more religious and political emails in English than other age groups (3.1%, p=0.035), whereas users aged 46 and older received more English language phishing and fraud emails than by other age groups (32.4%, p=0.01).

Table 4.18: Percentages of languages and types of Arabic and English email spam received by public users in different age groups

	Age Groups				**P***
Question	**15-25** **n=463**	**26-35** **n=358**	**36-45** **n=154**	**46+** **n=45**	
Languages of email spam					
English	51	63	69.9	66.5	**<0.001**
Arabic	41.4	31	24.8	25.8	**<0.001**
Unrecognised languages	4	3.9	4	5.8	0.752
Other languages	3.5	2	1.3	1.8	**0.024**
Types of Arabic email spam					
Business	29.3	32.1	33.6	35.1	0.180
Religious and political	5	4.4	6.4	5.7	0.342
Pornographic	12.6	8	5.6	4.7	**<0.001**
Forums	31.6	39.3	37.5	37.3	**0.002**
Products and services	12.5	9.7	12.1	11.5	0.146
Phishing and fraud	7.9	5	3.2	5	**0.002**
Other	1.1	1.4	1.4	0.7	0.947
Types of English email spam					
Business	29.3	30	28.4	30	0.918
Religious and political	3.1	1.6	2.3	1.5	**0.035**
Pornographic	26.5	21.9	23.3	21.9	0.054
Forums	3.6	3	2.4	3.1	0.546
Products and services	11.7	12.9	11.6	10.3	0.828
Phishing and fraud	24.8	30.2	31.2	32.4	**0.01**
Other	1	0.3	1	0.7	0.713

*P values are based on ANOVA test between public users in different age groups; P values < 0.05 were considered statistically significant.

4.1.3.4 The nature of email spam as perceived by Saudis and non-Saudis public users

Table 4.19 revealed significant differences between Saudis and non-Saudis in the email account provider used. Saudis used Hotmail more than non-Saudis (71.6% vs 50.3%, p<0.001), while non-Saudis used Yahoo more than Saudis (36.4% vs 13.5%, p<0.001).

Non-Saudis had more experience in using email than Saudis. There were more non-Saudis who had used email for eight years than Saudis (45.5% vs 25.3%, p<0.001).

Table 4.19: Percentages of email account providers used and the number of email spam received by Saudi and non-Saudi users

Question	Nationality		P*
	Saudi n=855	Non-Saudi n=165	
Email account provider			
Hotmail	71.6	50.3	**<0.001**
Yahoo	13.5	36.4	
Gmail	14.4	12.1	
Other	0.6	1.2	
Experience using email			
<8 years	74.7	54.5	**<0.001**
≥8 years	25.3	45.5	
Average number of spam / week			
<5	7.7	5.5	0.464
5-15	35.2	42.2	
16-25	10.1	9.4	
>25	47.1	43	

*P values are based on chi-square test between Saudi and non-Saudi users; P values <0.05 were considered statistically significant.

Non-Saudi users, as shown in Table 4.20, received more English email spam than Saudi users (72.2% vs 56.3%, p<0.001), and Saudi users received more Arabic email spam than non-Saudi users (37% vs 21.6%, p<0.001).

There were significant differences between Saudis and non-Saudis in terms of Arabic and English email spam. Saudi users received more pornographic emails in Arabic than did non-Saudi users (10% vs 6.5%, p=0.022), but non-Saudi users received more religious and political emails than Saudi users (6.8% vs 4.7%, p=0.031).

Saudi users received more products and service emails in English than non-Saudis (13% vs 7.6%, p<0.001), but non-Saudi users received more phishing and fraud emails than Saudi users (34.7% vs 26.7%, p<0.001).

Table 4.20: Percentages of languages and types of Arabic and English email spam received by Saudi and non-Saudi users

Question	Nationality		P*
	Saudi n=855	Non-Saudi n=165	
Languages of email spam			
English	56.3	72.2	**<0.001**
Arabic	37	21.6	**<0.001**
Unrecognised languages	4.1	3.7	0.640
Other languages	2.6	2	0.390
Types of Arabic email spam			
Business	30.7	34.5	0.112
Religious and political	4.7	6.8	**0.031**
Pornographic	10	6.5	**0.022**
Forums	35.4	36	0.841
Products and services	11.6	10.6	0.549
Phishing and fraud	6.4	4.1	**0.022**
Other	1.2	1.4	0.830
Types of English email spam			
Business	29.8	27.8	0.359
Religious and political	2.4	2.2	0.783
Pornographic	24.1	24	0.952
Forums	3.2	2.7	0.481
Products and services	13	7.6	**<0.001**
Phishing and fraud	26.7	34.7	**<0.001**
Other	0.7	0.8	0.939

*P values are based on independent-samples t-test between Saudi and non-Saudi users; P values < 0.05 were considered statistically significant.

4.1.3.5 The nature of email spam as perceived by public users in different education levels

Table 4.21 revealed that users completed high school (75.9%) used Hotmail more than other users completed other degrees while users completed PhD (34%) used Yahoo more than users completed other degrees (p<0.001).

Users completed PhD degree had more an experience in using email than users completed other degrees. The percentage of users, who had 8 years and more experience in using email, was higher in PhD degree than other degrees (57.4%, p<0.001).

Table 4.21: Percentages of email account providers used, and the number of email spam received by users in different education levels

Question	Education Level					P*
	HS n=145	D n=49	B n=588	M n=144	PhD n=94	
Email account provider						
Hotmail	75.9	67.3	73	56.3	44.7	**<0.001**
Yahoo	13.1	16.2	12.9	29.9	34	
Gmail	10.3	22.4	13.6	13.9	18.1	
Other	0.7	0	0.5	0	3.2	
Experience using email						
<8 years	90.3	63.3	77.4	50	42.6	**<0.001**
≥8 years	9.7	36.7	22.6	50	57.4	
Average number of spam / week						
<5	8.9	2.9	8.9	4.3	2.7	0.113
5-15	35.6	20	38	31.9	43.2	
16-25	12.9	8.6	9.6	10.3	8.1	
>25	42.6	68.6	43.5	53.4	45.9	

*P values are based on chi-square test between public users in different education level; P values <0.05 were considered statistically significant.
Abbreviations: HS = High School, D = Diploma, B = Bachelor, and M = Master.

Table 4.22 summarised the results about the languages and types of Arabic and English email spam received by users in different education levels. The percentage of English email spam received by users completed PhD degree was larger than that received by users completed other education degrees (72.9%, p<0.001). The percentages of Arabic email spam received by users completed high school education or less, were larger than that received by users completed other degrees (41.4%, p<0.001).

There were significant differences between users in different education levels in terms of the percentages of types of Arabic and English email spam. The percentage of pornographic emails in English received by users completed high school, was higher than that received by users in other degrees (27.2%, p=0.010), while the percentage of pornographic emails in Arabic received by users completed diploma, was greater than that received by users in other degrees (13.7%, p=0.003).

The percentage of phishing and fraud emails in English received by users completed PhD, was larger than that received by users completed other degrees (37.9%, p=0.002). Users who completed high school, received more other types of email spam such as fun, puzzles, competitions, greetings, and invitation for friendship by social network websites such as Facebook, than users in other degrees (2.7%, p=0.046).

Table 4.22: Percentages of languages and types of Arabic and English email spam received by public users in different education levels

Question	Education Level					P*
	HS n=145	D n=49	B n=588	M n=144	PhD n=94	
Languages of email spam						
English	50.7	59.1	56.8	65.5	72.9	**<0.001**
Arabic	41.4	30.3	36.7	28.4	22.4	**<0.001**
Unrecognised languages	3.1	9.1	3.6	4.8	3.8	**0.01**
Other languages	4.7	1.3	2.8	1.1	0.9	**0.003**
Types of Arabic email spam						
Business	28.1	29.2	30.7	34.3	35.3	0.174
Religious and Political	5.2	2.1	4.7	5.2	7.5	0.138
Pornographic	10.2	13.7	10.9	5.7	4	**0.003**
Forums	36.9	35.3	33.5	40.5	37.8	0.095
Products and Services	13.1	12.7	11.7	9.7	9.8	0.423
Phishing and Fraud	5.1	5.8	7	3.7	5	0.127
Other	1.4	1	1.4	1	0.5	0.922
Types of English email spam						
Business	26.5	34	29.9	31.6	25.1	0.136
Religious and Political	2.2	1	2.8	2.2	1	0.173
Pornographic	27.2	15.4	25.3	23	19.5	**0.01**
Forums	3.4	3.5	3.6	2.1	1.7	0.180
Products and Services	11.6	14.6	11.9	11.3	13.3	0.896
Phishing and Fraud	26.3	31.7	26.1	29	37.9	**0.002**
Other	2.7	0.6	0.2	0.7	1.4	**0.046**

*P values are based on ANOVA test between public users in different education levels; P values <0.05 were considered statistically significant.

Abbreviations: HS = High School, D = Diploma, B = Bachelor, and M = Master.

4.1.3.6 The nature of email spam as perceived by public users in different study disciplines

Table 4.23 showed that Hotmail was used significantly more by users who studied physical and biological sciences (82.8%) than users in other study disciplines, whereas who users studied social sciences used Yahoo (24.7%) more than users in other areas (p=0.002).

Those who had used email for eight years and more tended to be in the area of computer science and information technology (38.9%) than in other areas (p=0.006).

Table 4.23: Percentages of email account providers used, and the number of email spam received by users in different study disciplines

Question	Study Discipline						P*
	E&T n=159	CS&IT n=293	SS n=93	P&BS n=93	HS&M n=88	Other n=149	
Email account provider							
Hotmail	66.7	64.8	61.3	82.8	59.1	69.1	**0.002**
Yahoo	19.5	14	24.7	12.9	21.6	20.1	
Gmail	13.8	20.1	14	3.2	19.3	9.4	
Other	0	1	0	1.1	0	1.3	
Experience using email							
<8 years	71.7	61.1	64.5	77.4	78.4	69.8	**0.006**
≥8 years	28.3	38.9	35.5	22.6	21.6	30.2	
Average number of spam/ week							
<5	7.6	7.7	4.4	9.7	4.9	6.1	**0.002**
5-15	48.3	26.5	45.6	51.6	41	27.6	
16-25	6.8	12.8	7.4	4.8	9.8	9.2	
>25	37.3	53	42.6	33.9	44.3	57.1	

*P values are based on chi-square test between public users in different study disciplines; P values < 0.05 were considered statistically significant.

Abbreviations: E&T = Education and Teaching, CS&IT = Computer Science and Information Technology, SS = Social Sciences, P&BS = Physical and Biological Sciences, HS&M = Health Sciences and Medicine.

The percentage, summarised in Table 4.24, revealed that there was a significant difference in different study disciplines of users who received spam written in languages rather than Arabic and English. The percentage of other languages of email spam received, such as Chinese and Turkish, was higher for those who studied physical and biology sciences than other areas (3.6%, p=0.006).

When looking at the types of Arabic and English email spam received by users in different study disciplines, there were again significant differences. Users studying social sciences and health sciences received more email spam related to forums in Arabic than users in other areas (41.5% and 41.4%, p=0.001). Those who studied computer science and information technology received more products and services emails in English than users who studied other areas (15.5%, p=0.01).

Table 4.24: Percentages of languages and types of Arabic and English email spam received by public users in different study disciplines

Questions	Study Discipline						P*
	E&T n=159	CS&IT n=293	SS n=93	P&BS n=193	HS&M n=88	Other n=149	
Languages of email spam							
English	57.4	61.8	63.4	58.3	65.5	56.7	0.213
Arabic	35.2	32.7	30.9	33.6	29.1	36.1	0.552

	Study Discipline						**P***
Questions	**E&T n=159**	**CS&IT n=293**	**SS n=93**	**P&BS n=193**	**HS&M n=88**	**Other n=149**	
Languages of email spam							
Unrecognised languages	4.1	4.3	5.4	4.4	2.6	4	0.720
Other languages	3.2	1.1	0.4	3.6	2.8	3.2	**0.006**
Types of Arabic email spam							
Business	31.8	31.7	32.7	29.5	33.4	31.5	0.964
Religious and political	5.1	3.7	6.5	6.5	4.8	5.8	0.144
Pornographic	9.1	10	5.8	10.7	6.5	11.7	0.256
Forums	38.8	34.2	41.5	30.5	41.4	28	**0.001**
Products and services	10.8	10.6	9.8	15.5	11	11.2	0.240
Phishing and fraud	3.9	7.5	2.6	7	2.9	10	**<0.001**
Other	0.4	2.2	1	0.1	0	1.8	0.241
Types of English email spam							
Business	29.7	29.5	28.1	27.6	33.5	31.1	0.677
Religious and political	3.1	1.7	1.9	4	2.2	2.5	0.135
Pornographic	27.7	21.3	24.2	24.3	24.5	23	0.150
Forums	3.1	2.6	2	2.9	3.8	4.8	0.231
Products and services	8.5	15.5	11.2	11.2	6.9	12.5	**0.01**
Phishing and fraud	27.7	28.9	32.6	29.7	28.1	24.3	0.371
Other	0	0.5	0.1	0	0.1	1.6	0.261

*P values are based on ANOVA test between public users in different study disciplines; P values <0.05 were considered statistically significant.

Abbreviations: E&T = Education and Teaching, CS&IT = Computer Science and Information Technology, SS = Social Sciences, P&BS = Physical and Biological Sciences, and HS&M = Health Sciences and Medicine.

4.1.3.7 The nature of email spam as perceived by students and employees

Table 4.25 found that most of students used Hotmail compared to employees (80% vs 58.6%), while employees used Yahoo (22.8% vs 10.1%) and Gmail (17.7% vs 9.5%) more than students (p<0.001).

The experience of employees in using email was longer than the experience of students, which the results showed that the percentage of employees with an experience of 8 years and more in using email was greater than the percentage of students (39.8% vs 14.5%, p<0.001).

The average number of email spam received by employees was larger than that received by students, which the results showed that the percentage of employees received more than 25 spam weekly was higher than the percentage of students (48% vs 44.1%, p=0.034).

Table 4.25: Percentages of email account providers used, and the number of email spam received by student and employee users

Question	Work Status		P*
	Student n=455	Employee n=565	
Email account provider			
Hotmail	80	58.6	**<0.001**
Yahoo	10.1	22.8	
Gmail	9.5	17.7	
Other	0.4	0.9	
Experience using email			
<8 years	85.5	60.2	**<0.001**
≥8 years	14.5	39.8	
Average number of spam/ week			
< 5	10.5	4.9	**0.034**
5-15	35.8	36.8	
16-25	9.6	10.3	
>25	44.1	48	

*P values are based on chi-square test between student and employee users; P values <0.05 were considered statistically significant.

Employees received more English email spam than students (64% vs 52.3%, p<0.001), whereas students received Arabic email spam more than employees (40.2% vs 30%, p<0.001) (See Table 4.26).

Significant differences were found in the types of Arabic and English email spam received by students and employees. Employees received more business emails in Arabic than did students (33% vs 28.9%, p=0.015), and students received more pornographic emails in both Arabic and English than did employees (Arabic: 12% vs 7.6%, p=0.001, and English 27.1% vs 21.9%, p=0.001).

Table 4.26: Percentages of languages and types of Arabic and English email spam received by student and employee users

Question	Work Status		P*
	Student n=455	Employee n=565	
Languages of email spam			
English	52.3	64	**<0.001**
Arabic	40.2	30	**<0.001**
Unrecognised languages	4	4	0.894
Other languages	3.4	1.9	**0.016**

Question	Work Status		P*
	Student n=455	Employee n=565	
Types of Arabic email spam			
Business	28.9	33	**0.015**
Religious and political	4.9	5.1	0.759
Pornographic	12	7.6	**0.001**
Forums	32.6	37.7	**0.007**
Products and services	12.6	10.5	0.066
Phishing and fraud	7.9	4.6	**0.001**
Other	1.1	1.4	0.692
Types of English email spam			
Business	28.1	30.4	0.181
Religious and political	2.6	2.1	0.352
Pornographic	27.1	21.9	**0.001**
Forums	3.7	2.7	0.107
Products and services	10.7	13	0.116
Phishing and fraud	26.4	29.3	0.104
Other	1	0.5	0.378

*P values are based on Independent-Samples t-test between student and employee users; P values <0.05 were considered statistically significant.

4.1.3.8 The nature of email spam as perceived by public users in different work positions

Table 4.27 shows that there was a significant difference between users in different work positions in the average number of email spam received weekly. The percentage of users, who worked in other positions (e.g. academia and banking) that were not specified in this study, receiving more than 25 spam weekly was greater than that received by those who worked in other positions that highlighted in this study (e.g. educational and medical) (70.6%, p<0.001).

Table 4.27: Percentages of email account providers used and the number of email spam received by users in different work positions

Question	Work Position					P*
	EP n=274	MP n=58	TP n=91	MTP n=97	Other n=45	
Email account provider						
Hotmail	60.6	56.9	52.7	64.9	50	0.210
Yahoo	24.5	31	19.8	15.5	29.4	
Gmail	13.9	12.1	26.4	18.6	20.6	
Other	1.1	0	1.1	1	0	
Experience using email						
<8 years	62	69	51.6	58.8	55.9	0.252
≥8 years	38	31	48.4	41.2	44.1	

Question	Work Position					P*
	EP n=274	MP n=58	TP n=91	MTP n=97	Other n=45	
Average number of spam / week						
<5	5.6	0	3.9	7.6	2.9	**<0.001**
5-15	47.7	59.9	18.4	24.2	8.8	
16-25	7.9	10.8	11.8	12.1	17.6	
>25	38.9	29.7	65.8	56.1	70.6	

*P values are based on chi-square test between public users in different work positions; P values <0.05 were considered statistically significant.
Abbreviations: EP = Educational Positions, MP = Medical Positions, TP = Technical Positions, MTP = Management Positions.

Table 4.28 revealed no significant differences between users in different work positions in the languages of email spam and types of Arabic email spam that they received. On the other hand, there were significant differences between users in different work positions in the types of English email spam they received. Users who worked in medical and educational positions received more pornographic emails in English than users in other work positions (25.8% and 25.6% respectively, p<0.001), whereas those who worked in technical positions received more product and service emails in English than users in other work positions (16.5%, p=0.028).

Table 4.28: Percentages of languages and types of Arabic and English email spam received by users in different work positions

Question	Work Position					P*
	EP n=274	MP n=58	TP n=91	MTP n=97	Other n=45	
Languages of email spam						
English	62.6	70.4	66.4	59	70.1	0.123
Arabic	31.5	25	27.5	32.8	26.7	0.383
Unrecognised languages	3.8	3.8	4	5.6	2.2	0.522
Other languages	2	0.8	2	2.5	1	0.696
Types of Arabic email spam						
Business	30.5	35.8	37.9	35	33.5	0.170
Religious and political	5.6	5.3	2.9	5.1	6.1	0.403
Pornographic	8.7	4.7	7.7	6.6	4.8	0.525
Forums	38.9	41.7	33.4	35.6	37.1	0.468
Products and services	10.6	9.7	10.5	11.3	9.4	0.971
Phishing and fraud	4.5	2.5	3.9	5	8.4	0.264
Other	1	0.3	3.6	1.1	0.5	0.241
Types of English email						

Question	Work Position					P*
	EP n=274	MP n=58	TP n=91	MTP n=97	Other n=45	
spam						
Business	28	34.6	34.7	30.9	30.7	0.133
Religious and political	2.1	2.4	1.7	3.5	0.6	0.248
Pornographic	25.6	25.8	16.4	18.1	12	**<0.001**
Forums	2	3.5	3.2	3.8	4	0.301
Products and services	11.4	7.2	16.5	13.5	21	**0.028**
Phishing and fraud	30.5	26	27	30	29.3	0.733
Other	0.5	0	0.3	0.5	2.1	0.610

*P values are based on ANOVA test between public users in different work positions; P values <0.05 were considered statistically significant.

Abbreviations: EP = Educational Positions, MP = Medical Positions, TP = Technical Positions, and MTP = Management Positions.

4.1.4 How Public Users Deal with Email Spam

The results for how public email users dealt with email spam are summarised in Table 4.29. The highest percentage of participants (39.9%, 95%CI: 36.9%-42.9%) said that they sometimes read the entire email spam. Nearly a third (28%, 95%CI: 25%-30.5%) of the participants always deleted email spam without reading it. Only a few public users in Saudi Arabia (3.1%, 95%CI: 2.2%-4.3%) always contacted the ISPs and notified them about email spam.

About one-fifth of the participants (20.9%, 95%CI: 18.5%-23.5%) responded to offers made in email spam, and for them the most positive impact of email spam was fun (12.5%, 95%CI: 10.6%-14.7%).

Table 4.29: Percentages of distribution of dealing of public users with email spam

Question	Frequency	Percentage (%)	95% CI
Read entire email spam			
Never	249	24.4	21.9-27.1
Sometimes	407	39.9	36.9-42.9
Always	63	6.2	4.8-7.8
Delete email without reading			
Never	59	5.8	4.5-7.3
Sometimes	392	38.4	35.5-41.4
Always	282	27.6	25-30.5
Notify ISP about email spam			
Never	579	56.8	53.7-59.8
Sometimes	100	9.8	8.1-11.7
Always	32	3.1	2.2-4.3
Respond to email spam			
Yes	213	20.9	18.5-23.5

Question	Frequency	Percentage (%)	95% CI
No	807	79.1	76.5-81.5
Positive impact of email spam			
Purchasing and selling	38	3.7	2.7-5
Learning	92	9	7.4-10.9
Fun	128	12.5	10.6-14.7

4.1.4.1 Dealing of public users with email spam, based on geographic region

Significant differences were found in the way public users in different regions dealt with email spam, as seen in Table 4.30. About two-thirds of public users in the western region (62.2%) did not notify ISPs, and this percentages was significantly higher than the percentages of public users in other regions (p<0.001).

Public users in the southern region (33.6%) responded more to email spam than public users in other regions (p=0.002). Users in the southern regions responded more to learning and fun materials (15.7% and 23.9% respectively) offered by email spam than public users in other regions (p=0.023 and p=0.001 respectively).

Table 4.30: Percentages of dealing of public users with email spam, based on region

	Region					P*
Question	**E** **n=203**	**W** **n=201**	**C** **n=352**	**S** **n=134**	**N** **n=130**	
Read entire email spam						
Never	27.1	23.9	26.1	22.4	18.5	**0.045**
Sometimes	33	44.3	37.2	49.3	41.5	
Always	7.9	3.5	7.1	7.5	3.8	
Delete email SPAM without reading						
Never	7.4	4.5	4.8	10.4	3.1	**0.035**
Sometimes	34	43.3	36.4	41.8	40	
Always	27.6	25.4	31	28.4	21.5	
Notify ISP about email spam						
Never	51.7	62.2	58	56	53.8	**<0.001**
Sometimes	12.8	9	9.9	11.9	3.8	
Always	3	0.5	2	9	4.6	
Responding to email spam (%YES)	19.2	15.4	20.5	33.6	20	**0.002**
Positive impact of email spam						
Purchasing and selling	4.4	1.5	3.7	5.2	4.6	0.379
Learning	6.4	6	9.7	15.7	9.2	**0.023**
Fun	10.8	10.9	11.1	23.9	10	**0.001**

*P values are based on chi-square test between public users in different regions; P values <0.05 were considered statistically significant.
Abbreviations: E = Eastern, W = Western, C = Central, S = Southern, and N = Northern.

4.1.4.2 Dealing of public users with email spam, based on gender

As seen in Table 4.31, there were no significant differences between males and females in their dealing with email spam.

Table 4.31: Percentages of dealing of public users with email spam, based on gender

Question	Gender		P*
	Male n=608	Female n=412	
Read entire email spam			
Never	24.5	24.3	0.434
Sometimes	38.5	42	
Always	5.8	6.8	
Delete email SPAM without reading			
Never	5.6	6.1	0.086
Sometimes	35.5	42.7	
Always	29.9	42.3	
Notify ISP about email spam			
Never	54.1	60.7	0.135
Sometimes	9.7	10	
Always	3.6	2.4	
Respond to email spam (%YES)	20.6	21.4	0.758
Positive impact of email SPAM			
Purchasing and selling	4.4	2.7	0.143
Learning	8.6	9.7	0.527
Fun	11.8	13.6	0.406

*P values are based on chi-square test between male and female users; P values <0.05 were considered statistically significant.

4.1.4.3 Dealing of public users with email spam, based on age group

A significant difference, as shown in Table 4.32, was found between the ways users in different age groups dealt with email spam. Compared to other age groups, most older users (40%) always deleted it without reading it (p=0.005).

Table 4.32: Percentages of dealing of public users with email spam, based on age group

Question	Age Groups				P*
	15-25 **n=463**	**26-35** **n=358**	**36-45** **n=154**	**46+** **n=45**	
Read entire email spam					
Never	24	27.1	20.1	22.2	0.096
Sometimes	37.4	40.2	48.7	33.3	

Question	**Age Groups**				**P***
	15-25 **n=463**	**26-35** **n=358**	**36-45** **n=154**	**46+** **n=45**	
Always	7.8	4.7	5.8	2.2	
Delete email SPAM without reading					
Never	8.4	3.6	4.5	0	**0.005**
Sometimes	37.8	39.1	42.2	26.7	
Always	23.3	31	29.2	40	
Notify ISP about email SPAM					
Never	54.4	60.3	56.5	53.3	0.256
Sometimes	11	8.1	12.3	2.2	
Always	3	2.8	4.5	2.2	
Respond to email spam (%YES)	20.7	22.3	18.8	17.8	0.771
Positive impact of email spam					
Purchasing and selling	3.7	3.6	3.9	4.4	0.993
Learning	10.2	8.4	7.1	8.9	0.668
Fun	12.1	14.2	10.4	11.1	0.623

*P values are based on chi-square test between public users in different age groups; P values <0.05 were considered statistically significant.

4.1.4.4 Dealing of public users with email spam, based on nationality

Table 4.33 revealed significant differences between Saudis and non-Saudis in their dealing with email spam. The percentage of Saudis who always read email, was higher than the percentage of non-Saudis (6.9% vs 2.4%, p=0.006). The percentage of non-Saudis who always notified ISPs was higher than the percentage of Saudis (6.7% vs 2.5%, p=0.014).

The percentage of Saudis who responded to email spam was higher than the percentage of non-Saudis (22% vs 15.2%, p=0.048). Saudis interacted more with fun offered in email spam than non-Saudis (13.6% vs 7.3%, p=0.025).

Table 4.33: Percentages of dealing of public users with email spam, based on nationality

Question	**Nationality**		**P***
	Saudi **n=855**	**Non-Saudi** **n=165**	
Read entire email spam			
Never	24.8	22.4	**0.006**
Sometimes	**37.8**	**50.9**	
Always	6.9	2.4	
Delete email SPAM without reading			
Never	6.4	2.4	0.063
Sometimes	37.3	44.2	

Question	Nationality		P*
	Saudi n=855	Non-Saudi n=165	
Always	27.3	29.7	
Notify ISP about email spam			
Never	56.1	60	**0.014**
Sometimes	9.9	9.1	
Always	2.5	6.7	
Respond to email spam (%YES)	22	15.2	**0.048**
Positive impact of email SPAM			
Purchasing and selling	4	2.4	0.335
Learning	8.8	**10.3**	0.530
Fun	**13.6**	7.3	**0.025**

*P values are based on chi-square test between Saudi and non-Saudi users; P values <0.05 were considered statistically significant.

4.1.4.5 Dealing of public users with email spam, based on education level

The results, as summarised in Table 4.34, showed a significant difference between users in different education levels in terms of dealing with email spam. The percentage of users who had completed a PhD and who always deleted email spam was higher than the percentages of users who had completed other degrees (36.2%, p=0.005).

Table 4.34: Percentages of dealing of public users with email spam, based on education level

Question	Education Level					P*
	HS n=145	D n=49	B n=588	M n=144	PhD n=94	
Read entire email spam						
Never	22.8	26.5	25	24.3	22.3	0.113
Sometimes	35.9	32.7	37.8	49.3	48.9	
Always	9	10.2	6.5	3.5	2.1	
Delete email SPAM without reading						
Never	11	6.1	6	1.4	3.2	**0.005**
Sometimes	33.1	36.7	37.1	49.3	39.4	
Always	23.4	28.6	27	28.5	36.2	
Notify ISP about email spam						
Never	56.5	57.1	54.1	64.6	61.7	0.200
Sometimes	7.6	12.2	11.4	8.3	4.3	
Always	3.4	0	3.1	2.1	6.4	
Respond to email spam (%YES)	23.4	34.7	19.7	21.5	16	0.087
Positive impact of						

Question	Education Level					P*
	HS n=145	D n=49	B n=588	M n=144	PhD n=94	
email spam						
Purchasing and selling	4.8	8.2	2.6	4.9	5.3	0.157
Learning	12.4	16.3	8.2	8.3	6.4	0.157
Fun	15.2	18.4	11.6	13.2	10.6	0.508

*P values are based on chi-square test between public users in different education level; P values < 0.05 were considered statistically significant
Abbreviations: HS = High School, D = Diploma, B = Bachelor, and M = Master.

4.1.4.6 Dealing of public users with email spam, based on study discipline

Table 4.35 indicated that there were significant differences between users in different study disciplines in the way they dealt with email spam. The percentage of users who always read email spam was higher in the area of physical and biological sciences (15.1%) than in other areas (p=0.005). The percentage of users, who always deleted email spam, was greater in the area of computer science and information technology (36.5%) than in other areas (p=0.002).

Table 4.35: Percentages of dealing of public users with email spam, based on study discipline

Question	Study Discipline						P*
	E&T n=159	CS&IT n=293	SS n=93	P&BS n=93	HS&M n=88	Other n=149	
Read entire email SPAM							
Never	28.3	27.6	20.4	18.3	19.3	24.8	**0.005**
Sometimes	37.7	44.7	46.2	33.3	42	35.6	
Always	5.7	3.4	6.5	15.1	5.7	4	
Delete email SPAM without reading							
Never	3.8	3.4	4.3	7.5	8	6	**0.002**
Sometimes	44	38.2	48.4	45.2	35.2	29.5	
Always	25.2	36.5	20.4	15.1	25	30.9	
Notify ISP about email spam							
Never	59.1	63.8	54.8	47.3	54.5	49	0.072
Sometimes	6.9	7.8	15.1	16.1	6.8	13.4	
Always	3.8	3.4	2.2	3.2	3.4	2	
Respond to email spam(%YES)	18.2	20.1	23.7	26.9	18.2	18.8	0.553
Positive impact of email spam							
Purchasing and selling	4.4	3.1	2.2	4.3	0	6	0.208
Learning	8.8	8.2	10.8	15.1	5.7	4.7	0.09
Fun	11.9	11.6	15.1	14	11.4	10.7	0.919

*P values are based on chi-square test between public users in different study disciplines; P values <0.05 were considered statistically significant.
Abbreviations: E&T = Education and Teaching, CS&IT = Computer Science and Information Technology, SS = Social Sciences, P&BS = Physical and Biological Sciences, and HS&M = Health Sciences and Medicine.

4.1.4.7 Dealing of public users with email spam, based on work status

Table 4.36 showed a significant difference between students and employees in terms of deleting email spam. The percentage of students who did not delete email spam was higher than the percentage of employees (8.1% vs 3.9%, p<0.001).

Table 4.36: Percentages of dealing of public users with email spam, based on work status

Question	**Work Status**		**P***
	Student **n=455**	**Employee** **n=565**	
Read entire email spam			
Never	22.9	25.7	0.233
Sometimes	38.5	41.1	
Always	7.5	5.1	
Delete email spam without reading			
Never	8.1	3.9	**<0.001**
Sometimes	38.7	38.2	
Always	22.2	32	
Notify ISP about email spam			
Never	55.4	57.9	0.644
Sometimes	9.5	10.1	
Always	3.7	2.7	
Responding to email spam (%YES)	20	21.6	0.534
Positive impact of email spam			
Purchasing and selling	3.1	4.2	0.326
Learning	8.6	9.4	0.654
Fun	80	78.4	0.534

*P values are based on chi-square test between student and employee users; P values <0.05 were considered statistically significant.

4.1.4.8 Dealing of public users with email spam, based on work position

As summarised in Table 4.37, there was a significant difference between users in different work positions in terms of deleting email spam. The percentage of users who always deleted email spam without reading it was higher in technical positions than in other positions (47.1%, p=0.022).

Table 4.37: Percentages of dealing of public users with email spam, based on work position

Question	Work Position					P*
	EP n=274	MP n=58	TP n=91	MTP n=97	Other n=45	
Read entire email spam						
Never	25.2	19	34.1	20.6	35.3	0.102
Sometimes	44.2	41.1	37.4	36.1	29.4	
Always	5.1	3.4	2.2	10.3	2.9	
Delete email SPAM without reading						
Never	4.4	3.4	4.4	3.1	2.9	**0.022**
Sometimes	42	36.2	29.7	42.3	23.5	
Always	31	24.1	47.1	20.6	44	
Notify ISP about email spam						
Never	61.3	56.9	60.4	50.5	50	0.447
Sometimes	8.4	6.9	8.8	13.4	14.7	
Always	3.3	0	4.4	1	2.9	
Responding to email spam (%YES)	19	17.2	24.2	30.9	20.6	0.133
Positive impact of email spam						
Purchasing and selling	5.5	0	6.6	1	5.9	0.118
Learning	9.5	5.2	11	13.4	2.9	0.299
Fun	10.2	13.8	14.3	15.5	17.6	0.522

*P values are based on chi-square test between public users in different work positions; P values <0.05 were considered statistically significant.

Abbreviations: EP = Educational Positions, MP = Medical Positions, TP = Technical Positions, and MTP = Management Positions.

4.1.5 The Effects of Email Spam on the Performance of Public Users

This section describes the effects of email spam on the performance of users, and compares these effects among users based on demographic information. As summarised in Table 4.38, about half of the participants (45.1%, 95%CI: 42.1%-48.2%) had been affected negatively by spam. Most participants (28.1%, 95%CI: 25.4%-31%) said that email inboxes were filled with spam while the lowest percentage (8.7%, 95%CI: 7.1%-10.6%) said that they felt less confidence in using email.

Table 4.38: Percentages of effects of email spam on the performance of public users

Question	Frequency	Percentage (%)	95% CI
Effects of email spam on the performance of public users			
Yes	460	45.1	42.1-48.2
No	560	54.9	51.8-57.9
Negative impact of email spam			
Stealing personal information, e.g. password	92	9	7.4-10.9

Question	Frequency	Percentage (%)	95% CI
Losing time and reducing productivity	198	19.4	17.1-21.9
Less confidence in using email	89	8.7	7.1-10.6
Filling email inbox	287	28.1	25.4-31
Computer infection by malicious programs, e.g. viruses	250	24.5	21.9-27.2

4.1.5.1 The effects of email spam on the performance of public users in different regions

Table 4.39 showed that users in the northern region (52.3%) were more affected by email spam than users in other regions (p=0.029). Users in the southern region (36.6%) were more affected by email spam through inboxes filling email with spam, than users in other regions (p=0.046).

Table 4.39: Percentages of effects of email spam on the performance of public users in different regions

Question	Region					P*
	E n=203	W n=201	C n=352	S n=134	N n=130	
Effects of email spam on the performance of public users (%YES)	43.3	36.8	45.7	51.5	52.3	**0.029**
Negative impact of email SPAM						
Stealing personal information, e.g. password	9.9	8	8.2	11.9	8.5	0.708
Losing time and reducing productivity	19.7	18.9	20.2	18.7	18.5	0.990
Less confidence in using email	10.8	8.5	9.9	3.7	7.7	0.186
Filling email inbox	22.7	24.4	29.8	36.6	29.2	**0.046**
Computer infection by malicious programs, e.g. viruses	23.6	18.9	26.7	22.4	30.8	0.112

*P values are based on chi-square test between public users in different regions; P values <0.05 were considered statistically significant.

Abbreviations: E = Eastern, W = Western, C = Central, S = Southern, and N = Northern.

4.1.5.2 The effects of email spam on the performance of males and females

Table 4.40 revealed no significant differences between males and females in terms of the effects of email spam on their performance.

Table 4.40: Percentages of effects of email spam on the performance of male and female users

Question	Gender		P*
	Male n=608	Female n=412	
Effects of email spam on the performance of public users (%YES)	43.3	47.8	0.151
Negative impact of email spam			
Stealing personal information, e.g. password	8.9	9.2	0.852
Losing time and reducing productivity	20.1	18.4	0.521
Less confidence in using email	9.4	7.8	0.372
Filling email inbox	28.5	27.7	0.785
Computer infection by malicious programs, e.g. viruses	23	26.7	0.181

*P values are based on chi-square test between male and female users; P values <0.05 were considered statistically significant.

4.1.5.3 The effects of email spam on the performance of public users in different age groups

There were significant differences, as shown in Table 4.41, between age groups in terms of the effects of email spam on their performance. Users aged 36-45 were more affected loss of time and reduced productivity than other age groups (28.6%, p<0.001). Users aged 26-35 were more affected by inboxes filling with spam than other age groups (33.2%, p=0.019).

Table 4.41: Percentages of effects of email spam on the performance of public users in different age groups

Question	Age Groups				P*
	15-25 n=463	26-35 n=358	36-45 n=154	46+ n=45	
Effects of email spam on the performance of public users (%YES)	40.8	48.3	51.3	42.2	0.058
Negative impact of email spam					
Stealing personal information, e.g. password	9.5	7.5	9.7	13.3	0.531
Losing time and reducing productivity	14.5	22.9	28.6	11.1	**<0.001**
Less confidence in using email	9.9	9.2	4.5	6.7	0.205
Filling email inbox	23.5	33.2	30.5	26.7	**0.019**
Computer infection by malicious programs, e.g. viruses	23.3	26	25.3	22.2	0.813

*P values are based on chi-square test between public users in different age groups; P values <0.05 were considered statistically significant.

4.1.5.4 *The effects of email spam on the performance of Saudis and non-Saudis*

As seen in Table 4.42, non-Saudis were more badly affected by email spam than Saudis (53.9% vs 43.4%, p=0.013).

Table 4.42: Percentages of effects of email spam on the performance of Saudi and non-Saudi users

Question	Nationality		P*
	Saudi n=855	Non-Saudi n=165	
Effects of email spam on the performance of public users (%YES)	43.4	53.9	**0.013**
Negative impact of email spam			
Stealing personal information, e.g. password	8.4	12.1	0.129
Losing time and reducing productivity	18.8	22.4	0.285
Less confidence in using email	9	7.3	0.470
Filling email inbox	27.7	30.3	0.499
Computer infection by malicious programs, e.g. viruses	23.7	28.5	0.195

*P values are based on chi-square test between Saudi and non-Saudi users; P values <0.05 were considered statistically significant

4.1.5.5 *The effects of email spam on the performance of public users in different education levels*

Table 4.43 shows email spam wasted the time and reduced the productivity of users who had completed a PhD degree more than users with other degrees (29.8%, p=0.014).

Table 4.43: Percentages of effects of email spam on the performance of public users in different education levels

Question	Education Level					P*
	HS n=145	D n=49	B n=588	M n=144	PhD n=94	
Effects of email spam on the performance of public users (%YES)	44.1	40.8	42.5	52.8	53.2	0.095
Negative impact of email spam						
Stealing personal information, e.g. password	9	4.1	8.7	9	13.8	0.381
Losing time and reducing productivity	15.2	16.3	17.7	25	29.8	**0.014**
Less confidence in using email	13.1	8.2	7.5	11.1	6.4	0.176
Filling email inbox	29	30.6	25.5	34	33	0.217
Computer infection by malicious programs, e.g. viruses	26.2	32.7	21.8	30.6	25.5	0.122

*P values are based on chi-square test between public users in different education level; P values <0.05 were considered statistically significant.

Abbreviations: HS = High School, D = Diploma, B = Bachelor, and M = Master.

4.1.5.6 The effects of email spam on the performance of public users in different study disciplines

Table 4.44 shows significant differences between users in different study disciplines in the effects of email spam on their performance. Users who studied social sciences were more affected by email spam than users in other areas (59.1%, p=0.001). The main effect on the performance of users in the area of social science was inboxes filling with spam (35.5%), and this percentage was higher than the percentages in other areas (p=0.04).

Table 4.44: Percentages of effects of email spam on the performance of public users in different study disciplines

Question	Study Discipline						P*
	E&T n=159	CS&IT n=293	SS n=93	P&BS n=93	HS&M n=88	Other n=149	
Effects of email spam on the performance of public users (%YES)	49.1	47.1	59.1	30.1	44.3	38.9	**0.001**
Negative impact of email spam							
Stealing personal information, e.g. password	13.8	8.9	10.8	6.5	5.7	6.7	0.171
Losing time and reducing productivity	19.5	21.2	26.9	21.5	17	15.4	0.353
Less confidence in using email	5.7	10.2	6.5	7.5	4.5	9.4	0.377
Filling email inbox	31.4	30.7	35.5	18.3	22.7	23.5	**0.04**
Computer infection by malicious programs, e.g. viruses	23.9	24.6	31.2	16.1	25	24.2	0.323

*P values are based on chi-square test between public users in different study disciplines; P values <0.05 were considered statistically significant.

Abbreviations: E&T = Education and Teaching, CS&IT = Computer Science and Information Technology, SS = Social Sciences, P&BS = Physical and Biological Sciences, and HS&M = Health Sciences and Medicine.

4.1.5.7 The effects of email spam on the performance of students and employees

Employees were more affected by email spam than students (48.7% vs 40.7%, p=0.011), as shown in Table 4.45. They were also more affected by their inboxes filling email inboxes with spam than students (31.7% vs 23.7%, p=0.005). Employees were also more affected through loss of time and reduced productivity than students (24.1% vs 13.6%, p<0.001).

Table 4.45: Percentages of effects of email spam on the performance of student and employee users

Question	Work Status		P*
	Student n=455	Employee n=565	
Effects of email spam on the performance of public users (%YES)	40.7	48.7	**0.011**
Negative impact of email spam			
Stealing personal information, e.g. password	8.1	9.7	0.374
Losing time and reducing productivity	13.6	24.1	**<0.001**
Less confidence in using email	9.5	8.1	0.462
Filling email inbox	23.7	31.7	**0.005**
Computer infection by malicious programs, e.g. viruses	23.1	25.7	0.340

*P values are based on chi-square test between student and employee users; P values <0.05 were considered statistically significant.

4.1.5.8 The effects of email spam on the performance of public users in different work positions

Table 4.46 shows that no significant differences between users in different work positions in terms of the effects of email spam on their performance.

Table 4.46: Percentages of effects of email spam on the performance of public users in different work positions

Question	Work Position					P*
	EP n=274	MP n=58	TP n=91	MTP n=97	Other n=45	
Effects of email spam on the performance of public users (%YES)	51.1	43.1	50.5	47.4	35.3	0.408
Negative impact of email spam						
Stealing personal information, e.g. password	12	8.6	4.4	9.3	11.8	0.312
Losing time and reducing productivity	25.2	20.7	28.6	19.6	20.6	0.583
Less confidence in using email	8.4	3.4	6.6	11.3	8.8	0.497
Filling email inbox	33.2	17.2	39.6	28.9	29.4	0.063
Computer infection by malicious programs, e.g. viruses	26.3	27.6	27.5	26.8	17.6	0.837

*P values are based on chi-square test between public users in different work positions; P values <0.05 were considered statistically significant.

Abbreviations: EP = Educational Positions, MP = Medical Positions, TP = Technical Positions, and MTP = Management Positions.

4.2 Discussion

This section discusses public users' perceptions of email spam, their awareness of anti-spam filters and the efforts to combat it, its effects on their performances, and their dealing with it.

4.2.1 The Awareness of Public Users about Email Spam, Anti-spam Filters and the Efforts to Combat it

The results showed that public users had a relatively low level of awareness about email spam and anti-spam filters. As described in the results section, about two-thirds of the participants (62%, 95%CI: 59%-64.9%) were aware of email spam. A study of Malaysian users' experience with email spam revealed that 86.5% of Malaysian users were aware of email spam (Bujang & Hussin 2010), which indicates greater awareness among Malaysian than Saudi users. This could be because the Malaysian users who participated in the study were more experienced in using the Internet and email than the Saudi users were. It may be fruitful, then, for the Saudi Government to focus on public users' awareness about email spam, perhaps by conducting information campaigns (Pfleeger & Bloom 2005). According to Frost and Udsen (2006), a combination of a number of different effective efforts by businesses and ISPs, such as user education, is needed.

Most public users defined email spam as "email that was sent randomly and contained malicious programs such as Viruses". This definition was different from the international definition of spam as UBE (El-Halees 2009; Zaidan et al. 2011) and UCE (Boykin & Roychowdhury 2004; Cheng 2004) and that of public users in other countries, such as Greece, as UBE and UCE (Pallas & Patrikakis 2005). This suggests that there is scope for an agreed definition for email spam that could be used for designing strategies and policies to combat spam in Saudi Arabia, such as enacting laws and developing anti-spam techniques for different languages used in spam (e.g. Arabic) (Everett 2004).

The results showed that about one-third of the participants (37.9%, 95%CI: 35%-40.9%) were aware of anti-spam filters. The study by Bujang and Hussin (2010) of Malaysian public users revealed that 66.9% were aware of the filters. Al-A'ali (2007) showed that only 26% of the participants in Bahrain were aware of anti-spam filters. It is clear from these studies that Malaysian users were more aware of anti-spam filters than Saudi users, although Saudi users knew more about them than Bahraini users. Despite Saudi users being more aware of email spam than users in other countries such as Bahrain, their awareness is still low. This suggests that there should be a focus on raising the awareness of public users about anti-spam filters, and how they use these filters, to reduce the volume of email spam in Saudi Arabia (Dantin & Paynter 2005).

Compared with the finding by Bujang and Hussin (2010) that only 14.6% of Malaysian users were aware of the services and efforts provided by the Malaysian Government, in this study, nearly a quarter of the Saudi participants (24.4%, 95%CI: 21.9%-27.1%) were aware of government efforts to combat email spam. This indicates that the Saudi users were more aware of the efforts provided to combat email spam than Malaysian users, although Malaysian users knew more about email spam and anti-spam filters than Saudi users. This might be because the Malaysian Government's ways of informing users about the efforts were not promoted as well as those of the Saudi Government's ways were. Nevertheless, most Saudi users were unaware of these efforts, which suggests that the Saudi government's public awareness program needs to be improved (Frost & Udsen 2006).

The results also indicated that few public users (13.6%, 95%CI: 11.6%-15.8%) were aware of the ISPs' efforts to combat email spam. This might be because the ISPs did not promote their efforts well, or inform public users of their efforts. This suggests that the ISPs should use more effective ways to inform public users of their efforts to combat email spam (Pallas & Patrikakis 2005).

4.2.2 The Nature of Email Spam Received by Public Users

This section discuss the results of survey questions about the nature of email spam received by public users, such as the volume of spam that they received, its languages and the various types of Arabic and English spam.

4.2.2.1 The volume of email spam received by public users

About three-quarters of the participants (73.1%, 95%CI: 70.4%-75.8%) reported receiving email spam, and nearly half of the participants (46.4%, 95%CI: 42.6%-49.7%) received more than 25 emails spam weekly. A study conducted in the USA by Grimes, Hough and Signorella (2007) showed that 95% of the respondents received email spam, a higher percentage than for users in Saudi Arabia. This might be because American users used the Internet for online shopping and banking more than Saudi users (Hermanson 2003), which could generate more spam (Hassanein & Head 2007). It is also possible that some Saudi respondents received email spam without realising it.

4.2.2.2 The languages of email spam received by public users

The results found that most of the email spam received by public users in Saudi Arabia was written in English, followed by Arabic, which is the native language of Saudi Arabia. This finding agrees with the results of studies conducted in other countries, such as Bahrain, Malaysia and Greece, which found that most email spam was written in English, which is the most used language in the world (Altbach 2004; Huddleston & Pullum 2002; Kirkpatrick 2007), followed by the native languages of these countries. Al-A'ali (2007) indicated that 64% of the respondents in Bahrain received English email spam, whereas 18% received Arabic email spam. Bujang and Hussin (2010) found that most of the email spam received in Malaysia was written in English, followed by Malay. The study by Pallas and Patrikakis (2005) reported that in Greece most of the email spam received was written in English and Greek. From these studies, it can be clearly seen that spammers attempt to reach more recipients through English, followed by the native language of the country, such as Arabic in Saudi Arabia and Bahrain, Malay in Malaysia and Greek in Greece.

4.2.2.3 The types of Arabic and English email spam received by public users

The results found that significant differences between the Arabic and English email spam received by public users. The percentage of emails related to forums was larger in Arabic email spam than in English email spam. There are two possible reasons for this. The first is that forums are a popular way for Saudi people to discuss their everyday experiences and needs, such as purchasing and selling, housing, study, religion, or even just for fun and personal communications (Al-Saggaf 2004). This may prompt users to subscribe to Arabic forums to fulfil this social need, as Arabic is the formal language of the Saudi people (Chejne 2009). As a result, their addresses would be added to the forums' mailing lists, resulting in receiving more Arabic spam from this source than English spam. The second reason could be that some Arabic forum managers or owners seek to increase their subscriber numbers by using automated software to collect a large number of email addresses (Andreolini et al. 2005), or by buying them from other forums (Cook et al. 2006). Either way, forum managers could then send messages to the collected email addresses containing a welcome message and an activation link. This suggests that it is important for the government to control emails related to forums by providing guidelines for managers or owners, and applying penalties to those not in compliance.

The percentage of political and religious emails was higher in Arabic email spam than in English email spam. This could be because some Arab countries use Internet-based media for publishing their political campaigns to get more voters (Grossman 2004; Sweet 2003), or for religious purposes (Martinkova 2008). This could explain the higher percentage of this type of email spam in Arabic than in English.

The reason for pornographic spam appearing more frequently in English email spam than in Arabic email spam is likely to be because pornographic email is prohibited by the Islam religion (Al-A'ali 2007), and also conflicts with the Arabic culture, which forbids this type of content (Abdoh, Musa & Salman 2009). The reason for more phishing and fraud emails appearing more in English email spam than in Arabic may be because the organised criminal elements behind most email phishing and fraud attempts are not yet operating or as established in Arabic-speaking countries as much as in English-speaking countries (Ramanathan & Wechsler 2012). They did, however, exist in Arabic spam and would be more easily understood by users than English ones, and would be more likely to prompt users to interact with it (Alnajim & Munro 2009). This has the potential to grow and increase electronic fraud transactions in Saudi Arabia. It follows that the government should combat and control this type of email spam, as it aims to steal identities and money from the recipients, and has cost other countries millions of dollars. The IFCC in the USA, which deals with users' complaints about Internet fraud, estimated that the cost to consumers of online fraud was $17.8 million in 2001 (Hinde 2002).

4.2.3 How Public Users Deal with Email Spam

The highest percentage of participants (39.9%, 95%CI: 36.9%-42.9%) sometimes read the entire email spam, and only about a quarter of the participants (27.6%, 95%CI: 25%-30.5%) always deleted email spam without reading it. A study conducted in the USA by Grimes, Hough and Signorella (2007) revealed that 66% of 205 participants deleted spam email, and Hermanson (2003), also in the USA, found that 82% of the participants in his study deleted spam email. The results of this study showed that the percentage of participants who deleted the email spam was larger in the USA than the percentage in Saudi Arabia. This indicated that public users in Saudi Arabia were less aware of ways of dealing with email spam than users in the USA. This could be because the USA is a developed country and the Internet started in the USA in 1960 (Crystal 2001), long before Saudi Arabia (1994) (CITC 2012). This could mean users in the USA have more experience with the Internet and its application, such as using email and dealing with spam, than Saudi users, which suggests that it is necessary to educate public users in Saudi Arabia about appropriate ways to deal with it.

Only a few public users in Saudi Arabia (3.1%, 95%CI: 2.2%-4.3%) always contacted their ISPs and notified them about email spam. A study conducted in the USA by Grimes, Hough and Signorella (2007) showed that 11.7% of American users contacted their ISPs when they received email spam, a greater percentage than Saudi users. This suggests that, to reduce the volume of email spam the ISPs could better communicate to public users their efforts to combat it and the necessary procedures when they receiving it.

4.2.4 The Effects of Email Spam on the Performance of Public Users

About half of the participants reported being negatively affected by email spam. More participants were affected by email inboxes filling with spam than anything else. This can lead to the consumption of available email storage capacity, and then the loss of important emails (Zhang, Zhu & Yao 2004). It can also lead to wasting users' time (Chigona et al. 2005; Hinde 2002; Özgür, Güngör & Gürgen 2004) and reducing their productivity (Leng 2006). About one-fifth of the Saudi participants reported loss of time and productivity due to receiving a large volume of email spam. This can cost the government and companies millions of dollars. Cook et al. (2006) stated that deleting spam manually from a user's inbox wastes time, an estimated cost to US companies of $10 billion in lost productivity.

The second most reported effect of email spam on the performance of public users was the infection of computers by malicious programs such as viruses (24.5%). This can be a way to steal important information from the recipients, such as passwords and bank accounts (Cournane & Hunt 2004; Hermanson 2003). A study conducted in South Africa indicated that 56% of the participants said that they received viruses from email spam (Chigona et al. 2005). By comparing these two studies, it can be shown that South African users were more affected in this way than Saudi users. Although the number of Saudi users affected by malicious programs through email spam was lower than other countries, this number could be expected to be increased in future, which indicates a need to increase the awareness of public users about the appropriate ways in dealing with spam to avoid its effects (Lugaresi 2004).

4.2.5 Demographic Information of Public Users and Email Spam Characteristics

This section discusses the results based on the demographic information such as region, gender, age group, nationality, education level, study discipline, work status, and work position.

4.2.5.1 Discussion of results, based on region

Statistical significant differences have been found in the characteristics of email spam received by public users from different regions. The awareness about email spam, anti-spam filters and the efforts to combat it was higher in the central region than in other regions. This could be because all government sectors that are responsible for the Internet, technology and communication, such as (CITC 2014) and (KACST 2014), are located in the central region (Riyadh). As mentioned in the results, most users in the southern and northern regions were more affected by email spam than users in other regions. This could be because they were less aware of email spam and how to deal with it than users in other regions, and this could make them victims of email spam. Users from the southern and northern regions reported that the most common positive impacts of email spam were learning and fun.

The average number of email spam received per user for each region was greater in the central region than in other regions, possibly because the central region is the largest region in Saudi Arabia, with a greater population than other regions (CDSI 2013), resulting in public users receiving more email spam than users in other regions. As well, more English email spam was reported in the central region than other regions, possibly because it contains the capital city, Riyadh, (Hamner & Al-Qahtani 2009). Riyadh is the political and economic centre of the country (Al-Majed, Murray & Maguire 2001) and contains more foreign people coming from overseas for work, business or study than other cities (CDSI 2013). This might encourage more interaction with English email spam, as English is the second language required for work or study in Saudi Arabia, in the absence of Arabic (Halligan 2006). It could also be that users in the central region communicate with people outside of Saudi Arabia more often, and so their email addresses are more likely to be found by English language spammers (Andreolini et al. 2005). These reasons could result in receiving more English spam than Arabic in the central region. By contrast, the percentage of phishing and fraud emails received was larger in the western region than in other regions. This could be because users in the western region used online banking or shopping more than users in other regions, as online banking and businesses is one of the categories targeted by attackers (Ramanathan & Wechsler 2012). Hassanein and Head (2007) reported that "users provided their credit card numbers in online shopping could make them vulnerable to credit card fraud."

4.2.5.2 Discussion of results, based on gender

There were significant statistical differences between males and females in the reported characteristics of email spam. Males were more aware of email spam and anti-spam filters than females. Email spam is a security threat (Lam & Yeung 2007), therefore this finding supports a study conducted by Johnson and Koch (2006) which revealed that males were more aware of security threats than females (Johnson & Koch 2006). A possible reason could be that males had more experience in using the Internet and email than females (Sait et al. 2008) and so were more aware of email spam and anti-spam filters. This suggests the need to increase the awareness in females of email spam.

Males received more email spam than females. Significant differences were found in the percentage of pornographic emails reported by males and females, with males receiving more than females. These results are in line with the results of other studies by Al-A'ali (2007), and Grimes, Hough and Signorella (2007). It is possible that males visit and subscribe to pornographic websites and interact with the contents more than females do. Goodson et al. (2001) found that "males were significantly more likely to have accessed the Internet to view sexually explicit materials and to claim curiosity about sex as their motivation for this behaviour". It might also be that females felt ashamed to report visiting such websites (Al-A'ali 2007), resulting in males appearing to receive more pornographic emails.

Males also received more product and service emails than females, in this study. This conflicts with the results of Al-A'ali's (2007) study, which revealed that females in Bahrain received more commercial emails than males, and because females liked this type of email, they did not consider such emails as spam. It might also be that males used online shopping to buy products by the Internet more often than females, resulting in receiving more products and services spam than females (LaRose & Rifon 2007). According to Hassanein and Head (2007):

> In an online shopping context, consumers are vulnerable and likely to expose themselves to loss if they provide their email address (making themselves vulnerable to receiving spam email or other annoyances).

4.2.5.3 Discussion of results, based on age group

In terms of the average number of email spam, the results have shown that users aged 26-35 received more email spam than other age groups. Similarly, users aged 26-35, as mentioned in the results, were more affected by email spam filling inboxes than were users in other age groups. This could be because users aged 26-35 published, subscribed or entered their email addresses on websites that could be used by spammers to obtain the recipients' information (Wood 2013), making them a target for spammers.

Types of Arabic and English email spam received by users in different age groups were significantly different. Users aged 15-25 received more pornographic emails than other age groups, possibly because they were young and liked to interact with the pornographic materials from spam, and therefore subscribed and added their personal data to pornographic websites, more than other age groups. This could encourage spammers to send more pornographic emails to young users and make them more targeted than other age groups. According to Grimes, Hough and Signorella (2007), "certain types of spam may be more distasteful to some age groups than others". This result is compatible with the results of the Al-A'ali (2007) study, which showed that younger users received more pornographic emails than users in other age groups. The author stated as a possible reason that:

> … older users are probably married and with children and they are more concerned about other family members viewing pornography and they are more mature than younger users who tend to like viewing this kind of material (Al-A'ali 2007).

The percentage of emails related to forums received by users aged 26-35 was higher than that received by other age groups, possibly because they register in forums and add their email addresses, which spammers can use to send more emails related to forums, such as subjects, replies or joining new forums (Hayati et al. 2010).

Users aged 46 and older received more phishing and fraud emails than other age groups. One reason could be that users in this age group purchase online more than younger users (Abdul-Muhmin & Al-Abdali 2011), which lead to receiving a higher percentage of phishing and fraud emails than other age groups. It might also be that older people are seen as easier targets for this type of spam, as they lack the technical literacy that would enable them to detect it. For example, it is believed that the Nigerian 419 scams purposely craft spam that only the gullible would believe, because this is exactly the population that is likely to fall for the rest of the scam (Glickman 2005; Tive 2006). This finding agrees with that of Grimes, Hough and Signorella (2007), whose study results revealed that older respondents made more online purchases than younger users.

4.2.5.4 Discussion of results, based on nationality

The study found that non-Saudis had a greater level of awareness about email spam, anti-spam filters and efforts to combat it in Saudi Arabia than Saudis. This greater level of awareness of email spam and anti-spam filters could be because non-Saudis had more experience in using the Internet and making online transactions (Abdul-Muhmin & Al-Abdali 2011). Lacking experience and awareness, Saudis would be more likely to have responded to email spam than non-Saudis, thinking of it as fun. This suggests a need for greater awareness in Saudis about effects of email spam, and appropriate ways to deal with it.

Non-Saudi users also received more English email spam than Saudi users. This could be because non-Saudi users in Saudi Arabia were required to use English as a first or second language in their work or study if they did not speak Arabic (Halligan 2006), and be more likely to understand and interact with it. Thus, this could encourage spammers to send more English email spam to non-Saudi users. It might also be that non-Saudi users subscribed to or added their email addresses to English forums or websites, enabling their email addresses to be harvested by English spammers (Schryen 2007). The percentage of Arabic email spam received by Saudi users was larger than that received by non-Saudi users, possibly because Arabic is the first language of the Saudi people (Aldossary, While & Barriball 2008), enabling them to respond and deal with Arabic email spam. However, it might also be the case that Saudi email addresses are more exposed on Arabic forums and other websites, which would make them more likely to be harvested by purveyors of Arabic spam (Schryen 2007) and leading to them receiving more Arabic email spam than non-Saudi users.

The results indicated that Saudi users received more product and service emails than non-Saudi users. This could be because more Saudis purchased products from the commercial websites than non-Saudis, which could result in them receive more products and services emails (Hassanein & Head 2007; LaRose & Rifon 2007). A further reason that this is likely to be more common in Arabic is that product and service spam is usually sent by real businesses, and so it is easier to enforce anti-spam laws on them. Thus this type of spam has become much less common in English since the USA (Sorkin 2009), Australia (Cheng 2004) and other English-speaking countries have implemented such laws.

The percentage of phishing and fraud emails in Arabic received by Saudi users was larger than that received by non-Saudi users. This might be because Saudis use Internet banking more than non-Saudis, which could be a possible reason for them receiving more phishing and fraud emails than non-Saudis; banking online could be a way to receive more phishing and fraud emails (Ramanathan & Wechsler 2012). Mohamed (2011) found statistical differences between Saudis and non-Saudis in their use of Internet banking, with more Saudis than non-Saudis banking in this way.

4.2.5.5 Discussion of results, based on education level

There were significant differences in this study in public users' awareness of email spam and anti-spam filters, based on their level of education. More users who had completed PhD had a greater awareness of email spam and effective ways of dealing with it than users who had completed other degrees. Those with PhD degree may well have had better knowledge of using the Internet and its applications than users with other degrees. According to Burke (2002), "education is often positively correlated with an individual's level of Internet literacy". Al-Somali, Gholami and Clegg (2008) stated that the higher level of education of users had a significant impact on using the Internet. Stepanikova and Zheng (2004) noted that "using the Internet depends largely on a person's education level, and the more educated are more likely to use the Internet than those with less education". This reason could make users with a higher level of education level more aware of email spam and anti-spam filters than users with other education levels, compared with users with lower educational attainment, even if they each receive equal volumes of spam.

In this study, the percentage of pornographic emails received by users in high school was higher than that received by users with higher education. As those users in high school are younger than those in other degrees , they could be particularly attracted to subscribing to pornographic websites and dealing with the sexual materials involved in email spam (Grimes, Hough & Signorella 2007). This in turn might lead to them receiving more frequent pornographic emails.

The percentage of phishing and fraud emails received by users who had completed a PhD was larger than that received by users completed other degrees, possibly because those who had completed a PhD degree were more aware of the technology than users in other degrees (Al-Somali, Gholami & Clegg 2008). This might explain their higher level of experience in using online transactions, such as shopping and banking, than users in other education degrees. Studies such as Ramanathan and Wechsler (2012), and Hassanein and Head (2007) revealed that online shopping and banking could be a way for phishing and fraud spammers to target users. Hind (2003) stated that "security attacker is specifically targeting those who bank online". This could make users who had completed a PhD more prone to receiving phishing and fraud emails than users who had completed other degrees.

Users who had completed high school received more 'other' types of email spam such as fun, puzzles, competitions, greetings, and friendship invitations by social network websites such as Facebook, than users in other education degrees. This could be because users in high school used the Internet more for entertainment, such as playing online games and watching videos, than other users. A study conducted by

Huang and Chou (2010) showed that most common reason for high school students to use the Internet (87%) was entertainment, such as online games and listening to music.

4.2.5.6 Discussion of results, based on study discipline

Significant differences were found between users in different study disciplines in terms of the awareness about email spam and anti-spam filters, and types of email spam. Users who studied computer science and information technology had a greater awareness of email spam and anti-spam filters than those in other disciplines. The most common source of knowledge about email spam and anti-spam filters of users who studied computer science and information technology was through school and university education. The results showed that the percentage of users who gained their knowledge through school and university education was larger in this discipline than in all other study disciplines. This could be a significant indicator that the study of computer science and information technology can increase the awareness of users about email spam and anti-spam filters. Another possibility could be that users in the discipline of computer science and information technology worked in technical positions more than users in other disciplines. Such work might increase their relative level of knowledge about email spam and anti-spam filters (Awawdeh & Tubaishat 2014).

The results also showed a higher average number of email spam received by users who studied computer science and information technology than those in other study disciplines, and the most common type email spam they received was product and services emails. This might be because users in the discipline of computer science and information technology have more technical knowledge than users in other disciplines, which can help in using the Internet for online shopping, and banking, and making online payments (Nishioka, Murayama & Fujihara 2012), in turn attracting more products and services advertisements (Hassanein & Head 2007).

4.2.5.7 Discussion of results, based on work status

The results showed that employees' awareness of email spam, anti-spam filters and how to combat it was higher than that of students. This could be because they had more experience in using the Internet and email. A study by Grimes, Hough and Signorella (2007) revealed that "students took fewer actions against spam, used the computer less, and spent fewer hours online than employees".

The results showed that employees received more business emails than students. This might be because employees reply to, or become interested in the business offers made by spam emails (Ridzuan, Potdar & Talevski 2010), leading to loss of work time and productivity. The results showed that employees were more affected by time wasting and loss of productivity than students. The percentage of students who received pornographic emails in both Arabic and English was greater than employees. This could be because most students were younger than employees, and they like visiting and subscribing to pornographic websites and interacting with sexual materials involved in spam more than employees do (Joseph 2008). This could encourage spammers to send pornographic emails to the addresses added to students' profiles on those websites.

4.2.5.8 Discussion of results, based on work position

The results pointed to significant differences between users in different work positions in terms of the characteristics of email spam reported. Users who worked in technical positions showed greater awareness about email spam, anti-spam filters, ways of dealing with email spam and efforts to combat than that of users who worked in other work positions. This indicates that working in technical positions can increase users' awareness of about email spam and anti-spam filters. Awawdeh and Tubaishat (2014) reported that technical staff working in the Internet technology environment had more knowledge about information security issues than staff working in other environments.

The results also showed that users who worked in technical positions received more product and services emails than users in other work positions. Those in technical positions might have had more experience in using online shopping (Nishioka, Murayama & Fujihara 2012), making it easier for them to buy and sell in that way (LaRose & Rifon 2007), and in turn likely to receive more related emails than users in other work positions.

4.3 Conclusions

This chapter described, analysed and discussed perceptions of the public users about email spam and anti-spam filters and the efforts to combat it in Saudi Arabia, its effects, and their dealing with it, based on some demographic factors. These factors included: region, gender, age, nationality, education level, study discipline, work status, and work position. Some of these factors have been used in previous studies and this research sought to discover their effects on public email users in Saudi Arabia. The main conclusions of the user survey are matched to the research questions and described below.

The awareness of public users about email spam, anti-spam filters and efforts to combat it in Saudi Arabia was low. The results indicated that there was a deficiency in the efforts provided by relevant agencies in Saudi Arabia to increase the awareness of users, as the results found that the most common sources of their knowledge about spam and anti-spam filters were self-education through the Internet and forums. This indicates that government and relative agencies should focus on increasing the awareness of public users about email spam, anti-spam filters and how to combat it.

There was no consensus definition for email spam by public users in Saudi Arabia, and the most common definition of users for email spam was "an email that was sent randomly and contains malicious programs such as Viruses". Some definitions of email spam by public users agreed with the international definitions as UCE and as UBE. This suggests the need for an agreed definition of email spam in Saudi Arabia, and could help in enacting law to combat email spam in Saudi Arabia, and in developing anti-spam filters.

A number of email account providers were used by the public, the most common provider being Hotmail. The results indicated that most of the email spam received by public users was written in English, followed by Arabic. The results indicated that Saudi Arabia has its own spammers, Arabic being the second most common language of email spam received in Saudi Arabia, after English. The results found differences between Arabic and English email spam received by public users. Emails related to forums and religious and political emails were more common in Arabic spam, whereas pornographic emails and phishing and fraud emails were more common in English spam.

The results indicated that the Saudi society is at risk if the rate of phishing and fraud spam increases to match that seen in other countries, because of the lower awareness of spam and phishing. The potential negative economic impact is perhaps the single greatest reason for the Saudi Arabian Government to seek to improve Internet literacy rates among the population.

Public users in Saudi Arabia differed in how they dealt with email spam. The results showed that most users were not yet aware of appropriate ways to deal with it, such as by deleting it, or contacting ISPs about it. This suggests the need for relevant agencies to increasing users' awareness of about effective ways to dealing with email spam. Email spam had many negative effects on the performance of public users; and the most common of which was email inboxes filling with spam. This can waste users' time in reading it, deleting it, or filtering it, reducing productivity and potentially affecting the country's economic growth. This suggests the need for government and relative agencies to take action, such as by enacting law, to stop the spread of email spam in Saudi Arabia.

These results can be further interpreted by investigating the experiences of business users and ISPs about email spam and how they deal with it. Therefore, the following chapter will present and discuss the survey results for this demographic in Saudi Arabia: the nature of email spam, their awareness about it and effort to combat it, how businesses dealing business with it, and its effects on their performance. It will also analyse and discuss the effects on the results of many factors, such as business size, business sector, and establishment year of business.

Chapter 5: A Study of Email Spam Related Characteristics among Businesses in Saudi Arabia

This chapter presents the results of the email spam survey given to businesses in Saudi Arabia. The survey aimed to develop a better understanding of:

- the nature of email spam in Saudi Arabia from the businesses' perspective, including its impact on their performance
- the awareness of businesses about email spam, anti-spam filters and the efforts to combat it in Saudi Arabia.

Questionnaires were collected from 92 businesses located in the eastern, western and central regions. Businesses in the southern and northern regions were also invited to complete the survey, however, all requested that their head offices complete the survey; they had asked their head offices for permission to complete the survey, but it was not given. These businesses were active in different sectors, such as production and manufacturing, finance and investments, technology and telecommunication, and consulting services.

The chapter is divided into the following sections:

- Section 5.1: presents the results of the businesses' questionnaire.
- Section 5.2: discusses the results of the businesses' questionnaire.
- Section 5.3: presents the conclusions drawn from the results of the questionnire.

5.1 Results

This section presents the results revealing participants' regarding their awareness of email spam, anti-spam filters; efforts to combat spam; its effects on their performance; and how they dealt with it. This section analyses and describes the results based on factors such as business size, business sector and the year the business was established.

The following statistical tests were used to analyse the data: chi-square test (X^2), Fisher Exact test, independent samples t-test, paired sample t-test and one-way ANOVA test. A p-value less than 0.05 ($p<0.05$) was considered statistically significant. These tests are described and discussed in section 3.8.

5.1.1 Participants' Demographic Information

The demographic information of the businesses is described in Table 5.1. A total of 92 businesses from the eastern, western and central regions participated in this study, and the highest percentage of participants was from the central region (46.74%).

About 37% of businesses were classified as medium (50-249 employees), 32.6% as large (250 employees and more), and 30.4% as small (1-49 employees). Approximately 56% of businesses were old (established before 1994), whereas about 44% were classified as new (established 1994 or later). The largest sector of businesses (45.7%) was production and manufacturing; the smallest sector was finance and investment (8.7%).

Table 5.1: Percentage distribution of businesses in Saudi Arabia, based on their demographic information

Demographic Information	Frequency	Percentage (%)
Region		
Eastern	28	30.43
Western	21	22.83
Central	43	46.74
Business size		
Small (1-49 employees)	28	30.4
Medium (50-249 employees)	34	37
Large (250 employees and more)	30	32.6
Establishment year		
Before 1994[4] (old)	50	56.2
1994 till now (new)	39	43.8
Business Sector		
Production and manufacturing	42	45.7
Finance and investment	8	8.7
Technology and telecommunication	15	16.3
Consulting services	9	9.8
Other businesses[5]	18	19.6

5.1.2 The Awareness of Businesses about Email Spam and Anti-spam filters, and the Efforts to Combat it

This section describes the results for businesses awareness of email spam, anti-spam filters and efforts to combat it. It also compares the results based on business size,

[4] 1994 was the year Saudi Arabia began using the Internet (CITC 2012).

[5] Other businesses included bookstores, printing and packaging, cars sales and hiring, insurance companies, and housewares sales.

business sector and establishment year of business. These results are summarised in Table 5.2.

Most businesses (90.2%, 95%CI[6]: 82.9%-95%) were aware of email spam and anti-spam filters. There were a number of sources to inform businesses about email spam and anti-spam filters. As shown in Table 5.2, most businesses obtained information about email spam and anti-spam filters through the Internet and forums (75%, 95%CI: 65.5%-83%), while the lowest percentage obtained information from the government (7.6%, 95%CI: 3.5%-14.4%).

Most of the businesses (39.7%, 95%CI: 28.7%-51.6%) defined email spam as UCE, while relatively few businesses (5.9%, 95%CI: 2%-13.4%) defined email spam as annoying email that was not related to the recipients' work.

About a quarter of businesses (25%, 95%CI: 17%-34.5%) were aware of government efforts to combat email spam, while slightly fewer (23.9%, 95%CI: 16.1%-33.3%) were aware of ISPs efforts to combat email spam. The highest percentage of businesses (26.1%, 95%CI: 11.7%-46.1%) said that most of the efforts undertaken by the government to combat email spam were technical, and identified the government sectors responsible for conducting these efforts as the CITC and KACST. Most Saudi businesses thought that Saudi ISPs used mainly anti-spam filters to block email spam (54.5%, 95%CI: 34.3%-73.7%). Approximately 40%, (95%CI: 29.6%-49.3%) of businesses informed their customers and employees about email spam and the appropriate methods to combat it.

Table 5.2: Percentages of distribution of the awareness of businesses about email spam, anti-spam filters, and efforts to combat it

Question	Frequency	Percentage (%)	95% CI
Aware of email spam and anti-spam filters			
Yes	83	90.2	82.9-95
No	9	9.8	5-17.1
Knowledge source for email spam and anti-spam filters			
ISPs	23	25	17-34.5
Internet and forums	69	75	65.5-83
Broadcast media, e.g. TV	22	23.9	16.1-33.3
Government	7	7.6	3.5-14.4

[6] 95% confidence interval

Question	Frequency	Percentage (%)	95% CI
Other companies and organisations	23	25	17-34.5
Definition of email spam			
UBE	22	32.4	22.1-44
Sent by unknown senders without permission	10	14.7	7.8-24.5
Sent randomly, contain malicious programs, e.g. viruses	5	7.4	2.9-15.4
UCE	27	39.7	28.7-51.6
Annoying email unrelated to recipients' work	4	5.9	2-13.4
Aware of government efforts to combat spam			
Yes	23	25	17-34.5
No	69	75	65.5-83
Government efforts to combat spam			
Technical efforts by CITC and KACST	6	26.1	11.7-46.1
Awareness efforts by CITC	4	17.4	6.2-36.2
Receiving ISPs' reports regarding spam issues	2	8.7	1.9-25.1
Aware of ISPs' efforts to combat spam			
Yes	22	23.9	16.1-33.3
No	70	76.1	66.7-83.9
ISPs' efforts to combat spam			
Using anti-spam filters	12	54.5	34.3-73.7
Providing awareness information	3	13.6	4-32.1
Reporting spam-related issues to CITC	2	9.1	1.9-26.1
Educate employees and customers about email spam and anti-spam filters			
Yes	36	39.1	29.6-49.3
No	56	60.9	50.7-70.4

5.1.2.1 The awareness of businesses about email spam, and the efforts to combat it, by business size

As shown in Table 5.3, there was a significant difference between small, medium and large businesses in terms of their awareness of email spam and anti-spam filters. Large businesses were more aware than medium and small businesses (100%, p<0.001).

The percentage of large businesses that gained their knowledge about email spam from broadcast media and government was higher than the percentages of small and medium businesses (33.3%, p=0.042, 16.7%, p=0.043 respectively). Medium-sized businesses thought that the government provided no information about email spam and anti-spam filters.

Some businesses provided awareness programs about email spam and anti-spam filters for their employees and customers, and the percentage of businesses that provided these programs was greater for large businesses than for small and medium businesses (60%, p=0.015).

Table 5.3: Percentages of distribution of the awareness of businesses about email spam, anti-spam filters, and efforts to combat it, based on size

Question	**Number of Employees Business Size 1-49 Small n=28**	**50-249 Medium n=34**	**250+ Large n=30**	**P***
Aware of email spam and anti-spam filters (%YES)	71.4	97.1	100	**<0.001**
Knowledge source for email spam				
ISPs	17.9	35.3	20	0.214
Internet and forums	71.4	79.4	73.3	0.745
Broadcast media, e.g. TV	7.1	29.4	33.3	**0.042**
Government	7.1	0	16.7	**0.043**
Other companies and organisations	14.3	23.5	36.7	0.140
Aware of government efforts to combat spam (%YES)	14.3	20.6	40	0.059
Aware of ISPs' efforts to combat spam (%YES)	17.9	26.5	26.7	0.666
Educate employees and customers about email spam and anti-spam filters (%YES)	32.1	26.5	60	**0.015**

*P values are based on chi-square test between businesses in different sizes; P values <0.05 were considered statistically significant.

5.1.2.2 The awareness of businesses about email spam, and the efforts to combat it, based on business sector

Table 5.4 shows that the finance and investment, and technology and telecommunication sectors were more aware of email spam and anti-spam filters than other sectors (100%, p=0.014). The sector that relied the most on information from Internet and forums, and broadcast media, was the production and manufacturing sector (88.1%, p=0.005, 38.1%, p=0.038 respectively).

The finance and investment sector was more aware of the government's awareness information than were the other sectors (37.5%. p=0.009), while the consultation sector had no knowledge of the government's awareness programs. The percentage of businesses that educated their employees and customers about email spam and anti-spam filters was highest in the finance and investment sector (100%, p=0.002).

Table 5.4: Percentages of distribution of the awareness of businesses about email spam, anti-spam filters, and efforts to combat it, based on sector

Question	Business Sector					P*
	P&M n=42	F&I n=8	T&T n=15	CS n=9	Other n=18	
Aware of email spam and anti-spam filters (%YES)	95.2	100	100	66.7	77.8	**0.014**
Knowledge source of email spam						
ISPs	33.3	0	40	11.1	11.1	0.066
Internet and forums	88.1	25	73.3	66.7	72.2	**0.005**
Broadcast media, e.g. TV	38.1	0	6.7	22.2	16.7	**0.038**
Government	4.8	37.5	13.3	0	0	**0.009**
Other companies and organisations	19	50	33.3	11.1	27.8	0.286
Aware of government efforts to combat spam (%YES)	23.8	50	26.7	33.3	11.1	0.297
Aware of ISPs' efforts to combat spam (%YES)	23.8	12.5	33.3	33.3	16.7	0.688
Educate employees and customers about email spam and anti-spam filters (%YES)	35.7	100	40	44.4	16.7	**0.002**

*P values are based on Fisher's exact test between businesses in different sectors; P values <0.05 were considered statistically significant.

Abbreviations: P&M = Production and Manufacturing, F&I = Finance and Investment, T&T = Technology and Telecommunication, CS = Consulting Services.

5.1.2.3 The awareness of businesses about email spam, and the efforts to combat it, based on establishment year

As can be seen in Table 5.5, old businesses were more aware of email spam and anti-spam filters than were new businesses (98% vs 79.5%, p=0.004). The most common source of knowledge for all businesses about email spam and anti-spam filters was through the Internet and forums. New businesses learnt more about email spam and anti-spam filters than old businesses from other companies and organisations (28.2% vs 24%, p=0.004).

Table 5.5: Percentages of distribution of the awareness of businesses about email spam, anti-spam filters, and efforts to combat it based on establishment year

Question	Establishment Year		P*
	Before 1994 (old) n=50	1994 and later (new) n=39	
Aware of email spam and anti-spam filters (%YES)	98	79.5	**0.004**
Knowledge source about email spam			
ISPs	28	17.9	0.268
Internet and forums	80	66.7	0.154

Question	Establishment Year		P*
	Before 1994 (old) n=50	1994 and later (new) n=39	
Broadcast media, e.g. TV	32	15.4	0.071
Government	6	10.3	0.459
Other companies and organisations	24	28.2	**0.004**
Aware of government efforts to combat spam (%YES)	24	25.6	0.859
Aware of ISPs' efforts to combat spam (%YES)	22	23.1	0.904
Educate employees and customers about email spam and anti-spam filters (%YES)	44	30.8	0.202

*P values are based on chi-square test between businesses based on the establishment year; P values <0.05 were considered statistically significant.

5.1.3 The Nature of Email Spam as Perceived by Businesses

Table 5.6 shows that most businesses (94.6%, 95%CI: 88.5%-97.9%) received email spam, and most of the email spam received by businesses was in English (66.1%, 95%CI: 61.1%-71.1%), followed by Arabic (20.5%, 95%CI: 16.6%-24.5%).

Table 5.6: Percentages of distribution of the number of businesses received email spam, and the languages of email spam they received

Question	Frequency	Percentage (%)	95% CI
Receive email spam			
Yes	87	94.6	88.5-97.9
No	5	5.4	2.1-11.5
Languages of email spam			
English	85	66.1	61.1-71.1
Arabic	72	20.5	16.6-24.5
Unrecognised languages	31	8	5.2-10.9
Other languages	26	5.1	3-7.4

There were significant differences, as summarised in Table 5.7, between types of Arabic and English email spam. Religious and political emails (7% vs 3.6%, p=0.003), and emails related to forums (31.8% vs 5.9%, p<0.001) were more common in Arabic than English, whereas in the English language there were more pornographic emails (15.3% vs 5.6%, p<0.001), products and services (18.3% vs 10.1%, p=0.003), and phishing and fraud emails (25.7% vs 6%, p<0.001).

Table 5.7: Percentages of types of Arabic and English email spam received by businesses

Types of Email Spam	Arabic (%)	English (%)	P*
Business	35.4	30.9	0.155
Religious and political	7	3.6	**0.003**
Pornographic	5.6	15.3	**<0.001**
Forums	31.8	5.9	**<0.001**
Products and services	10.1	18.3	**0.003**
Phishing and fraud	6	25.7	**<0.001**
Other	2.6	0.4	0.083

*P values are based on paired-samples t-test between types of Arabic and English email spam; P values <0.05 were considered statistically significant.

Table 5.8 summarises the percentages of businesses that received email spam based on size, sector and establishment year. Medium and large businesses received more email spam than small businesses did (100% vs 82.1%, p=0.002); the finance and investment, technology and telecommunication, and consultation sectors received more email spam than “other” sectors (100%, p=0.014); and more old businesses received email spam than new businesses (100% vs 87.2%, p=0.009).

Table 5.8: Percentages of businesses that received email spam, based on size, sector and establishment year

Business Classification	Receive email spam (%YES)	P*
Business size		
Small	82.1	**0.002**
Medium	100	
Large	100	
Business sector		
P&M	97.6	**0.014**
F&I	100	
T&T	100	
CS	100	
Other	77.8	
Establishment year		
Before 1994 (old)	100	**0.009**
1994 and later (new)	87.2	

*P values are based on chi-square test between businesses based on size, sector and establishment year; P values <0.05 were considered statistically significant.

Abbreviations: P&M = Production and Manufacturing, F&I = Finance and Investment, T&T = Technology and Telecommunication, CS = Consulting Services.

5.1.3.1 The nature of email spam as perceived by businesses, based on size

As shown in Table 5.9, small businesses reported receiving more Arabic email spam than medium and large businesses (26.7%, p=0.046). There were no significant differences between small, medium and large businesses in the types of Arabic spam received, although there were significant differences in the types of English spam. The percentages showed that small businesses received more English religious and political emails than medium and large businesses (6.5%, p=0.014).

Table 5.9: Percentages of languages and types of Arabic and English email spam received by businesses in different sizes

	Number of Employees Business Size			P*
Question	**1-49 Small n=19**	**50-249 Medium n=36**	**250+ Large n=37**	
Languages of email spam				
English	60.1	73	62.5	0.078
Arabic	26.7	14.8	22.2	**0.046**
Unrecognised languages	6.1	8.5	8.9	0.733
Other Languages	6.9	3.5	5.7	0.459
Types of Arabic email spam				
Business	38.3	37.8	29.4	0.359
Religious and political	7.2	4	9.6	0.216
Pornographic	3.1	3.9	8.9	0.239
Forums	31.1	35	32.3	0.900
Products and services	9.5	9.9	10.1	0.989
Phishing and fraud	9.5	3.4	5.7	0.341
Other	1	5.9	0	0.080
Types of English email spam				
Business	29.5	29.1	38.3	0.171
Religious and political	6.5	1.2	2.8	**0.014**
Pornographic	18.7	12	14.7	0.223
Forums	8.5	5.6	5.7	0.665
Products and services	14	22.8	15	0.289
Phishing and fraud	22.8	27.8	23.9	0.733
Other	0	1.1	0	0.130

*P values are based on ANOVA test between businesses in different sizes; P values <0.05 were considered statistically significant.

5.1.3.2 The nature of email spam as perceived by businesses, based on sector

Table 5.10 shows significant differences in the types of Arabic and English email spam received by businesses in the different sectors. The finance and investment sectors received higher percentages of Arabic emails related to forums (63.3%, p=0.016), and English language emails related to phishing and fraud (51.2%, p=0.035), than other sectors.

Table 5.10: Percentages of languages and types of Arabic and English email spam received by businesses in different sectors

Question	Business Sector					P*
	P&M n=42	F&I n=8	T&T n=15	CS n=9	Other n=18	
Languages of email spam						
English	67.1	57.5	74.3	50	9.6	0.106
Arabic	21.9	15	15	27.8	21	0.458
Unrecognised languages	6.2	20	6	10	7.5	0.095
Other Languages	4.5	7.5	4.7	12.2	1.8	0.179
Types of Arabic email spam						
Business	39.1	26.7	45	30.7	20	0.071
Religious and political	4.8	10	8.7	12.1	6.2	0.502
Pornographic	7.7	0	6.4	2.8	1.9	0.492
Forums	27.9	63.3	19	39.2	42.9	**0.016**
Products and services	10	0	8.3	13.6	13.5	0.295
Phishing and fraud	5.4	0	12.5	1.4	6.2	0.355
Other	2	0	0	0	9.2	0.148
Types of English email spam						
Business	35.7	33.7	25.7	28.7	31.5	0.595
Religious and political	2.4	1.2	3	9.4	2.7	0.084
Pornographic	13	8.7	19.1	18.7	15	0.383
Forums	5.4	0	9.9	8.7	7.7	0.415
Products and services	20.4	5	24.3	5.6	18	0.187
Phishing and fraud	22	51.2	18	27.5	25.8	**0.035**
Other	0.7	0	0	1.2	0	0.698

*P values are based on ANOVA test between businesses in different sectors; P values <0.05 were considered statistically significant.

Abbreviations: P&M = Production and Manufacturing, F&I = Finance and Investment, T&T = Technology and Telecommunication, CS = Consulting Services.

5.1.3.3 The nature of email spam as perceived by businesses, based on establishment year

Table 5.11 reveals that there were no significant differences between old and new businesses in the languages and types of Arabic and English email spam that old and new businesses received.

Table 5.11: Percentages of languages and types of Arabic and English email spam received by old and new businesses

Question	Establishment Year		P*
	Before 1994 (old) n=50	1994 and later (new) n=39	
Languages of email spam			
English	67.6	61.8	0.259
Arabic	18.6	24.5	0.151
Unrecognised languages	8.7	7.5	0.691

Question	Establishment Year		P*
	Before 1994 (old) n=50	1994 and later (new) n=39	
Other Languages	4.8	6.1	0.564
Types of Arabic email spam			
Business	36.4	32.5	0.515
Religious and political	7.3	6.4	0.749
Pornographic	6.2	3.3	0.334
Forums	30.1	38.5	0.248
Products and services	9.5	10	0.881
Phishing and fraud	6.8	4.6	0.528
Other	1.2	4.6	0.241
Types of English email spam			
Business	32.5	29.8	0.561
Religious and political	2.7	3.7	0.536
Pornographic	14.8	14	0.796
Forums	7.1	3.9	0.216
Products and services	20.6	12.6	0.091
Phishing and fraud	21.8	29.4	0.185
Other	0.4	0.6	0.751

*P values are based on independent-samples t-test between businesses based on the establishment year; P values <0.05 were considered statistically significant.

5.1.4 How Businesses Deal with Email Spam

The ways in which businesses dealt with email spam are presented in Table 5.12. More than half of businesses (58.7%, 95%CI: 48.5%-68.5%) had businesses units or teams to manage network security. The most common responsibility of these units or teams (48.6%, 95%CI: 32.7%-64.7%), as reported by the businesses, was setting up and updating Internet security software and hardware. The least common means of dealing with spam (11.4%, 95%CI: 4%-25%) was to report security attacks to CITC.

About one-fifth of businesses (18.5%, 95%CI: 11.6%-27.3%) had employees with a specific responsibility to combat email spam, whose most common task was to apply and update anti-spam filters (73.3%, 95%CI: 48.3%-90.3%). Most businesses (80.4%, 95%CI: 71.5%-87.5%) used anti-spam filters to block email spam.

Table 5.12: Percentages of distribution of ways businesses deal with email spam

Question	Frequency	Percentage (%)	95% CI
Having business unit or team to manage network security			
Yes	54	58.7	48.5-68.4
No	38	41.3	31.6-51.5

Question	Frequency	Percentage (%)	95% CI
Responsibilities of business units or teams regarding network security			
Setting up and updating Internet security software and hardware	17	48.6	32.7-64.7
Reporting security attacks to CITC	4	11.4	4-25
Designing security policies for businesses	9	25.7	13.6-41.7
Providing technical support for users regarding security issues	5	14.3	5.7-28.5
Having specific employees to combat email spam			
Yes	17	18.5	11.6-27.3
No	75	81.5	72.7-88.4
Tasks of employees regarding email spam			
Applying and updating anti-spam filters	11	73.3	48.3-90.3
Reporting emails spam to CITC	2	13.3	2.9-36.3
Adding emails spam into blacklists	2	13.3	2.9-36.3
Using anti-spam filters to block email spam			
Yes	74	80.4	71.5-87.5
No	18	19.6	12.5-28.5

5.1.4.1 How businesses dealt with email spam, based on business size

As shown in Table 5.13, there were significant differences between the ways in which businesses dealt with email spam. More large businesses than small and medium ones created business units or teams to manage network security (86.7%, p=0.001). Large businesses also used anti-spam filters more than small and medium businesses (90%, p=0.032).

Table 5.13: Percentages of dealing of businesses with email spam, based on size

	Number of Employees Business Size			P*
How business deal with email spam	**1-49 Small n=28**	**50-249 Medium n=34**	**250+ Large n=30**	
Having business unit or team to manage network security (%YES)	46.4	44.1	86.7	**0.001**
Having specific employees to combat email spam (%YES)	14.3	14.7	26.7	0.371
Using anti-spam filters to block email spam (%YES)	64.3	85.3	90	**0.032**

*P values are based on chi-square test between businesses in different sizes; P values <0.05 were considered statistically significant.

5.1.4.2 How businesses dealt with email spam, based on business sector

As seen in Table 5.14, there were more businesses with business units or teams to manage network security in the finance and investment sector than in other sectors (100%, p=0.031). The finance and investment sectors used anti-spam filters to block email spam more than other sectors did (100%, p=0.017).

Table 5.14: Percentages of dealing of businesses with email spam, based on sector

How business deal with email spam	Business Sector P&M n=42	 F&I n=8	 T&T n=15	 CS n=9	 Other n=18	P*
Having business unit or team to manage network security (%YES)	57.1	100	73.3	44.4	38.9	**0.031**
Having specific employees to combat email spam (%YES)	21.4	25	6.7	22.2	16.7	0.739
Using anti-spam filters to block email spam (%YES)	88.1	100	86.7	66.7	55.6	**0.017**

*P values are based on Fisher's exact test between businesses in different sectors; P values <0.05 were considered statistically significant.

Abbreviations: P&M = Production and Manufacturing, F&I = Finance and Investment, T&T = Technology and Telecommunication, CS = Consulting Services.

5.1.4.3 How businesses deal with email spam, based on establishment year

Table 5.15 reveals that more old businesses established business units or teams to manage network security than new businesses (70% vs 41%, p=0.006). The results have shown no significant differences between old and new businesses in the number of special employees they assign to combat email spam, or in using anti-spam filters to block it.

Table 5.15: Percentages of dealing of old and new businesses with email spam

Dealing businesses with email spam	Establishment Year Before 1994 (old) n=50	 1994 and later (new) n=39	P*
Having business unit or team to manage network security (%YES)	70	41	**0.006**
Having specific employees to combat email spam (%YES)	22	12.8	0.263
Using anti-spam filters to block email spam (%YES)	84	74.4	0.261

*P values are based on Chi-square test between businesses based on the establishment year; P values <0.05 were considered statistically significant.

5.1.5 The Effects of Email Spam on the Performance of Businesses

The effects of email spam are summarised in Table 5.16. The biggest effect on businesses' performance was reduction in the efficiency of organisation's email server due to the large volume of spam (82.6%, 95%CI: 73.9%-89.3%), while the smallest effect to the expense of buying or updating anti-spam filters (54.3%, 95%CI: 44.2%-64.3%).

Table 5.16: Percentages for the effects of email spam on the performance of businesses

Effects of email spam	Frequency	Percentage (%)	95% CI
Losing time and reducing productivity	66	71.7	62-80.2
Spending money to buy or update anti-spam filters	50	54.3	44.2-64.3
Reducing the efficiency of organisation's email server	76	82.6	73.9-89.3
Computer infection by malicious programs, e.g. viruses	59	64.1	54-73.4

5.1.5.1 The effects of email spam on the performance of businesses, based on business size

As summarised in Table 5.17, large businesses spent more money to buy or update anti-spam filters used to block email spam than small and medium businesses (70%, p=0.014)

Table 5.17: Percentages for the effects of email spam on the performance of small, medium and large businesses

Effects of email spam	Number of Employees Business Size			P*
	1-49 Small n=28	50-249 Medium n=34	250+ Large n=30	
Losing time and reducing productivity	71.4	73.5	70	0.951
Spending money to buy or update anti-spam filters	32.1	47.1	70	**0.014**
Reducing the efficiency of organisation's email server	78.6	82.4	86.7	0.718
Computer infection by malicious programs, e.g. viruses	57.1	73.5	60	0.346

*P values are based on chi-square test between businesses in different sizes; P values <0.05 were considered statistically significant.

5.1.5.2 The effects of email spam on the performance of businesses, based on business sector

Table 5.18 reveals that the percentage of businesses that spent money to buy or update anti-spam filters was greater in the finance and investment sectors than the other sectors (100%, p=0.04).

Table 5.18: Percentages for the effects of email spam on the performance of businesses in different sectors

Effects of email spam	**Business Sector**					**P***
	P&M n=42	**F&I n=8**	**T&T n=15**	**CS n=9**	**Other n=18**	
Losing time and reducing productivity	59.5	87.5	86.7	88.9	72.2	0.134
Spending money to buy or update anti-spam filters	50	100	33.3	44.4	44.4	**0.04**
Reducing the efficiency of organisation's email server	83.3	100	80	88.9	72.2	0.498
Computer infection by malicious programs, e.g. viruses	71.4	25	66.7	77.8	55.6	0.106

*P values are based on Fisher's exact test between businesses in different sectors; P values <0.05 were considered statistically significant.

Abbreviations: P&M = Production and Manufacturing, F&I = Finance and Investment, T&T = Technology and Telecommunication, and CS = Consulting Services.

5.1.5.3 The effects of email spam on the performance of businesses, based on establishment year of business

As seen in Table 5.19, there were no significant differences between old and new businesses in the effects of email spam on their performance.

Table 5.19: Percentages for the effects of email spam on the performance of old and new businesses

Effects of email spam	**Establishment Year**		**P***
	Before 1994 (Old) n=50	**1994 and later (New) n=39**	
Losing time and reducing productivity	76	66.7	0.331
Spending money to buy or update anti-spam filters	54	46.2	0.463
Reducing the efficiency of organisation's email server	86	79.5	0.415
Computer infection by malicious programs, e.g. viruses	68	59	0.379

*P values are based on Chi-square test between businesses based on the establishment year; P values <0.05 were considered statistically significant.

5.2 Discussion

This section discusses results of the survey questions about the nature of email spam, the awareness of businesses about it, anti-spam filters, and the efforts to combat it. It also discusses the results of the questions about the effects of email spam on their performances, and the way they deal with spam.

5.2.1 The Awareness of Businesses about Email Spam, Anti-spam Filters, and the Efforts to Combat it

The results revealed that most businesses (90.2%, 95%CI: 82.9%-95%) knew about email spam and anti-spam filters, and about a quarter of businesses were aware of the government's and ISPs' efforts in Saudi Arabia to combat it. When comparing these results with other countries, such as Australia, it was found that all businesses in Australia were aware of email spam, its impacts, and the government legislation to combat it (ACMA 2011). The most common source of knowledge of businesses about email spam and anti-spam filters was self-education through the Internet and forums. Businesses thought that there was a deficiency in the efforts provided by government to raise their awareness about email spam and methods of combatting it. Previous research has revealed that the EU, Denmark and India conducted programs to raise awareness of online threats such as spam. The member states of the EU has conducted campaigns to make users aware of spam and the appropriate ways to deal with it (Pfleeger & Bloom 2005). The Confederation of Danish Industries and the Danish Consumer Ombudsman office has provided awareness programs for private individuals and companies explaining how to combat spam (Frost & Udsen 2006). Jidiga and Sammulal (2013) stated that:

> In India, the government organizations like MCIT (Ministry of Communication and Information technology) setup separate divisions to conduct a security awareness programs to the people, employees, students about spam.

It can be concluded that collaboration between the Saudi Government, broadcast media, and ISPs in the education of businesses about email spam and methods of combating it could increase the awareness of businesses about email spam.

In this study, most businesses in Saudi Arabia defined email spam in line with international definitions of email spam as UCE (Boykin & Roychowdhury 2004; Cheng 2004; Sakkis et al. 2003). However, this result differs from the definitions by researchers found in other studies such as Ahmed and Oppenheim (2006), Adam (2007), Polanski (2008), Fogel and Raghupathi (2013) and Arutyunov (2013) which considered UCE not as email spam, but as a quick and easy tool to advertise products and services to customers.

The results showed that about two-fifths of businesses (40%, 95%CI: 29.6%-49.3%) educated their employees and customers about email spam and methods of combating it. Previous studies have indicated that the education and users' awareness about spam are important tools in combatting it. D'Ambra (2007) stated that "education needs to play a larger role in the fight against spam as computer users either lack the understanding or are not interested in computer security". Jidiga and Sammulal (2013) reported that private organisations in India were helping the government to conduct the awareness programs. The results suggest, therefore, that businesses should focus on education of their customers and employees about spam and anti-spam filters.

5.2.2 The Nature of Email Spam Received by Businesses

This section discusses the survey results about the nature of email spam received by businesses, such as the volume of spam, its languages, and types of Arabic and English email spam.

5.2.2.1 The volume of email spam received by businesses

Most Saudi businesses reported receiving email spam, with an average 4,400 spam received per week. A previous study conducted on 500 US and Finnish companies found that an average of 1,987,000 spam emails were received by companies each week (Siponen & Stucke 2006). Another study revealed that UK companies received an average of 101,500 email spam per week (Computer Fraud and security 2004). This indicated that American, Finnish and British companies, on average, received more email spam per week than Saudi companies. This could be because US, Finland and UK companies were more popular (Milletary & Center 2005) and have more customers and employees than Saudi companies (Ridzuan, Potdar & Talevski 2010), and so provide better targets than Saudi companies.

5.2.2.2 The languages of email spam received by businesses

In this study, most email spam received by businesses in Saudi Arabia was written in English. This could be because English is the most used language in the world (Kirkpatrick 2007) and is the most popular language for email spam in English-speaking and non-English–speaking countries (Ermakova 2010). This agrees with the finding of Shrivastava and Bindu (2012) that the most popular language for email spam around the world was English. Moreover, Pfleeger and Bloom (2005) reported that most of the email spam received in the EU was written in English, even though the EU includes about 12 different official languages. However, this result conflicts with other studies, such as a study by Symantec (2010) that showed that the highest percentage of email spam received in Brazil was in Portuguese (33%), while English (25.6%) was the second most-used language for spam in that country

In this study, Arabic was the second most popular language used in email spam received by businesses. This agrees with the finding of a study by El-Halees (2009), which revealed that most email spam received in the Arab countries was written in Arabic, English, or mixed Arabic and English, possibly because Arabic is the official language of Saudi society (Chejne 2009). This could encourage Arabic spammers to write email spam in Arabic to make them more understandable for recipients (Zaidan et al. 2011) and reaping more financial benefits for spammers (Cook et al. 2006).

5.2.2.3 The types of email spam received by businesses

In this study, the most common type of Arabic email spam related to forums; while the most common type of English email spam was phishing and fraud. The types of Arabic and English email spam received in Saudi Arabia were different from that in other languages received in different countries. This might be explained by the spammers' culture, religion and country of origin (Abdoh, Musa & Salman 2009). Yamakawa and Yoshiura (2010) revealed that the most common types of English email spam received in Japan were related to commercial advertising, while the most common types of Japanese email spam were related to sexuality. Lev and Goldin (2006) described different types of spam for four countries: Russia, China, Germany and Korea. Email spam in Russia targets food, accessories, education and construction. In China, the most common type of email spam was the sale of fake invoices that are designed to reduce the tax burdens of different businesses, and anti-government spam. Subjects of spam in Germany included racist and white supremacist spam. In Korea, typical types of spam included financial or mortgage-related emails (Lev & Goldin 2006). Ermakova (2010) reported that the most common type of English email spam in non-English-speaking countries was for English courses.

The results showed that there were significant differences between Arabic and English email spam received by businesses. More emails related to forums and political and religious emails were received in Arabic than in English. It might be that forums are a favourite way for Saudis to communicate with each other in Saudi Arabia (Al-Saggaf 2004) and are used for educational, financial (Stone-Gross et al. 2011), religious and political purposes (Grossman 2004; Martinkova 2008; Sweet 2003).

In this study, more pornographic emails received by businesses were in English than in Arabic. In Saudi Arabia, the religion is Islam, and Islam prohibits pornography (Al-A'ali 2007), but pornography is also forbidden by Arabic culture (Abdoh, Musa & Salman 2009). Both could serve to reduce the number of pornographic emails received in Arabic. This finding agrees with a study conducted in the USA by Hind (2003), which revealed that the most frequent type of email spam received by American users was pornographic.

The results of this study revealed that percentage of products and services emails was higher in English than Arabic. This is similar to the results of a study by Yamakawa and Yoshiura (2010), which revealed more commercial advertising in email spam in English than in Japanese. A possible reason could be that businesses' email addresses are added to the mailing lists of other sectors that the businesses dealt with previously, or that they contracted with them to achieve financial benefits (i.e. make a commercial partnership between Saudi businesses and foreign sectors in English-speaking countries) (Ramady & Sohail 2010). This might lead to the harvesting of email addresses of Saudi businesses by English spammers and a larger number of products and services advertisements in English than in Arabic.

There were more phishing and fraud emails in English than in Arabic in the results of this survey. It is possible that most businesses designed online payment portals with

English interfaces (AlGhamdi & Drew 2012), which encourages spammers to design fake portals comparable to the original ones, and attach them in email spam. This could result in more phishing and fraud emails in English than Arabic.

5.2.3 How Businesses Deal with Email Spam

The results showed that over half of the businesses (58.7%, 95%CI: 48.5%-68.5%) in surveyed had a business unit or team to manage network security. Previous research has revealed that establishing business management or teams to manage information security is important to protect the organisation's network from potential security attacks. According to Vroom and Von Solms (2004):

> … with the introduction of information technology and the resulting security challenges that organizations face daily, it has become essential to ensure the security of the organization's information and other valuable assets.

The results of this study showed that the most common responsibility of employees in businesses units charged with fighting spam was to set and update Internet security software and hardware. Several tasks have been undertaken by information security management or units in businesses to combat security attacks, and one of these tasks was Internet security software management. Von Solms (2005) stated that examples of the information security operational management activities in the organisations were installing and updating Internet security software, and renewing software licences. Johnson and Koch (2006) reported that about 12% of the IT department budgets of American organisations were spent on network security.

Another task specified for combatting security attacks in businesses, in the results of this study, was to design security policies (Sorkin 2001). Pfleeger and Bloom (2005) reported that some companies developed security policies, and an example of these companies was the ePrivace Group, which developed security policies such as the Trusted Email Open Standard (TEOS). This study found that a bit over quarter of Saudi businesses (25.7%, 95%CI: 13.6%-41.7%) had developed security policies for their organisations. A study conducted by Sunner (2005) on 182 IT security professionals in the UK revealed that 51% had formal policies regarding security attacks. This finding suggests that Saudi businesses should have a focus on designing security policies by to combat different security attacks (Sorkin 2001). However, employees' negligence or ignorance of the security policies in the organisations can also result in the organisations being more targeted for security breaches. Vroom and Von Solms (2004) found that 48% of the security breaches in organisations were due to employees' ignorance of the security policies. In this study, about two-thirds of businesses (64.1%, 95%CI: 54%-73.4%) have been affected by malicious programs such as trojans and viruses. This is a high percentage that indicates the need for further efforts by Saudi businesses to protect their networks from potential security attacks.

About one-fifth of businesses (18.5%, 95%CI: 11.6%-27.3%) had employees with a specific responsibility to combat email spam. Previous studies have demonstrated that employing qualified staff can help in fixing problems related to email spam. Arutyunov (2013) reported that an American company allocated one full-time IT person to fix spam-related problems for every 690 employees. Ridzuan, Potdar and Talevski (2010) stated that companies need to spend money to buy the necessary anti-spam filters, recruit employees to deal with spam problems, and provide the required training for those employees to improve their understanding of email spam. It is clear that either the recruitment of qualified employees with expertise in the field of network security in general and spam in particular, or outsourcing the maintenance and management of the anti-spam filters, would help in reducing the effects of email spam on the performance of businesses (Frost & Udsen 2006).

Most Saudi businesses (80.4%, 95%CI: 71.5%-87.5%) used anti-spam filters to combat email spam. A technology consultant at one of the companies that sell email security products recommended that network managers use an email firewall with anti-spam and anti-virus software to monitor and clean machines, update the software regularly, and implement intrusion detection software to prevent spammers' activities from taking place within the firewall (Everett 2004). The use of effective anti-spam filters by businesses could reduce the volume of email spam and save millions of dollars (Osterman Research Inc. 2008). Osterman Research Inc. (2008) has indicated that the cost of email spam for a company with 1,200 employees could be $2.4 million, but by using anti-spam filters, they could save $1.2 million. Renewing the licence or updating anti-spam filters can cost businesses a lot of money, especially small companies, but this cost is still lower than the cost to productivity caused by spam (Ridzuan, Potdar & Talevski 2010). In this study, more than half of Saudi businesses (54.3%, 95%CI: 44.2%-64.3%) spent money to apply or update anti-spam filters in the company.

5.2.4 The Effects of Email Spam on the Performance of Businesses

The results have shown that the greatest effect of email spam on the performance of Saudi businesses was reduction in the efficiency of the organisation's email server due to receiving a large volume of email spam (82.6%, 95%CI: 73.9%-89.3%). The huge volume of email spam could be a burden on the email server (Mo et al. 2006), as spam is received and stored in email inboxes, and some spam might have attachments. Downloading these attachments can result in consumption of the organisation's bandwidth (Cook et al. 2006). This suggests that developing anti-spam filters to block email spam before it arrives companies' networks would be more efficient. Another suggestion is that "companies need to deploy security measures on network servers to prevent spammers from hacking the server" (Ridzuan, Potdar & Talevski 2010). These suggestions can be implemented to reduce the effects of email spam on the efficiency of email server systems in Saudi businesses.

The second effect of email spam on the performance of Saudi businesses was loss of time and reduced productivity (71.7%, 95%CI: 62%-80.2%). Employees spent time isolating spam emails from legitimate emails and fixing problems caused by spam (Bujang & Hussin 2013; Pérez-Díaz et al. 2012). Employees also waste time checking spam folders to avoid losing important emails that are misclassified by anti-spam filters (Ridzuan, Potdar & Talevski 2010). A study by Siponen and Stucke (2006) indicated that employees in the USA and Finland wasted an average of 13 minutes daily in fixing email spam problems. Another study conducted in Germany indicated that the time spent to identify and delete email spam was about1,200 minutes per employee per year (Caliendo et al. 2008). When employees waste time dealing with email spam it can cost companies in productivity. The Singapore Infocomm Development Authority (IDA) indicated that the total cost of spam for consumers was about S$23 million in lost productivity each year (Leng 2006). The GDP loss due to processing email spam in Japan was about 500 billion yen a year

(Takemura & Ebara 2008). It can be concluded that email spam can reduce companies' productivity and in turn affect the economy of countries. This suggests that Saudi Arabia needs to focus further on mitigating email spam.

5.2.5 Demographic Information of Businesses and Email Spam Characteristics

This section discusses email spam characteristics based on different factors, such as business size, business sector and establishment year of the business.

5.2.5.1 Discussion of results, based on businesses size

Statistically significant differences have been found in the awareness of small, medium and large businesses about email spam and anti-spam filters. The results have shown that large businesses had more knowledge about email spam and anti-spam filters than small and medium businesses. The reason may be due to the lack of knowledge of small and medium businesses about the Internet and technology. Burgess (2002) found that the awareness and knowledge of small and medium enterprises (SMEs) about ICT was low. This finding is supported by studies by Dojkovski, Lichtenstein and Warren (2007) and Margariti et al. (2007), which revealed that small and medium businesses were less aware of security issues than large businesses. This suggests that additional efforts are required by relevant agencies in Saudi Arabia to increase the awareness of small and medium businesses about security issues such as email spam. According to Dojkovski, Lichtenstein and Warren (2007), "SMEs need external support in order to develop the necessary proactivity to promote and support information security culture internally".

Large businesses provided more education and awareness for their employees and customers about email spam and methods of combating it, than small and medium businesses. This could be because small and medium businesses lack the budget and the specialised trainers to conduct education and awareness programs. This is supported by Furnell, Gennatou and Dowland (2000), who stated that small and medium sized businesses provide inadequate awareness, training and education programs regarding information security issues, and this might be because of the lack of funds, time and specialised knowledge and instructors to coordinate these programs.

The results indicated that large and medium-sized businesses received more email spam than small businesses. This finding is inconsistent with a previous study by Siponen and Stucke (2006), which revealed that small businesses were targeted more by email spam than were medium and large businesses. This might be because large and medium businesses were more popular (Milletary & Center 2005), and the average number of employees and customers who deal with them was greater than the average number in small businesses (i.e. it is profitable for spammers to reach more recipients and this can be achieved through customers in large and medium businesses) (Ridzuan, Potdar & Talevski 2010). This could result in large and medium businesses being targeted more than small businesses by spammers.

The percentage of large businesses that had business units or team to manage network security, and which used anti-spam filters to block email spam, was greater than the percentages in small and medium businesses. Possibly larger businesses can afford to allocate more money to establish units for network security and allocate qualified employees to manage network security measures, which is less achievable for small and medium businesses (Johnson & Koch 2006; Yeniman Yildirim et al. 2011). However, the discrepancy might also be explained by small and medium businesses outsourcing the management of their networks and network security to technical companies or ISPs, or hiring employees or experts to deal with security issues such as spam (Ridzuan, Potdar & Talevski 2010).

5.2.5.2 Discussion of results, based on businesses sector

The results have shown that businesses working in the finance and investment, and technology and telecommunication sectors, were more aware of email spam and anti-spam filters than other sectors. There are several possible reasons for this. The finance and investment sectors deal with important customer information, such as credit card numbers, and they are required to keep customers' information confident and secure (Hwang, Chen & Lee 2007; Schwartz & Janger 2007). This requires greater understanding about security threats, such as email spam, and appropriate ways to combat them. Another reason could be that the finance and investment sector relies more than other sectors on information technology for their business operations, and are thus more aware of email spam and anti-spam filters than other sectors. According to Yeh and Chang (2007), "banking and finance appeared heavily reliant on Information Technology more than other businesses”.

One would expect that the technology and telecommunication sectors are necessarily more aware of email spam and anti-spam filters because of the nature of the field. A study conducted by Sathiyaseelan and Filmore (2011) found that most of the employees in the information technology organisations were aware about technologies and technical methods used in their organisations.

All finance and investment sectors (100%) provided awareness programs for their employees and customers about email spam and anti-spam filters, and this percentage was higher than in all other sectors. This might be because the finance and investment sectors are targeted more often by attackers than other sectors (Ramanathan & Wechsler 2012), increasing the need for awareness and education to avoid the effects of security attacks caused by spam. Conducting the security awareness programs for employees and customers who work in or deal with this sector is necessary to guarantee the safety of confidential information. Kumar (2005) argued for awareness and education programs for employees and customers in the financial sectors. Previous research has indicated the need for financial institutions to provide more customer awareness programs about information security. Mahdi, Rezaul and Rahman (2010) found that approximately half of the participants asserted the need for the financial and investment sectors in the UK to increase their efforts in customer and employee awareness of security issues.

A significant difference was found in the results of this study between types of email spam received in different sectors. The finance and investment sector received a higher percentage of phishing and fraud emails than the other sectors. One possible reason might be that attackers target the security vulnerabilities of the finance and investment sector systems and infect employees’ computers in order to obtain important information. This is supported by a report that “If an employee’s desktop from one of the specific banks becomes infected, the virus recognizes this and attempts to steal data to compromise the bank” ('New Bugbear targets banks' 2003). This could make the finance and investment sector a favoured target of criminals. Previous studies have also indicated that this sector may be more targeted by attackers and phishers than other sectors. Ramanathan and Wechsler (2012) found that banking services are one of the sectors targeted by attackers. Another study by

Hind (2003) indicated that over 1,200 web addresses of banks and financial institutions were inserted to phishers' and attackers' lists. Kumar (2005) also suggested that "Banks should take steps to make their customers and employees aware of basic security practices". It is clear, then, that the finance and investment sector in Saudi Arabia needs to increase its employees' and customers' awareness of about the security issues caused by email spam.

The results of this study have shown that the finance and investment sector in Saudi Arabia, more than the other sectors, has established business units and created teams to manage network security of the organisation, and used anti-spam filters to block email spam. Its security efforts are better than those of the other sectors. Yeh and Chang (2007) reached a similar conclusion, finding that banking and finance businesses in Taiwan did more to combat security attacks than other businesses. This may be because of the importance of the finance and investment sectors' customer information, which requires additional efforts to protect them (Hwang, Chen & Lee 2007). A study conducted by Mahdi, Rezaul and Rahman (2010) revealed that about half of the participants in their UK study said that financial institutions are required to protect their information from potential security attacks. However, technical efforts can impact on the budget of this sector (Ridzuan, Potdar & Talevski 2010). The results of this study found that the finance and investment sector spends a lot of money on network security compared to other sectors.

5.2.5.3 Discussion of results, based on establishment year of businesses

The results showed that significant differences between old and new businesses in terms of their awareness of email spam and anti-spam filters, and also in their dealing with it. Old businesses knew more about email spam and anti-spam filters than new businesses. One explanation for this could be that old businesses have probably expanded geographically, which could require connecting and linking branches in different places with the head office by the Internet, and involves providing the necessary security measures (Mazidah & Burairah 2014). This is supported by the results of the researcher's study, which found that more old businesses than new businesses created business units or teams to manage network security. This could increase their knowledge about information technology. It may also be that more old businesses than new businesses use e-commerce to offer products and sales via the Internet (Passari, Radmand & Batoie 2013; Turban et al. 2009). This may help old businesses to gain a better understanding of the Internet and its application, such as anti-spam filters. There is clearly a need, then, for new businesses to increase their awareness of email spam and effective methods to combat it.

5.3 Conclusions

This chapter described, analysed and discussed the survey results regarding the awareness of Saudi businesses about email spam, anti-spam filters and the efforts to combat it in Saudi Arabia. It also presented, analysed and discussed the results for the nature of email spam as perceived by Saudi businesses, its effects on their performance, and how they dealt with it, based on some factors that have been used in some previous studies, including business size, business sector and establishment year of the business. The main conclusions of the business survey addressing the research questions are presented in the following paragraphs.

Most Saudi businesses were aware of email spam, as evidenced by defining it as UCE and UBE, which agreed with most of the international definitions. The most common source of their knowledge was through the Internet. The efforts of the government and relevant agencies in Saudi Arabia to combat it and raise awareness were limited. Although few Saudi businesses were unaware of email spam and anti-spam filters, some businesses provided awareness programs for their employees and customers. This suggests the need for additional efforts by the government and relevant agencies to inform businesses about email spam and related issues.

Not surprisingly, the highest percentage of email spam received by Saudi businesses was written in English, with Arabic the second most frequent language. However, there were many differences between the Arabic and English email spam. There were more political and religious emails and emails related to forums in Arabic than in English, but more pornography, products and services, and phishing and fraud emails in English than in Arabic. Although some of the dangerous types of email spam (e.g. phishing and fraud) are more limited in Arabic, there were attempts from Arab spammers to develop it and spread it in Saudi Arabia for financial gain. This suggests the need for further efforts and preparation by government and industry to combat this type of email spam.

Saudi businesses had made some technical attempts to deal with security issues such as email spam. These efforts included establishing business units or teams to manage network security, allocating specific employees to deal with spam problems, and using anti-spam filters. However, some businesses had not made any effort to use technical methods or tools to limit the effect of security attacks on their work and productivity, which implies that businesses should concentrate on technical measures and tools to protect their networks.

Email spam had some effects on the performance of Saudi businesses. The main effect was to reduce the efficiency of the organisation's email server due to the receipt of huge volumes of spam. This wastes employees' time in fixing ensuing problems, and can result in loss of productivity. Reduced productivity can also affect Saudi economic growth, as indicated in the large financial cost of email spam to countries such as the US (Cook et al. 2006) and Singapore (Leng 2006) for the same reason. Again, this raises the need for additional efforts by Saudi Arabian Government authorities to combat email spam. These efforts could be legal, technical and educational.

The results of this chapter described the experiences of Saudi businesses as email users and how they deal with it. If the email spam issue is investigated from point of view the sector (i.e. ISPs) that is responsible for providing the Internet service in Saudi Arabia, fruitful results can be expected. The next chapter will describe the survey results of the nature of email spam as perceived by Saudi ISPs, its effects, how the ISPs deal with it, and their evaluation of the effectiveness of anti-spam filters in detecting Arabic and English email spam.

Chapter 6: An Assessment of Email Spam among Saudi ISPs

This chapter presents the results of the survey of Saudi ISPs. It includes the awareness efforts about email spam provided by ISPs for their customers and employees, the nature of email spam that they blocked, and how the ISPs in Saudi Arabia dealt with it. Information is provided about the anti-spam filters used by ISPs to block email spam and their effectiveness in detecting Arabic and English spam as perceived by Saudi ISPs.

The chapter is divided into the following sections.

- Section 6.1: presents the results of the questionnaire for ISPs.
- Section 6.2: discusses the results of the questionnaire for ISPs.
- Section 6.3: describes the conclusions of this chapter.

6.1 Results

This section describes the results of the survey of Saudi ISPs. It includes responses covering the ISPs activities in increasing employee and customer awareness of SPAM, the nature of the SPAM that they block, and how the ISPs go about attempting to block SPAM. ISPs also reported which spam filtering systems they were using, and provided empirical feedback on the relative effectiveness of these systems against English and Arabic spam.

The statistical test, paired sample t-test, was used to analyse the data. This test was employed to compare the means between two related groups (two dependent variables). In this study, the paired sample t-test compared the means between Arabic and English in the types of email spam (Table 6.4), their sources or origins (Table 6.5), and the effectiveness of content-based and origin-based filters in detecting email spam for both languages (Table 6.8).

6.1.1 Participants' Demographic information

This section provides demographic information about Saudi ISPs. The information is presented in Table 6.1. Overall, 11 ISPs participated in this study, most of whom were from the central region (63.64%).

Most of the ISPs that participated in this study (45.5%) were classified as medium-sized (50-249 employees). Approximately 54.5% of the ISPs were new (established in 1994 and later), while 45.5% were classified as old (established before 1994)[7] ISPs.

Table 6.1: Distribution of ISPs in Saudi Arabia based on their demographic information

Demographic Information	Frequency	Percentage (%)
Region		
Eastern	2	18.20
Western	2	18.20
Central	7	63.64
ISP size		
Small (1-49 Employees)	3	27.3
Medium (50-249 Employees)	5	45.5
Large (250 Employees and more)	3	27.3
Establishment Year		
Before 1994 (old)	5	45.5
1994 till now (new)	6	54.5

6.1.2 Saudi ISPs Definition of Email Spam, their Awareness of Government Efforts to Combat it, and ISPs' Efforts to Educate Employees and Customers about it

Table 6.2 summarises the ISPs' definition of email spam, their awareness of it, and education efforts to inform their employees and customers about email spam and anti-spam filters

Most of the Saudi ISPs (62.5%, 95%CI[8]: 29.5%-88.1%) defined email spam as UCE. Saudi ISPs thought that the technical efforts conducted by CITC and KACST represented the main efforts of government to combat spam (63.6%, 95%CI: 34.8%-86.3%).

Approximately a quarter of ISPs (27.3%, 95%CI: 8.3%-56.5%) conducted

[7] 1994 was the year Saudi Arabia entered the Internet Service(CITC 2012). Some ISPs were established before 1994 and worked in different business activities before providing Internet services. This included engineering and technology consultation services, photocopiers and computer equipment sales, software and technology products sales and mobile services. For this reason, they are classified as old, while ISPs that were established in 1994 or later were classified as new.

[8] 95% Confidence Interval

workshops and training for employees and customers about email spam and anti-spam filters. These workshops and training sessions were conducted every 4-6, 7-9 and 10-12 months respectively (33.3%, 95%CI: 3.9%-82.3%). Approximately (54.5%, 95%CI: 27%-80%) of the ISPs provided awareness programs for employees and customers about email spam and methods of combatting it.

Table 6.2: Percentages of distribution of the efforts of Saudi ISPs to educate their employees and customers about email spam, and anti-spam filters

Question	Frequency	Percentage (%)	95% CI
Definition of email spam			
UBE	2	25	5.6-59.2
Email was sent from unknown senders and without recipients' permission to receive it	1	12.5	1.4-45.4
UCE	5	62.5	29.5-88.1
Government efforts to combat spam			
Technical efforts by CITC and KACST	7	63.6	34.8-86.3
Awareness efforts by CITC	1	9.1	1-35.3
Receiving ISPs' reports regarding spam issues	3	27.3	8.3-56.5
Conducting workshops and training for employees about email spam and anti-spam filters			
Yes	3	27.3	8.3-56.5
No	8	72.7	43.5-91.7
Period of conducting workshops and training about email spam			
Every 4-6 months	1	33.3	3.9-82.3
Every 7-9 months	1	33.3	3.9-82.3
Every 10-12 months	1	33.3	3.9-82.3
Awareness of customers of email spam and anti-spam filters			
Yes	6	54.5	27-80
No	5	45.5	20-73

6.1.3 The Nature of Email Spam As Perceived by ISPs

The results summarised in Table 6.3 show that all ISPs (100%, 95%CI: 80%-100%) reported blocking email spam, and the most popular language that they blocked was English (58.6%, 95%CI: 43.4%-73.8%). The second most common language of blocked email spam was Arabic (24.2%, 95%CI: 14.7%-33.7%).

Table 6.3: Percentages of ISPs that blocked email spam, and the languages used for the email spam that they blocked

Question	Frequency	Percentage (%)	95% CI
Blocking email spam			
Yes	11	100	80-100
No	0	0	0-20
Languages used for email spam			
English	11	58.6	43.4-73.8
Arabic	11	24.2	14.7-33.7
Unrecognised languages	5	6.8	0-13.6
Other languages	6	10.4	1.4-22.2

Table 6.4 showed a significant difference between the types of Arabic and English email spam blocked by Saudi ISPs. The percentage of pornographic email spam was higher in English than in Arabic (14% vs 4.4%, p=0.012).

Table 6.4: Percentages of types of Arabic and English email spam blocked by Saudi ISPs

Types of Email Spam	Arabic (%) n=11	English (%) n=11	P*
Business	40	20	0.685
	50	55	
	30	70	
	40	40	
	10	10	
	60	30	
	40	30	
	70	30	
	60	55	
	20	5	
	10	50	
Mean	**39**	**35.9**	
Religious and political	0	0	0.6
	10	0	
	30	10	
	0	0	
	0	0	
	10	10	
	10	20	
	0	0	
	0	1	
	0	5	
	0	0	
Mean	**5.4**	**4.2**	
Pornographic	0	20	**0.012**
	10	20	
	10	5	
	0	0	

Types of Email Spam	**Arabic (%)** **n=11**	**English (%)** **n=11**	**P***
	0	20	
	0	0	
	15	20	
	0	20	
	4	30	
	10	10	
	0	10	
Mean	**4.4**	**14**	
Forums	60	60	0.78
	5	0	
	10	5	
	0	20	
	60	50	
	0	30	
	10	20	
	0	0	
	12	0	
	20	30	
	80	20	
Mean	**23.4**	**21.4**	
Products and services	0	0	0.071
	15	15	
	20	10	
	30	0	
	30	15	
	30	0	
	20	10	
	30	50	
	20	10	
	40	30	
	10	10	
Mean	**22.3**	**13.6**	
Phishing and fraud	0	0	0.148
	10	10	
	0	0	
	0	0	
	0	5	
	0	30	
	5	0	
	0	0	
	4	4	
	10	20	
	0	10	
Mean	**2.6**	**7.2**	
Other	0	0	0.341
	0	0	
	0	0	
	30	40	
	0	0	

Types of Email Spam	**Arabic (%) n=11**	**English (%) n=11**	**P***
	0	0	
	0	0	
	0	0	
	0	0	
	0	0	
	0	0	
Mean	**2.7**	**3.6**	

*P values are based on paired-samples t-test between types of Arabic and English email spam; P values <0.05 were considered statistically significant.

Significant differences were found between sources of Arabic and English email spam (see Table 6.5). Most of the Arabic email spam was sent from Saudi Arabia, and the percentage of Arabic email spam that sent from Saudi Arabia was higher than the percentage of English email spam sent from Saudi Arabia (41.4% vs 16.4%, p=0.031). The highest percentage of English email spam was sent from non-Arabic countries, and the percentage of English email spam that originated from non-Arabic countries was greater than the percentage of Arabic email spam (52.7% vs 8.2%, p=0.002).

Table 6.5: Percentages of Arabic and English Email Spam from Various Sources Blocked by Saudi ISPs

Source (Origin) of Email Spam	**Arabic (%) n=11**	**English (%) n=11**	**P***
Saudi Arabia	41.4	16.4	**0.031**
Other Arabic countries	29.5	11.4	0.064
Non-Arabic countries	8.2	52.7	**0.002**
Unknown	20.9	19.5	0.914

*P values are based on paired-samples t-test between sources of Arabic and English email spam; P values <0.05 were considered statistically significant.

6.1.4 How Saudi ISPs Deal with Email Spam

The results for survey questions on how Saudi ISPs deal with email spam are shown in Table 6.6. Most Saudi ISPs (81.8%, 95%CI: 53%-96%) had business units or teams to manage network security, and the greatest responsibility of employees in these units (55.6%, 95%CI: 25.4%-82.7%) was setting up and updating Internet security software and hardware. About half of Saudi ISPs (54.5%, 95%CI: 27%-80%) did not have specific employees to combat email spam.

All ISPs (100%, 95%CI: 80%-100%) used anti-spam filters to block email spam. As well, all ISPs (100%, 95%CI: 80%-100%) used content-based filters, of which the most popular was Iron Port (45.5%, 95%CI: 20%-73%). Ten of the ISPs (90.9%, 95%CI: 64.7%-99%) used origin-based filters, of which the most popular were blacklists (100%, 95%CI: 78.3%-100%).

Table 6.6: Percentages of Saudi ISPs dealing with email spam

Question	Frequency	Percentage (%)	95% CI
Having business unit or team to manage network security			
Yes	9	81.8	53.3-96
No	2	18.2	4-46.7
Responsibilities of business units or teams regarding network security			
Setting up and updating Internet security software and hardware	5	55.6	25.4-82.7
Reporting security attacks to CITC	2	22.2	4.9-54.4
Providing technical support for users regarding security issues	2	22.2	4.9-54.4
Having specific employees to combat email spam			
Yes	5	45.5	20-73
No	6	54.5	27-80
Using anti-spam filters to block email spam			
Yes	11	100	80-100
No	0	0	0-20
Types of anti-spam filters used to block email spam			
Content-based filters	11	100	80-100
Origin-based filters	10	90.9	64.7-99
Types of content-based filters			
Iron Port	5	45.5	20-73
Brightmail	2	18.2	4-46.7
Barracuda	1	9.1	1-35.3
McAfee	1	9.1	1-35.3
Norman	1	9.1	1-35.3
Sophos	1	9.1	1-35.3
Forefront	1	9.1	1-35.3
Symantec	1	9.1	1-35.3
Mfiltro	1	9.1	1-35.3
Kaspersky	1	9.1	1-35.3
Types of origin-based filters			
Blacklists	10	100	78.3-100
Whitelists	3	30	9.3-60.6
Challenge response systems	2	20	4.4-50.3
Updating anti-spam filters regularly			
Yes	11	100	80-100
No	0	0	0-20

6.1.5 The Effects of Email Spam on the Performance of Saudi ISPs

Table 6.7 summarises the effects of email spam on the performance of Saudi ISPs. Most Saudi ISPs (90.9%, 95%CI: 64.7%-99%) have been affected by email spam through spending money to buy or update anti-spam filters, while the fewest reported losing customers (36.4%, 95%CI: 13.7%-65.2%) due to the receipt of a large volume of email spam.

Table 6.7: Percentage distribution of the effects of email spam on the performance of ISPs in Saudi Arabia

Effects of email spam	Frequency	Percentage (%)	95% CI
Loss of time and reduced productivity	5	45.5	20-73
Spending money to buy or update filters	10	90.9	64.7-99
Loss of customers due to receiving a large volume of spam	4	36.4	13.7-65.2
Consumption of bandwidth by excessive spam	7	63.6	34.8-86.3

6.1.6 The Effectiveness of Anti-spam Filters in Detecting Arabic and English Email Spam, as Perceived by Saudi ISPs

Table 6.8 shows that there were significant differences in the effectiveness of content- and origin-based filters in detecting Arabic and English email spam, as evaluated by Saudi ISPs. Both content- and origin-based filters were more effective in detecting English email spam than Arabic email spam (84.1% vs 70.4%, p=0.025 and 85% vs 70%, p=0.024 respectively).

Table 6.8: The effectiveness of anti-spam filters in detecting Arabic and English email spam

Types of anti-spam filters	Effectiveness in detecting Arabic email spam (%)	Effectiveness in detecting English email spam (%)	P*
Content-based filters	25	75	**0.025**
(n=11)	100	100	
	100	100	
	75	75	
	100	100	
	100	100	
	25	50	
	50	75	
	50	75	
	75	75	
	75	100	
Mean	**70.4**	**84.1**	

Types of anti-spam filters	Effectiveness in detecting Arabic email spam (%)	Effectiveness in detecting English email spam (%)	P*
Origin-based filters (n=10)	25	75	**0.024**
	100	100	
	100	100	
	75	100	
	100	100	
	75	75	
	25	50	
	75	75	
	50	75	
	75	100	
Mean	**70**	**85**	

*P values are based on paired-samples t-test between the effectiveness of content- and origin-based filters in detecting Arabic and English email spam; P values <0.05 were considered statistically significant.

6.2 Discussion

This section discusses the survey results for ISPs' definition of email spam, their awareness of government efforts to combat spam in Saudi Arabia, and the awareness and education programs they provide their customers and employees. It also discusses the nature of email spam that ISPs blocked, their dealing with it, its effects on their performance, the anti-spam filters used in blocking email spam, and their perceived effectiveness in detecting Arabic and English email spam.

6.2.1 Saudi ISPs' Definition of Email Spam, and Their Efforts to Educate Employees and Customers

The results indicated that Saudi ISPs' most common definition of email spam was UCE. This was similar to the definitions found in other studies (Boykin & Roychowdhury 2004; Bujang & Hussin 2010; Cheng 2004; Pallas & Patrikakis 2005). By contrast, other studies such as Ahmed and Oppenheim (2006), Adam (2007), Polanski (2008) and Arutyunov (2013) did not consider UCE as a definition of email spam and defined email spam as a way to promote business and product advertisements. This suggests the need in Saudi Arabia for an agreed definition of email spam that could be used to design policies, enact laws and develop anti-spam filters to combat it (Everett 2004).

In this study, about a quarter of Saudi ISPs (27.3%, 95%CI: 8.3%-56.5%) conducted workshops and training for their employees about email spam and methods of combatting it, and just over half of the ISPs (54.5%, 95%CI: 27%-80%) provided awareness programs for their customers about email spam and anti-spam filters. This indicates that more Saudi ISPs need to provide such programs for their employees and customers and in the process, they might improve their knowledge of email spam and reduce its effects on their performances. Raising employee and customer awareness about email spam is an important step to minimise it (Dantin & Paynter 2005) and can be done by establishing awareness programs such as workshops, seminars and training (Refai & Nyanchama 2007). Šolić et al. (2011) stated that:

> As there are law regulations nowadays and technical solutions like spam filters on both users' and providers' side attention should be paid to the users' behaviour and their awareness of how to suppress spam.

This is supported by studies of ISPs in Greece (Pallas & Patrikakis 2005), Singapore (Leng 2006), who suggested workshops and newsletters, and Denmark, where Frost and Udsen (2006) reported that nine Danish ISPs had setup an "ISP Security Forum" organisation, and one of the responsibilities of this organisation was to provide common guidelines for customers about email spam and filters used to combat it.

Overall, it can be concluded that the efforts of Saudi ISPs to inform employees and customers was low, and that they should focus on the awareness of Saudi society. This could be achieved by conducting campaigns to make users aware of email spam and anti-spam filters (Europa 2007), and creating awareness centres in the ISPs for providing workshops and training for their employees and customers (Refai & Nyanchama 2007).

6.2.2 The Nature of Email Spam Blocked by Saudi ISPs

This section discusses the nature of email spam blocked by Saudi ISPs, such as its volume, its languages, and types and sources (origins) of Arabic and English email spam.

6.2.2.1 The languages of email spam blocked by ISPs

The average number of email spam blocked by Saudi ISPs was 1,500,000 per week. English was the most used language of the blocked email spam, while Arabic ranked second. One reason for this could be that English is the most used language in the world (Kirkpatrick 2007) and Arabic is the official language in Saudi Arabia (Chejne 2009). Another reason could be that the anti-spam filters used by Saudi ISPs were more effective in detecting English email spam than Arabic. This is supported by other studies (Çıltık & Güngör 2008; El-Halees 2009; Nguyen, Tran & Nguyen 2008), which found that the anti-spam filters performed better in detecting English than non-English spam.

Another possibility for spammers using English in writing email spam rather than other languages is that English may be understandable by most people in different countries, so English spam could reach more recipients and be more profitable (Ermakova 2010). This finding is similar to those of previous studies (Pfleeger & Bloom 2005; Shrivastava & Bindu 2012), which have indicated that most of the world's email spam is written in English.

6.2.2.2 The types of Arabic and English email spam blocked by ISPs

The most common type of email spam blocked, for both Arabic and English, were business advertisements. This finding concurs with those of studies conducted in other countries. Dantin and Paynter (2005) found that New Zealand ISPs blocked mostly product and business offers. Chigona et al. (2005) found that South African ISPs blocked mostly business emails.

However, these types differ from the types of email spam blocked in Japan and Russia. Yamakawa and Yoshiura (2010) found that most common type of Japanese email spam were related to sexuality. Ermakova (2010) indicated that the most common types of English, French, Russian and Italian email spam blocked in Russia were focused on medicine, tourism and education. This indicates that the types of email spam in different language that were blocked in different countries could rely on factors such as the culture, religion and motivation of the spammers. Abdoh, Musa and Salman (2009) reported this effect, with email spam content (e.g. commercial, pornographic, malicious programs) differing from one country to another due to the motivations and cultures of spammers. The results of this study support this conclusion, revealing a higher percentage of pornographic emails in English than in Arabic, possibly due to the Islamic religion and the culture of Saudi society, which prohibits pornography content (Al-A'ali 2007).

6.2.2.3 *The sources (origins) of Arabic and English email spam blocked by ISPs*

Most of the Arabic email spam was sent from Saudi Arabia, while the highest percentage of English email spam was sent from non-Arabic countries. According to Lev and Goldin (2006), factors used to identify the source of email spam or spammers include spammer's Internet Protocol (IP), domain location, and the email content. A study conducted by Yamakawa and Yoshiura (2010) revealed that most Japanese email spam was sent from China and Taiwan, while only 10% of Japanese spam originated from Japan. By comparing the results of these two studies, it can be clearly seen that both Saudi Arabia and Japan have their own spammers, but the percentage of Japanese email spam originating in Japan was lower than the percentage of Arabic email spam originating in Saudi Arabia. This might be explained by the strict laws against spam in Japan; which could make Japanese spammers resort to using other countries that do not enact laws against spam, to avoid legal punishment. Lev and Goldin (2006) suggest that "the spam to legitimate email ratio in Japan is much lower than average due to the strict attitude towards law enforcement". Khong (2004) pointed to the existence and nature of laws to combat spam in some countries but not others, which may lead more spammers to choose to send spam from countries that do not legislate against sending spam to other countries. This suggests the need to enact laws against spam in Saudi Arabia to help reduce the volume of email spam.

Other studies have also investigated the origin of email spam in different countries. Hinde (2002) pointed out that most of the fraud emails received in the USA were sent from Africa, especially Nigeria. Pfleeger and Bloom (2005) reported that most of the email spam received in the EU originated from North America. Another study conducted in South Africa showed that most of the email spam were sent from China, India, and North Korea (Chigona et al. 2005). In Singapore, about 77% of email spam received was sent from outside of Singapore (Leng 2006). It can be concluded, then, that most of the email spam received in different countries such as Saudi Arabia, Japan and South Africa was originated from Asian countries, as found by Computer Fraud and Security (2008). This suggests the need for regional or international collaboration to combat spam legally, educationally and technically (Moustakas, Ranganathan & Duquenoy 2005).

6.2.3 How Saudi ISPs Dealt with Email Spam

Although technical efforts, such as security measures, are one of the essential methods ISPs should provide to protect the security of organisations and networks of customers, Saudi ISPs seemed not to consider these measures: some Saudi ISPs had not established businesses units or created teams to manage network security. This might be because Saudi ISPs outsource the task, such as to technical companies, or hire external employees to manage security issues (Frost & Udsen 2006; Ridzuan, Potdar & Talevski 2010). Studies have demonstrated the importance of security departments in ISPs. Vroom and Von Solms (2004) emphasised the security challenges emerging with the information technology revolution and the importance of assuring the organisation's security. Such departments install and update Internet security software and hardware to block security attacks (von Solms 2005), and require an adequate budget (Johnson & Koch 2006). It is suggested, therefore, that Saudi ISPs focus on technical efforts to prevent security attacks. This could be achieved by establishing security units, departments or teams.

Approximately half of Saudi ISPs (45.5%, 95%CI: 20%-73%) had employees specifically to combat email spam. Employing expert employees to deal with the issue is an important way to mitigate it. Alongi (2004) suggested that "ISPs hire employees to screen spam, install filtering programs, terminate spammer accounts, and file lawsuits". Alepin (2004) suggested that ISPs hire personnel to solve problems caused by email spam, provide technical support for customers and deal with users' complaints about spam. In different countries, ISPs create forums, groups, or teams of employees to combat spam. In the USA, one of the largest ISPs UUNET, created a team of six employees with a budget of one million dollars and with a specific responsibility to combat spam (Khorsi 2007). On a broader level, about 150 ISPs in the UK established LINX, a forum to combat spam and tackle spammers. Malcolm Hutty, a LINX regulation officer, stated that LINX was the best practice to combat spam and it contributed to reducing the volume of spam to less than 1% in the UK ('ISPs get tougher on spam' 2004). It can be seen from this discussion that the volume of email spam could potentially be reduced by creating a specific group, forum, or team of employees within ISPs or between ISPs across the country.

6.2.4 The Effects of Email Spam on the Performance of Saudi ISPs

One of the responsibilities of the ISPs is to use advanced anti-spam filters to block email spam, and these filters can be software or hardware (Lam & Yeung 2007). However, applying and updating these filters comes at a cost (Ridzuan, Potdar & Talevski 2010). Moustakas, Ranganathan and Duquenoy (2005) found that ISPs pay a lot of money for infrastructure to develop anti-spam filters (hardware or software). The results of this study indicated that the greatest impact of spam on the performance of Saudi ISPs was the substantial expense of buying and updating anti-spam filters (90.9%, 95%CI: 64.7%-99%). Nevertheless, spammers continuously develop methods to bypass these filters (Wittel & Wu 2004), which necessitates updating the filters. Previous studies have discussed the cost to ISPs of combatting spam ISPs to combat spam in different countries. The US FTC forum reported that American ISPs spent billions of dollars to stop spam (Allman 2003), and an EU study revealed that ISPs paid about 10 billion euros a year (Garcia, Hoepman & Nieuwenhuizen 2004).

Ridzuan, Potdar and Talevski (2010) claimed that although deploying or updating anti-spam filters is expensive, this is still lower than the cost of email spam to productivity. The use of anti-spam filters by ISPs can filter and delete email spam automatically, saving ISP employees’ time and increasing their productivity (Kohn 2002). The results of this study showed that approximately half of Saudi ISPs (45.5%, 95%CI: 20%-73%) reported that fixing problems related to email spam wasted on average 4 hours a week and reduced productivity. In Germany, Caliendo et al. (2012) revealed that ISPs spent an average of 25 minutes each week solving email spam issues. A US study reported a cost of 40 minutes weekly to fix such problems (Brod 2004). These results found that Saudi ISPs lost a higher average number of hours to fix problems caused by spam than those in countries such as Germany and the US. The loss of productivity has the potential to affect the economic growth of Saudi Arabia, hence the need for additional efforts by relevant agencies and government authorities to combat it.

The second impact of email spam on the performance of Saudi ISPs was consumption of the bandwidth of excessive email spam (63.6%, 95%CI: 34.8%-86.3%). This can result in the need to spend more money, to buy extra bandwidth, as

Androutsopoulos et al. (2000) and Chigona et al. (2005) have pointed out. Cournane and Hunt (2004) point to the choice between providing subscribers with a slower Internet service and paying more money to increase the bandwidth which can result in increasing subscribers' charges. This can affect the reputation of the ISPs and result in the loss of customers or subscribers (Khong 2001; Moustakas, Ranganathan & Duquenoy 2005; Potashman 2006; Smith 2004). In this study, about one-third of Saudi ISPs (36.4%, 95%CI: 13.7%-65.2%) reporting losing customers as a result of the large volume of email spam. A study conducted in the USA found that 7% of customers switched their ISPs because of an email spam issue (Gartner Group 1999). There is clearly a need for Saudi ISPs to increase their technical effort to reduce the volume of email spam, for the financial benefit not only of the ISPs themselves, but also their customers, and in turn, the economic growth of the country.

6.2.5 The Anti-spam Filters Used by Saudi ISPs to Block Email Spam, and Their Effectiveness in Detecting Arabic and English Email Spam

All Saudi ISPs used anti-spam filters to block email spam. The most common content-based filter used was Iron Port, with blacklists were the most common origin-based filters. Although all Saudi ISPs updated their anti-spam filters regularly, these filters (both content- and origin-based filters), were not perceived to be completely effective in detecting English and Arabic email spam. Spammers are constantly releasing new types of Arabic and English email spam, which can be difficult for the existing filters to block (Hayati & Potdar 2009; Wang et al. 2007). Previous studies investigated the anti-spam filters used by ISPs in several countries and their effectiveness in detecting email spam. In the USA, the most common filter used to block email spam was Brightmail, and its effectiveness in detecting email spam, as perceived by American ISPs, was 95% (Gartner Group 1999). South African ISPs applied anti-spam filters such as Postfix, Sender Policy Framework (SPF), SpamAssassin, Bayesian filters, distributed blacklists, heuristic engines, and statistical classification filters, and they found that the Bayesian filters were more effective than other filters in detecting email spam (Chigona et al. 2005). The ISPs in Greece deployed anti-spam filters such as Domain Name System Blacklists (DNSBLs), heuristic techniques, and custom technique to block email spam. They indicated that these filters misclassified some emails i.e. spam as legitimate and vice versa) (Pallas & Patrikakis 2005). In New Zealand, two ISPs (TelstraClear and Xtra) used commercial anti-spam filters to block email spam and they reported effectiveness levels of 30%-60% (Dantin & Paynter 2005). It can be concluded, therefore, that anti-spam filters are not effective in detecting all email spam, hence the need for the development and improvement of the current filters to increase their effectiveness in detecting new types of email spam in different languages. One strategy is for ISPs and companies that develop anti-spam filters to cooperate with each other. Discussion about the strengths and weakness of filters could lead to the production of filters with better performance (Potashman 2006).

In this study, Saudi ISPs perceived the anti-spam filters to be more effective in detecting English than Arabic email spam. This is in line with the findings of El-Halees's study (2009). Other studies have also found that anti-spam filters were more effective for English than other languages: Çıltık and Güngör (2008) found this result for Turkish, and Nguyen, Tran and Nguyen (2008) for Vietnamese. Subramaniam, Jalab and Taqa (2010) suggested that "anti-spam methods used for English language spam detection may not produce higher performances given the nature of different human languages". This discussion suggests that work is required to produce more effective anti-spam filters for detecting email spam in non-English languages, especially Arabic.

6.3 Conclusions

This chapter presents and discusses the efforts of Saudi ISPs to educate their customers and employees about email spam, the nature of email spam as perceived by ISPs, its effects on their performance, and how they dealt with it. This chapter also described the anti-spam filters used by Saudi ISPs to block email spam, and their effectiveness in detecting Arabic and English email spam. The following paragraphs describe the main conclusions of the ISP's survey, which addresses some of the questions of this research.

Saudi ISPs had no specific definition of email spam. And the most common definition was UCE, which is similar to the international definition (Boykin & Roychowdhury 2004; Bujang & Hussin 2010; Cheng 2004; Pallas & Patrikakis 2005). A clear and specific definition of email spam could be used in designing policies and strategies to combat email spam in Saudi Arabia.

Saudi ISPs have made several efforts to educate their employees and customers about spam, such as conducting workshops for employees and providing awareness programs for customers. However, these efforts are still few. Greater efforts are needed, and the agreement on a common definition for email spam could contribute to finding solutions to the problems it causes in Saudi Arabia.

The greatest percentage of email spam blocked by Saudi ISPs was written in English, and business advertisement emails were the most common type observed by Saudi ISPs in Arabic and English email spam. As most of the spam blocked by Saudi ISPs was written in English, this indicated that the anti-spam filters used by ISPs were more effective in detecting English email spam than non-English email spam. Therefore, work is needed to develop anti-spam filters that are more effective in detecting non-English email spam.

Arabic and English email spam were sent from different countries. Most of the Arabic email spam was sent from Saudi Arabia, which suggests that the implementation of laws against spam in Saudi Arabia and penalties for creating it could reduce the incidence of spam. On the other hand, the highest percentage of English email spam was sent from non-Arabic countries. This indicates the need for international cooperation to trace spammers' origins and to combat spam through the provision of awareness and education programs for email users about spam, the enactment of international legal regulations against spammers and development of more effective technical measures to block it (Moustakas, Ranganathan & Duquenoy 2005).

All of the Saudi ISPs used technology, mainly anti-spam filters, to combat email spam. Applying and updating anti-spam filters costs ISPs a lot of money, but the cost was a disincentive to buying and updating filters and neither content- and origin-based filters were perceived to be completely effective in detecting English but more particularly Arabic spam.

The reason why anti-spam filters were not completely effective in detecting spam could be that spammers continuously develop new tricks to bypass these filters (Hayati & Potdar 2009; Wang et al. 2007). This encouraged the researcher to investigate tricks used by spammers in Arabic and English email spam, and to suggest possible filters against spammers’ tricks, in particular those used in Arabic spam. To suggest appropriate filters against Arabic spammers’ tricks, the researcher reviewed previous studies of anti-spam filters and their effectiveness. The literature revealed that many filters have been proposed, but mostly to detect English language spam. Some of these filters have been classified by other researchers into subgroups based on the method used (e.g. reputation or content). The researcher decided to collect filters proposed or classified by other researchers and cluster them in a taxonomy (Chapter 7). This taxonomy helped the researcher in the following tasks:

- To suggest possible filters against spammers' tricks, especially for Arabic email spam, which will be presented and discussed in Chapter 8.
- To help the researchers or other future developers to improve or produce new filters for Arabic email spam.

On this basis, the proposed taxonomy is presented and discussed in the next chapter, Chapter 7. It is followed by the description and discussion of the investigation of spammers’ tricks in Arabic and English emails, in Chapter 8.

Chapter 7: A Proposed Taxonomy of Email Spam Filters

Several filters have been developed to detect email spam in different languages. However, these filters had a higher performance in detecting English email spam than non-English email spam such as Arabic (Çıltık & Güngör 2008; El-Halees 2009; Nguyen, Tran & Nguyen 2008). As mentioned in the previous chapter (Chapter 6), Saudi ISPs reported that the existing anti-spam filters were more effective in detecting English email spam than Arabic email spam. This encouraged the researcher to review previous studies that propose filters to detect email spam. The literature revealed that a wide variety of filters have been proposed to detect English spam, and that some of the proposed filters have been classified by other researchers into subgroups based on the method used (e.g. content or reputation-based filters). This resulted in collecting these filters that proposed or classified by other researchers to detect email spam, and cluster or organise them, in this study, into a taxonomy.

Taxonomy plays a significant role in research and content management because it helps researchers understand and analyse complex domains. Taxonomy is important for the organisation of any information (Sujatha & krishna Rao 2011). The major advantages of taxonomies include the reduction of complexity and the identification of similarities and differences among objects (Bailey 1994). Consequently, the taxonomy proposed in this study could provide suggestions for new filters for the spammers' tricks observed in Arabic email spam, which will be presented and discussed in Chapter 8. This could be achieved by looking at filters used and their effectiveness in detecting English email spam, and then deciding which of these filters could be used to develop new filters to detect Arabic email spam.

This chapter is divided into the following sections:

- Section 7.1: describes the methodology used to develop the proposed taxonomy.
- Section 7.2: further explains the email spam filters included in the taxonomy.
- Section 7.3: discusses the effectiveness of reputation- and content-based filters, classified in the taxonomy, in detecting email spam.

- Section 7.4: concludes this chapter.

7.1 Methodology Followed in the Development of the Proposed Taxonomy

This section describes the process used to develop a taxonomy of email spam filters. It begins by reviewing taxonomies proposed by other researchers to participate in the fight against network security attacks, including spam. It also describes how the proposed taxonomy of email spam filters in this study has been developed by explaining the methodology followed to build the current taxonomy.

7.1.1 Previous Taxonomies in the Field of Network Security

Before discussing the methodology used to develop the taxonomy of email spam filters, it is necessary to review the literature on this subject. In the field of network security, including spam, a few taxonomies have been proposed. The aim of these taxonomies was to raise the awareness and interest of the research community about security issues (Gyongyi & Garcia-molina 2004), and to develop cost-effective countermeasures against security attacks (Mirowski, Hartnett & Williams 2009).

Hansman and Hunt (2003) proposed a taxonomy of network and computer attacks. The authors divided the attacks into ten types: viruses, worms, trojans, buffer overflows, denial of service attacks, network-based attacks, physical attacks, password attacks, information gathering attacks, and blended attacks. Each type of attacks was divided into subtypes. The authors hoped that a taxonomy would increase the users' knowledge about different types of security attacks, and help different bodies such as information and advisory bodies to understand and develop methods to combat them.

Weaver et al. (2003) presented a taxonomy of computer worms based on worm target discovery and selection strategies, worm carrier mechanisms, worm activation, possible payloads, and plausible attackers. The authors noted that this taxonomy could help in understanding the classes of worms, the attackers who may employ them, and the potential payloads, which in turn could help understand the threats they pose.

Mirowski (2009) developed a taxonomy of radio frequency identification (RFID) attacker behaviours. The attacker behaviour was divided into two major types: attacker behaviour in a RFID authorisation system, and attacker behaviour in a RFID monitoring system. The authorisation system attacker behaviour was divided into two types: original tag and clone tag; and the monitoring system attacker behaviour was classified into three types: deny tag identification, deny reader and deny middleware database. The authors believed that “this taxonomy will help security practitioners understand RFID system security requirements and lead to more cost-effective security countermeasures”.

In the field of spam, a few taxonomies have been proposed, mostly about classification of web spam methods. Gyongyi and Garcia-Molina (2004) produced a comprehensive taxonomy of current web spam techniques proposed in previous studies to combat web spam. The proposed taxonomy included two main techniques: boosting techniques and hiding techniques. The boosting techniques were classified into two techniques: term spamming and link spamming; and hiding spamming was divided into three techniques: content hiding, cloaking and redirection. The authors explained each technique in the taxonomy and they believed that this taxonomy could help develop appropriate countermeasures against web spam.

Another taxonomy of web spam detection methods was developed by Ghiam and Pour (2012). It is an update of a previous taxonomy presented by Gyongyi and Garcia-Molina (2004) which was described above. In this taxonomy, the authors added some techniques that were not included in the previous taxonomy, and modified others. They classified web spam detection into three types of techniques: link-based, hiding, and content-based. The link-based techniques were divided into link and ranking, and link farm. The hiding techniques were organised into two methods: cloaking and redirection; and content-based techniques were classified into two methods: link and content, and content analysis.

Alperovitch, Judge and Krasser (2007) presented a taxonomy that examined email reputation systems properties, and they surveyed some of the approaches in previous works. The author categorised email reputation systems into two main identifiers: address-based identifiers and content-based identifiers. The address-based identifiers were classified into three sub-identifiers: IP-based, domain-based and email address–based identifiers. Content-based identifiers were categorised into two methods:

fingerprinting and approximate text addressing.

From the previous paragraphs, it can be clearly seen that a number of taxonomies in the field of network security have already been developed by previous studies. The researchers claimed that these taxonomies could help in developing methods to combat different attacks of network security, including spam (Gyongyi & Garcia-molina 2004; Mirowski, Hartnett & Williams 2009). In the field of spam, a few taxonomies were found in previous studies, most of them related to the classification of web spam techniques. The literature reveals many methods that have been developed to combat email spam. However, no single taxonomy could be found to classify, organise or cluster these different methods. Therefore, this study tried to cover this gap by presenting a taxonomy of major email spam filters that could help in developing methods against email spam, particularly Arabic email spam. The development of proposed taxonomy is described in the following section.

7.1.2 The Proposed Taxonomy

One of the objectives of this study was to propose a taxonomy of email spam filters, comprising most of the anti-spam filters used to detect email spam. The purpose is to suggest which of these filters might be selected to develop new filters for Arabic email spam. The literature review found that methods have been developed to detect email spam, but no taxonomy that clusters or organises these filters. Therefore, this study cover this gap by presenting a taxonomy of current email spam filters to aid understanding of the anti-spam filters used to detect email spam, and by then suggesting the appropriate filters to detect spammers' tricks used in Arabic email spam (presented and discussed in Chapter 8).

The development of a taxonomy is a difficult process. The discipline of biology, which has a well-known taxonomy of living organisms (i.e. the Linnaean taxonomy), provides some guidance. According to Nickerson et al. (2009), "the traditional Linnaean taxonomy classifies organisms based on a predefined hierarchy of categories from kingdom to species". There are also two types of taxonomy development in biology: phenetics and cladistics (Stevens 2003). Phenetics (also called numerical taxonomy) classifies organisms on the basis of similarity between their characteristics (Nickerson, Muntermann & Varshney 2010). Cladistics, however, "looks at the evolutionary relationships among organisms, not just their common features" (Nickerson et al. 2009).

The phenetics approach or numerical taxonomy was used, in the informatics, to classify objects based on their similarities (Nickerson, Varshney & Muntermann 2013). This study also adopted the phenetic approach to classify anti-spam filters on the basis of the similarity between the methods used in filters to detect email spam.

7.1.2.1 Taxonomy Requirements

Before starting the work on designing a new taxonomy, it is important to define what a good taxonomy should include, or the requirements needed to produce a useful taxonomy (Hansman & Hunt 2003). Several requirements or attributes have been defined in previous studies (Hansman & Hunt 2003; Nickerson et al. 2009), and can help in producing a good taxonomy. These requirements are:

1. The taxonomy should be structured.
2. The taxonomy should be comprehensible, so that it can be understood by experts in the field of network security, and also by other interested people.
3. The terminology should be based on existing knowledge and usage (in the field of email spam), to avoid confusion.
4. The terms used in the taxonomy should be well-defined, to guarantee that there is no confusion in understanding its meaning.
5. The taxonomy should be concise, with a limited number of characteristics, as overly complex classification with many characteristics might be difficult to apply.
6. The taxonomy should be extendible to allow for adding new characteristics when new types of objects appear.

Therefore, all previous requirements or attributes were followed in producing the taxonomy proposed in this study. To meet the second requirement, the taxonomy was tested by two experts in the field of network security and by other interested two persons. A checklist of these requirements was forwarded to them. Their feedback, mostly about terminology and the taxonomy structure, was received and the taxonomy was modified, based on their comments.

7.1.2.2 Taxonomy Elements (Constructs)

This study used the phenetics approach, which classifies objects based on the similarities between characteristics. Consequently, this study classified the anti-spam filters based on the similarity between the methods used by filters to detect email spam. Previous studies have found that two major techniques are used to develop filters to combat email spam: reputation- and content-based techniques. The reputation-based techniques rely on information outside of the content of email messages (e.g. IP address or sender domain) to detect spam (Golbeck & Hendler 2004), while the content-based techniques detect spam by examining the content of email messages (Cook et al. 2006). A number of filters proposed by previous studies were listed (see Table 7.1), under these two major techniques.

Table 7.1: Types of email spam detection techniques

Email SPAM Detection Techniques	
Reputation-based Techniques	**Content-based Techniques**
1. Blacklist	1. Rule-based
2. Whitelist	2. Bayesian method
3. Challenge response system	3. Chi-squared method
4. Origin diversity analysis	4. Support vector machine
5. Implicit method	5. Boosting method
6. Explicit method	6. Maximum entropy model
7. Country-based	7. Memory-based learning
8. Open proxy detection	8. Genetic algorithms
9. Grey list	9. Artificial immune system
10. ProMail	10. Artificial neural network
11. SMTP logs mining	11. k-nearest neighbours clustering
12. Mail volume	12. Density-based clustering
	13. Decision trees
	14. Fingerprinting method
	15. Perceptron method
	16. Multi-layer networks
	17. Learning vector quantisers
	18. LINGER
	19. Honeypot method
	20. Zombie-based
	21. Keywords-based
	22. Phonetic string matching
	23. Random forest
	24. Classification and regression trees
	25. Random trees
	26. C4.5 (J48)
	27. Signature/checksum scheme
	28. K-mer
	29. Bayesian Mail Filter

Email SPAM Detection Techniques	
Reputation-based Techniques	**Content-based Techniques**
	30. MN TF NB 31. Bogofilter 32. SpamCop 33. SpamBayes 34. Flexible Bayes 35. MV Gauss NB 36. MN Boolean NB 37. MV Bernoulli NB 38. RIPPER 39. REPTree 40. Decision stump 41. ID3 42. ADTree

As well as their use in different sciences such as medicine and biology (Gopalakrishnan et al. 2006; Jin et al. 2006), some well-known techniques, such as decision trees and Bayesian, shown in Table 7.1, have also been used to develop filters to combat email spam. These methods have been classified by researchers who work in the field of spam, into different subgroups. The classification of these methods is discussed in the following two sections.

7.1.2.2.1 Classification by previous studies of reputation-based filters

Several filters, based on the information outside of content of messages such as IP address and sender domain, have developed to detect email spam (See Table 7.1). Some of these filters have been clustered by other researchers in one main group or classified into subgroups.

Whitelist, blacklist and challenge response system filters have been clustered by previous researchers into one main group, named origin-based filters (Garcia et al. (Cook et al. 2006; Garcia, Hoepman & Nieuwenhuizen 2004; Pfleeger & Bloom 2005). They defined origin-based filters as methods that classify email spam, based on the network information, such as the source of IP and email addresses. Based on the origin of email spam, Gardner-Stephen (2009) proposed a new filter that detects spam by clustering similar messages, and then considering the claimed origins of the messages.

Some subgroups of origin-based filters, such as blacklists and challenge response systems, were divided into sub-filters. Blacklist filters were defined as methods that list IP addresses of suspected or known spammers, and they were divided by Cook et al. (2006) and Heron (2009) into two subgroups: country-based filters and open proxy detection filters. Challenge response systems aimed to send an automated reply or challenge to senders requiring some action to prove that they are real users (Heymann, Koutrika & Garcia-Molina 2007). Islam et al. (2009) developed a filter based on the challenge response systems, called grey list.

Garcia, Hoepman and Nieuwenhuizen (2004) identified a type of reputation-based method called mail volume filter, which depends on the analysis of network traffic stream (traffic analysis–based filters) to detect spam. Implicit and explicit filters were clustered by Boykin and Roychowdhury (2004), and Okolica, Peterson and Mills (2008) in one main group called social network–based filters, which were defined as methods that detect email spam by assigning to each message a probability of it being spam, based on the past history of the participants (Boykin & Roychowdhury 2004). On basis of the implicit social network techniques that analyse fields of emails headers such as ‘To’, ‘Cc’ and ‘Bcc’ to build a graph of social relations of users and classify new emails based on this graph (Blanzieri & Bryl 2008), Tseng, Huang and Chen (2007) proposed a ProMail filter to detect email spam through constructing a social network graph of email passing through a SMTP server, often by mining log files.

7.1.2.2.2 Classification by previous studies of content-based filters

As seen in Table 7.1, different filters have been proposed to detect spam based on the analysis of the message’s content. Cook et al. (2006) defined a type of content-based filter called heuristic filters. The authors defined heuristic-based filters as methods that search for patterns that are commonly identified in spam, and they organised them in one sub-filter called a rule-based filter. Rule-based methods classify email spam by the occurrence of critical keywords. Keyword-based filters were the most common type of these methods (Cook et al. 2006). The phonetic string matching filter was an improved type of keyword-based filter proposed by Freschi, Seraghiti and Bogliolo (2006) as a solution to combat word obfuscation, which had been developed by spammers to bypass keyword filters. On the basis of rule-based filters, Cohen (1995) developed a filter to combat email spam, called Repeated Incremental

Pruning to Produce Error Reduction (RIPPER).

Guzella and Caminhas (2009) and El-Halees (2009) have organised the following well-known methods into one main group, called machine learning. These methods included statistical methods, genetic algorithms, artificial immune systems, artificial neural networks, clustering methods, and decision trees. Some of these filters have been classified into subgroups. Machine learning methods aim to avoid the human labour required to maintain rule-based filters by automatically deriving a non-spam/spam classifier (Guzella & Caminhas 2009). Statistical methods classify email spam based on the statistical properties: if properties are closer to the corpus of spam emails, email is classified as a spam; other email is classified as non-spam (Zhang, Zhu & Yao 2004). Zhang, Zhu and Yao (2004) classified statistical filters into six subgroups: Bayesian, support vector machine (SVM), chi-squared, boosting, maximum entropy, and memory-based learning. Artificial neural networks detect spam from common features in emails (Garcia, Hoepman & Nieuwenhuizen 2004). Artificial neural networks were divided by El-Halees (2009) into two methods: perceptron and multi-layer perceptron. On the basis of artificial neural networks, two filters were also developed: LINGER (Clark, Koprinska & Poon 2003b), and learning vector quantisers (LVQ) (Chuan et al. 2005).

Clustering techniques are used to detect email spam by clustering emails into groups (clusters). The filter or classifier is trained on each group, and then detects spam by identifying its cluster (Saeedian & Beigy 2009). Clustering techniques were classified by Ungar and Foster (1998), and Yoshida et al. (2004) into two methods: K-nearest neighbours (K-NN) clustering and density-based clustering. Decision tree methods were used to detect email spam from the analysis of each email categorisation (e.g. sexual, financial or advertising) to find the association rules of spam emails in the categorisation (Sheu 2009). Decision tree methods were divided by Abu-Nimeh et al. (2008), Sharma and Sahni (2011) and Chaudhary, Kolhe and Kamal (2013) into eight techniques: C 4.5 (J48) tree, ID3, ADTree, REPTree, random tree, decision stump, classification and regression tree, and random forest.

Bayesian is a statistical method proposed by Sahami et al. (1998) to detect email spam by considering the historical probability of each word in the message occurring in either spam or non-spam messages (Schneider 2003). Previous studies, such as

Metsis, Androutsopoulos and Paliouras (2006), and Almeida, Yamakami and Almeida (2009), organised the Bayesian method into six filters: Multivariate Bernoulli Naïve Bayes (MBNB), Multinomial Term Frequency Naïve Bayes, Multinomial Boolean Naïve Bayes, Multivariate Gauss Naïve Bayes, Boolean Naïve Bayes, and Flexible Bayes. Five filters based on the Bayesian method have also been developed to detect email spam: SpamCop (Pantel & Lin 1998), Bayesian Mail Filter (BMF) (Garcia, Hoepman & Nieuwenhuizen 2004), SpamBayes (Meyer & Whateley 2004), and Bogofilter (Raymond 2005).

Fingerprinting is another content-based filters that detects spam by computing and comparing the finger print of any incoming email (Goodman, Heckerman & Rounthwaite 2005). Using fingerprinting methods, various studies have developed filters to detect email spam. These filters included honeypot (Oudot 2003), digest-based filters (Damiani et al. 2004), signature/checksum schemes (Kołcz, Chowdhury & Alspector 2004) and zombie host detection filters (Duan et al. 2012). Using honeypot methods, a BrightMail filter has been developed to detect email spam (Allman 2003).

It can be concluded that numerous filters have been developed to combat email spam, of which there are several main groups, some of which can be further divided into subgroups. The literature revealed that all of these filters could be classified as using either a reputation- or a content-based method, but no taxonomy was found that organises these filters. Therefore, this study covered this gap by presenting a taxonomy that organises the different filters that have been developed, based on the similarities between their methods of detecting email spam. The proposed taxonomy could help anti-spam developers and people who work in the field of security to develop countermeasures against email spam. In particular, the taxonomy could be used to suggest new filters against spammers' tricks observed in Arabic email spam, as most of the filters included in the taxonomy were most successful detecting English spam. Arabic spammers' tricks are presented and discussed in Chapter 8.

The proposed taxonomy is presented in Figure 7.1. More details about each method and filter abbreviations included in the taxonomy are discussed in the next section.

7.2 The Proposed Taxonomy of Email Spam Filters

This section describes in detail the proposed taxonomy of major email spam filters.

7.2.1 Reputation-based Filters

This major class of email spam filter relies on information outside of the content of the individual email messages to detect spam (Golbeck & Hendler 2004; Prakash & O'Donnell 2005; Zheleva, Kolcz & Getoor 2008). These filters make assessments about the reputation of one or more of the participants (sender, recipient and intermediaries) in the email transaction. The various methods differ in the subject in how the reputation is classified, for example IP address, sender domain and sender address, and also in the nature of the reputation calculation (Golbeck & Hendler 2004). The reputation-based filters are divided into three major techniques: (a) origin-based; (b) social-based, and; (c) traffic analysis–based.

7.2.1.1 Origin-based filters

Origin-based filters classify spam from the network information, such as the source IP and email addresses (Cook et al. 2006). Such analyses have the advantage that they can be performed before an email is received by recipients, potentially saving network and computational resources (Garcia, Hoepman & Nieuwenhuizen 2004). Numerous origin-based techniques exist, including: (1) blacklists; (2) whitelists; (3) challenge response systems (CRS), and; (4) origin diversity analysis.

7.2.1.1.1 Blacklists (BL)

Blacklists include the realtime blackhole lists (RBL) and domain name system blacklists (DNSBLs). These databases list IP addresses of suspected spammers or known spammers (Cook et al. 2006). With these blacklists spam can be blocked at the SMTP connection phase (Ramachandran, Feamster & Vempala 2007).

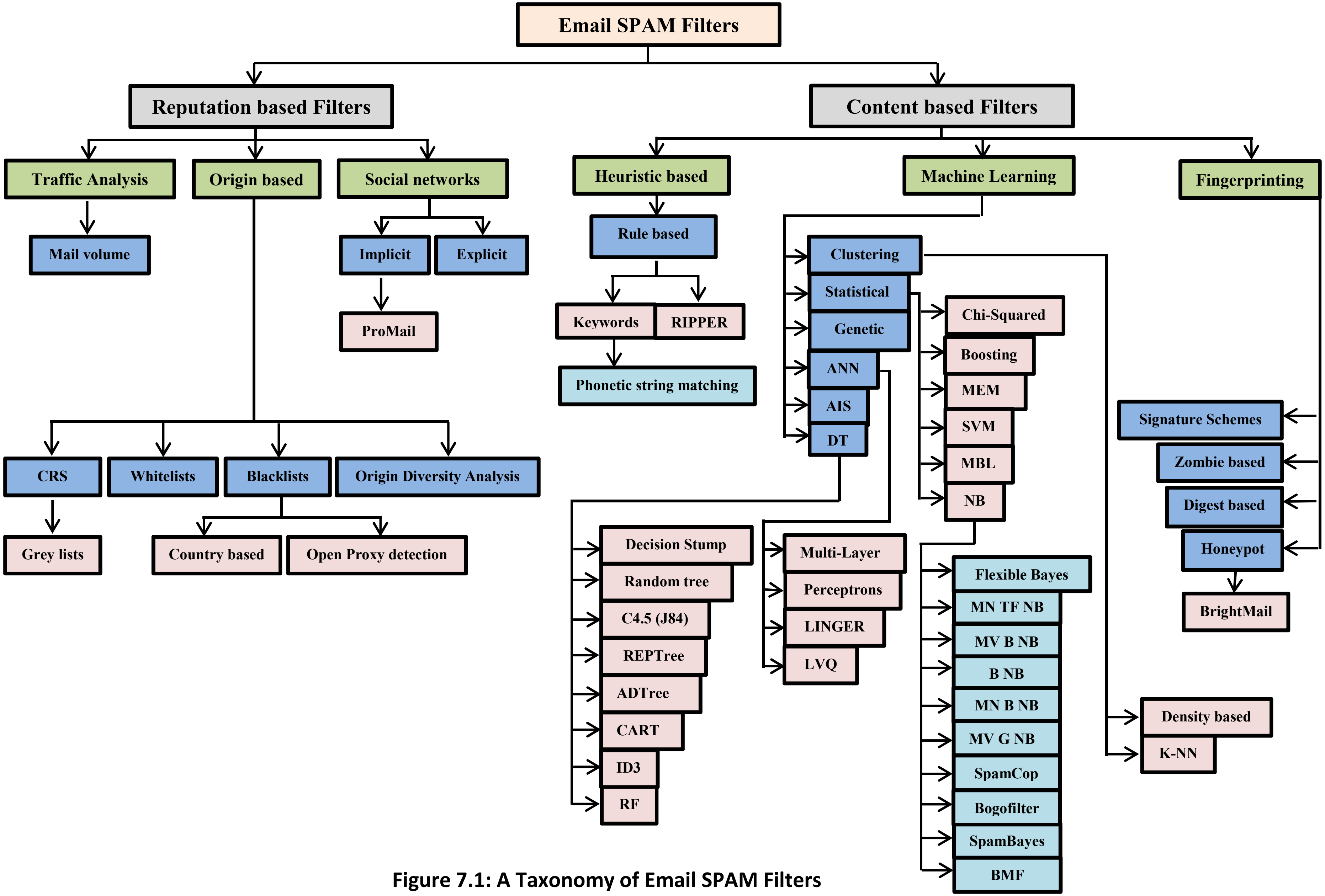


Figure 7.1: A Taxonomy of Email SPAM Filters

However, while blacklists are reasonably effective and efficient, they have disadvantages. First, blacklists are maintained by an entity distinct from the user, introducing an external dependency into any spam filter that relies on them (Cook et al. 2006). Second, the effectiveness of blacklists depends on the timeliness and methods of those who manage them (Dudley, Barone & While 2008). Finally, because most blacklists are usually queried via the Domain Name System (DNS), this can result in substantial DNS traffic and consequent delays in spam processing, especially for mail servers that reference more than one blacklist (Sanz, Gómez Hidalgo & Cortizo Pérez 2008).

Many methods are used to produce and maintain blacklists, including open-relay detection and country-based. Firstly, an open relay or proxy is a SMTP server that indiscriminately relays all email it is presented with, without validating the headers of those messages. Such servers allow spammers to not only hide their origin, but also to add falsified headers to misdirect anyone seeking to identify their source (Andreolini et al. 2005). For this reason several blacklists routinely list all open proxies that they identify (Garcia, Hoepman & Nieuwenhuizen 2004). Secondly, some countries are considered to be particularly notorious sources of spam. Some blacklists reject mail from any SMTP server in that country, which exacerbates the false-positive problem (Heron 2009). Indeed, false positives are a problem for many spam countermeasures, and thus it is common to combine the opinion of multiple filtering methods before making a final classification (Sinha, Bailey & Jahanian 2008).

7.2.1.1.2 Whitelists (WL)

Whitelists enable users to create a list of trusted addresses that they nominate for exclusion from spam filtering. Email originating from all other addresses are filtered as normal (Pfleeger & Bloom 2005). Whitelists reduce the cost of filtering spam, because some email messages are allowed to bypass the filtering process (Cook et al. 2006). However, whitelists require user interaction. Also, many whitelist implementations rank white–listed email above all other email in the inbox. Thus, whitelists can make it difficult to identify email that is not white-listed, but is nonetheless legitimate, due to its reduced visibility, especially when a large proportion of spam is present (Golbeck & Hendler 2004).

7.2.1.1.3 CRS

Whereas whitelists place the burden on the receiver for determining trustworthy senders, CRSs transfer the burden of authentication to the sender. Senders receive an automated reply or challenge, which requires some action to prove that they are real users (Heymann, Koutrika & Garcia-Molina 2007; Templeton 2004; Xie, Yin & Wang 2006). For example, the system might send an image that contains a picture of animals to the sender. The sender is asked to count the number of animals in the picture. Such tasks are chosen to be trivial for a real person, but too difficult for a computer to perform quickly enough to facilitate effective spamming (O'Brien & Vogel 2003).

Thus, one advantage of CRSs is that they protect against automated spam-sending programs. A second advantage is that they require that the spam originate from a functional mailbox that is monitored by the spammer (Hird 2002). If it is monitored, it provides an easy means of identifying and filtering spam from that sender. Finally, CRSs can be used to populate whitelists (Cook et al. 2006).

A critical disadvantage of challenge response filters is that if both sending and receiving mail servers implement them, a deadlock results, as each server wait for the other to respond to its challenge (Cook et al. 2006). One way to reduce, but not eliminate, this deadlock problem is to use CRSs in a grey-listing filter. In this way, the challenge is sent only for email that is considered possible spam (Harris 2003).

7.2.1.1.4 Origin diversity analysis

The origin diversity analysis method is a hybrid origin- and content-based technique that focuses on the behaviour of spam emails instead of the content of spam messages. This method clusters similar messages and then considers the claimed origins of the messages. If there are many putative origins, then it is assumed that most of them must be fraudulent, and hence likely to be spam (Gardner-Stephen 2009). The claimed advantage of this scheme is that it relies only on a distinguishing behaviour of spam in that it arrives in quantity from many apparent locations, since any deviation from that combination would reduce the effectiveness of the spam. Thus the origin diversity analysis method has the potential to be robust in the face of the continuing evolution of spam (Gardner-Stephen 2009).

7.2.1.2 Social network-based filters

Social network–based filters aim to assign to each message a probability of it being spam, based on the past history of the participants (Boykin & Roychowdhury 2004). Social network–based filters are classified into: (a) implicit filters, and; (b) explicit filters.

7.2.1.2.1 Implicit filters

Implicit filters are used to combat email spam by analysing fields of emails headers, such as 'To' , 'Cc' and 'Bcc', to build a graph of the social relations of users and classify new emails based on this graph (Blanzieri & Bryl 2008; Boykin & Roychowdhury 2004). ProMail filter is a type of implicit filter. ProMail and related methods construct a social network graph of email passing through an SMTP server, often by mining log files (Tseng, Huang & Chen 2007). Typically, nodes in the graph represent email accounts, while edges represent email transactions (Hayati & Potdar 2008). These graphs are used to make decisions about whether or not a message is likely to be from a source in the recipients' social network, and hence more likely to be legitimate (Lam & Yeung 2007).

7.2.1.2.2 Explicit filters

In contrast to the implicit methods, there exist methods that explicitly build the social network through user interaction and may also utilise user-supplied or automatically computed reputation ratings (Golbeck & Hendler 2004). These methods are naturally complementary with white listing and CRSs (Boykin & Roychowdhury 2004).

7.2.1.3 Traffic Analysis-based Filters

Traffic analysis methods detect anomalies and patterns in the network traffic stream by mining the log files of an SMTP server (Bindu & Thomas 2012). Although other analyses are possible, one common analysis (called mail volume filter) that is used to detect spam is to identify when a host or network issues an abnormally large amount of email. However, this technique results in a very high false acceptance rate (FAR) (Garcia, Hoepman & Nieuwenhuizen 2004).

7.2.2 Content-based filters

In contrast to reputation-based filters, content-based filters detect spam by examining the content of email messages, irrespective of the origin (Cook et al. 2006). Common traits of these techniques include: (1) they require the body of a message before they can classify messages as spam or legitimate, and thus incur the use of more network bandwidth compared with reputation-based filters, and; (2) they are immune to the originating location of message, unlike origin-based techniques (Garcia, Hoepman & Nieuwenhuizen 2004). There exist several families of content-based filtering techniques, including: (a) heuristics; (b) machine learning, and; (c) fingerprinting.

7.2.2.1 Heuristic-based filters

In these filters, email can be classified as spam by searching for patterns that are commonly identified in spam. Patterns can be specific words, phrases, malformed message headers, exclamation marks and capital letters (Cook et al. 2006). Perhaps the most common type of heuristic filter is the rule-based filter.

7.2.2.1.1 Rule-based filters

Rule-based filters were very common and popular until 2002, when Bayesian filters were released (Graham 2003). The classification of spam emails relied on user-specified rules, which characterise known unwanted emails (Cook et al. 2006). Rule-based filters depend on the occurrence of critical keywords (i.e. keyword-based filters) to classify spam. They not only analyse the content of email, but also the email header, which contains list of the recipients, IP address's source and subject (Freschi, Seraghiti & Bogliolo 2006).

An example of the rule-based filters that used to filter email spam was RIPPER (Cohen 1995, 1996). The RIPPER filter is:

> … a propositional learner designed for efficient performance on large, noisy datasets. RIPPER is designed to handle set- and bag-valued attributes equivalently by generating keyword-spotting rules (Cohen 1995 cited in Provost 1999)

However, there is a problem with keyword-based filters – word obfuscation. For example, a keyword-based filter might have a rule to match the word '*Free*', but that rule would not necessarily match the strings *'f*r*e*e*' or '*bonus*' (Cook et al. 2006). One partial solution to word obfuscation is phonetic string matching, which provides a more robust pattern-matching based on its phonetic transcription. This technique seeks to address the problems experienced by keyword-based filters when confronted with word obfuscation (Freschi, Seraghiti & Bogliolo 2006).

7.2.2.2 Machine learning filters

Machine learning techniques (ML) aim to avoid the human labour required to maintain rule-based filters by automatically deriving a legitimate/spam classifier. By definition, these techniques need to be fed pre-classified training data, although once primed, many can provide their own forward training to attempt to keep abreast of the evolution of spam (Guzella & Caminhas 2009). Some categories of machine-learning spam techniques are: statistical filters, genetic algorithms, artificial immune systems, artificial neural networks, clustering filters and decision trees.

7.2.2.2.1 Statistical filters

Statistical filters rely on a corpus of spam emails and legitimate emails to conclude features, which can be used to classify incoming emails. If the statistical properties are closer to the corpus of spam emails, the email is classified as a spam. If the statistical properties are closer to the legitimate emails corpus, the email is classified as legitimate (Zhang, Zhu & Yao 2004). As with rule-based filters, many statistical filters also consider the header portion of messages (Freschi, Seraghiti & Bogliolo 2006). A selection of statistical spam filters is: Bayesian, chi-squared, SVM, boosting, maximum entropy models (MEM) and memory-based learning filters.

Naïve Bayesian filters

Naïve Bayesian spam filters consider the historical probability of each word in the message occurring in either spam or non-spam messages (Androutsopoulos, Koutsias, et al. 2000; Sahami et al. 1998; Schneider 2003). They calculate the probability that the email is spam or non-spam by combining the individual spam/legitimate probability of each word, or k-mers (Sculley, Wachman & Brodley 2006) inside the message to produce a final probability estimate that an email is spam or non-spam (Cook et al. 2006). The percentage of false positives generated by Naïve Bayesian filters is low, and they are self-adapting to stop new spam by receiving ongoing training form the user. While extremely effective for a time, more recently Naïve Bayesian filters have become less effective due to the common practice of including random blocks of text in spam messages to reduce the accuracy of this detection technique (Garcia, Hoepman & Nieuwenhuizen 2004).

There exist different versions of the Naïve Bayesian methods used in filtering spam. Multinomial term frequency Naïve Bayes (MN TF NB) was defined by Almeida and Yamakami (2009) as a method that represents each message as a set of terms and computes each term by how many times it appears in a message. Multivariate Bernoulli Naïve Bayes (MV Bernoulli NB) is a statistical method used to classify spam, based on the probability of each message being represented by calculating the presence or absence of each term from the message. The Boolean Naïve Bayes (Boolean NB) classifier is similar to the MV Bernoulli NB, except that it does not consider the absence of the terms (Almeida & Yamakami 2010).

Multinomial Boolean Naïve Bayes (MN Boolean NB) is similar to the MN TF NB methods in estimating probability and differs from the MV Bernoulli NB in that it does not compute the absence of each term from the message. Multivariate Gauss Naïve Bayes (MV Gauss NB) used real-valued attributes by assuming that each attribute follows a Gaussian distribution for each category (Metsis, Androutsopoulos & Paliouras 2006). Flexible Bayes was defined by Méndez et al. (2008) as a method that:

> … works in a similar way than MV Gauss NB, but the distributions of each attribute are estimated by means of Gaussian distributions representing each different value for the attribute in each class.

Many programs or open-source filters have been developed on basis of the Naïve Bayesian methods to detect email spam. SpamCop is a program that was developed by Pantel and Lin (1998) to classify email spam. It treats a message as a multi-set of words and uses Naïve Bayesian to determine if a message is spam or non-spam. SpamBayes is a program developed by Tim Peters and others on 2002 to classify email spam using Naïve Bayesian methods (Meyer & Whateley 2004). Bogofilter is an open-source filter that employs the Naïve Bayesian to classify email spam (Raymond 2005). Another Naïve Bayesian filter used to detect email spam is BMF (Garcia, Hoepman & Nieuwenhuizen 2004).

Chi-squared filters

Chi-squared filters are used in the field of authorship identification and to detect spam emails. These filters can detect email spam by applying the chi-based authorship identification technique to the spam identification problem (O'Brien & Vogel 2003).

SVM filters

Used in text classification, SVMs are supervised learning methods that have more recently been applied to the spam identification problem (Cristianini & Shawe-Taylor 2000; El-Halees 2009; Guzella & Caminhas 2009; Lynam, Cormack & Cheriton 2006; Zhang, Zhu & Yao 2004).

Boosting filters

Boosting is a ML algorithm that is based on the idea of combination of many weak hypotheses (Ali & Yang 2007; He & Thiesson 2007; Jin et al. 2003). For example, in the AdaBoost system (Zhang, Zhu & Yao 2004), a learner is trained in each stage of the classification procedure, and the output of each stage is used to reweigh the data for the future stages (Blanzieri & Bryl 2008). Boosting algorithms with confidence-rated predictions have been proposed as being well suited to the spam-filtering problem and capable of outperforming both Bayesian and decision tree methods (Carreras & Marquez 2001).

Maximum entropy models (MEM) filters

Another ML technique from natural language processing is the MEM, which has also been applied to spam filtering (Khorsi 2007; Zhang & Yao 2003; Zhang, Zhu & Yao 2004).

Memory-based learning filters

Memory-based learning filters (MBL) are a “non-parametric inductive learning paradigm that stores training instances in a memory structure on which predictions of new instances are based” (Zhang, Zhu & Yao 2004) that have also been applied to the problem of spam (Sakkis et al. 2003).

7.2.2.2.2 Genetic algorithm filters

Genetic spam detection algorithms use feature detectors, often evolve over time, and are used to score emails. The classification of emails as spam or non-spam is based on the integration of one or more such feature scores (Garcia, Hoepman & Nieuwenhuizen 2004; Goweder, Rashed & Alhamammi 2008).

7.2.2.2.3 Artificial immune system (AIS) filters

Artificial immune systems are ML methods used to fight spam and viruses of computers. They use methods that are in some way based on the immune system of biological organisms (Guzella & Caminhas 2009; Nicosia 2004; Oda 2005; Oda & White 2005; Secker, Freitas & Timmis 2003; Sirisanyalak & Sornil 2007; Yue et al. 2007). The classification of emails as spam or legitimate in this technique can be based, for example, on artificial lymphocytes created from a gene database, where the genes represent mini-languages to include keywords, which are checked in spam (Khorsi 2007).

7.2.2.2.4 Artificial Neural Networks

Artificial Neural Networks (ANNs) are a common classification technique in artificial intelligence applications (Guzella & Caminhas 2009). They represent networks of virtual neuron cells and are trained to perform some tasks (Puniškis, Laurutis & Dirmeikis 2006). For spam detection, ANNs typically classify incoming emails based on common features of emails (Garcia, Hoepman & Nieuwenhuizen 2004). There are many types, including perceptrons, multi-layer networks, LVQs (El-Halees 2009) and LINGER (Clark, Koprinska & Poon 2003a, 2003b).

Perceptron filters

Perceptrons are generated by trying to find a linear function f(x) for feature vector, which ideally produces distinct ranges of output values for a given message (i.e., f(x) > 0 for vectors if the message is spam, and f(x) < 0 for vectors if the message is legitimate (Khorsi 2007).

Multi-layer network filters

A multi-layer neural net is "a network of connected perceptrons which from a network with successive layers" (Khorsi 2007), as such they are potentially more powerful than perceptrons (Wesley-Smith 2006).

Learning vector quantiser filters

Learning vector quantisers cultivate a set of neurons, selecting the best neuron for each classification task and preening those neurons to increase their accuracy. LVQs are well suited to text classification tasks, and have been applied to the spam classification problem with results superior to both Bayesian and various other forms of ANNs (Chuan et al. 2005).

LINGER filter

LINGER is a system that depends on neural networks to classify email spam. It is flexible and more accurate than other methods used to detect spam. LINGER includes two modules: a pre-processing module (which contains pre-processing for word extraction, feature selection, weighting and normalisation), and a classification module (Clark, Koprinska & Poon 2003b).

7.2.2.2.5 Clustering filters

Clustering filters can be effective ML tools to detect email spam. They cluster emails into groups, and the filter or classifies is trained on each group and then detect spam by identifying its cluster (Saeedian & Beigy 2009). Two types of clustering filters that have been applied to spam classification are K-NN and density-based clustering.

K-NN clustering filters

K-nearest neighbours (K-NN) clustering indexes and converts emails to a high-dimensional vector and then measures the distance between the vectors of each email cluster formed of neighbouring (i.e. relatively close) vectors. Once clusters have been formed, spam classification need only be performed for a sub-set of any cluster population, as the result can then be inferred to apply to the other members of the cluster (El-Halees 2009).

Density-based clustering filters

Density-based clustering is another form of document clustering that has been applied to spam classification. A claimed advantage is the ability to process hashed versions of messages, thus preserving user privacy. This method depends on having sensitive comparators. These comparators are usually either fast or sensitive. The challenge is to find comparators that are both sufficiently fast and sufficiently sensitive (Yoshida et al. 2004).

7.2.2.2.6 DT filters

Decision trees are a classification technique commonly used in data mining, where the interior nodes of the tree represent observations, and leaf nodes represent decisions or conclusions (Carreras & Marquez 2001). The decision tree technique can be used to detect email spam from the analysis of each email categorisation (sexual, financial and job-hunting, and marketing and advertising) to find the association rules of spam emails in the categorisation (Sheu 2009). Decision trees have many advantages. According to Zhao and Zhang (2008), decision trees are easy to understand, can be easily converted to a set of production rules, can classify both categorical and numerical data, and contain no a priori assumptions about the nature of the data.

Different types of decision tree techniques have been applied to spam classification. The C4.5 (J84) tree uses attribute values of the available training data to create a decision tree (El-Halees 2009; Youn & McLeod 2007a, 2007b, 2007c). The J84 tree is an open-source implementation of C4.5 decision tree (Sharma & Sahni 2011) that was used to create a binary tree (Youn & McLeod 2007a). The classification and regression tree (CART) is a data exploration and prediction algorithm (Abu-Nimeh et al. 2008). It was defined by Sharma and Sahni (2011) as “is a classification method which in order to construct decision trees uses historical data”.

The random forests (RF) tree was employed to produce filters to detect email spam (DeBarr & Wechsler 2009). Chaudhary, Kolhe and Kamal (2013) defined RF as a collection of pruned classification or regression trees induced from bootstrap samples of the training data using random feature selection in the tree induction process. Iterative Dichotomiser 3 (ID3), one of the most effective decision trees (Sheu 2009),was developed on the basis of the Concept Learning System (CLS). According to Sharma and Sahni (2011):

> … the Concept Learning System algorithm is the basis for the ID3 algorithm. By adding a feature selection heuristic ID3 improves on

CLS. The attributes of the training instances are searched through by ID3 and the attribute that best separates the given examples is extracted by it.

The alternating decision tree (ADTree) is a ML method used for classification (Sharma & Sahni 2011). The data structure and algorithm of ADTree are generalisations of decision tree and have connections to boosting (Zhao & Zhang 2008). Reduced Error Pruning Tree (REPTree) is a fast decision tree learner that uses information gain as the basis of splitting to create a decision tree (Chaudhary, Kolhe & Kamal 2013). It uses Reduced Error Pruning to prune the tree (Zhao & Zhang 2008). A random tree was defined by Zhao and Zhang (2008) as a tree drawn at random from a set of possible trees, which means that each tree in the set of trees has an equal chance of being sampled. A decision stump is a single-level decision tree where the split at the root level is based on a specific attribute or value pair. It includes only one internal node (the root), which is connected to terminal nodes (its leaves) (Chaudhary, Kolhe & Kamal 2013).

7.2.2.3 Fingerprinting filters

Finger printing techniques make use of a list of “finger prints” of known types of spam, by computing and comparing the finger print of any incoming email. Various schemes for generating finger prints are possible, for example via an exact or approximate digest (digest-based filters) (Damiani et al. 2004), or a hashing algorithm (signature schemes) (Attenberg et al. 2009; Garcia, Hoepman & Nieuwenhuizen 2004). In any case, lists of the resulting fingerprints are usually propagated to mail servers and any message received that has a matching fingerprint is assumed to be spam (Allman 2003). Particular challenges in fingerprint-based spam detection is making them reliable in the face of polymorphic spam and ensuring that a fingerprint does not disclose any content of a given message (Takesue 2009). Honeypots and zombie-based approaches are considered types of fingerprinting filters.

7.2.2.3.1 Honeypot filters

A common method of collecting known spam messages for a fingerprinting system is via a honeypot, which is a machine or system that exists solely to collect spam (Oudot 2003). Honeypots are also of value to researchers by identifying new species of spam as they emerge, as well as analysing email harvesting activity and detecting email relays (Andreolini et al. 2005). BrightMail is one type of honeypot method. BrightMail filters email addresses before placing them in the Post Office Protocol (POP) mailbox. It allows spammers to detect email addresses left on web pages, news groups or subscription to mailing lists and send spam email to these addresses (Allman 2003).

7.2.2.3.2 Zombie-based filters

Spammers can send their emails by spambots or zombie machines (Hayati & Potdar 2008). Many zombie machines often use non-standard optimisations to the SMTP protocol, which can be detected by the receiving SMTP server. Thus it is possible to classify some spam based on the content of the SMTP session (Lieven et al. 2007).

7.3 The Effectiveness of Anti-spam Filters in Detecting Email Spam

The previous section presented the taxonomy proposed in this study, and explained and discussed each method used in filtering email spam. This section discusses studies that have used reputation and content techniques, described above, to produce filters against email spam, and their effectiveness in detecting email spam.

7.3.1 The Effectiveness of Reputation-based Filters in Detecting Email Spam

Reputation-based methods rely on information outside of the content of email, such as the reputation of one or more of the participants, the IP source, and senders' email addresses, to classify messages as spam or legitimate (Cook et al. 2006; Garcia, Hoepman & Nieuwenhuizen 2004). This section describes previous proposals for email spam filters based on the reputation methods described in the previous section (taxonomy of email spam filters), and discusses their effectiveness in detecting email spam.

Boykin and Roychowdhury (2004) used the properties of social networks to develop an automated anti-spam tool to classify senders as spammers or non-spammers. This tool has two advantages: the first advantage is that it does not require user intervention or supervised training. The second advantage is that there are no false negatives or false positives when using this tool. The results showed that the anti-spam tool classified 53% of all emails as spam or legitimate, while 47% of emails were not classified. The authors recommended that "this method is extremely important in achieving accurate and automated spam filtering".

Golbeck and Hendler (2004) presented a method for spam filtering based on whitelists and social networks. The users assigned a "reputation" or "trust" score to known senders. This yielded a large reputation network containing many users. The method applied an algorithm to infer reputation relationships between users, which were used to score emails. The score value is used to sort messages in the users' inboxes. This method was highly accurate and any valid email from unknown senders connected within the social network could receive high scores. This method was complementary to other spam filters and did not replace other filtering systems. It helped other spam filters by identifying legitimate emails that can be undistinguishable from spam emails.

Cook et al. (2006) proposed a new spam detection technique based on the intrusion detection system (IDS). This technique used audit logs analysis to block spam emails before they entered the network, at the network gateway. Port scans were used as evidence against suspicious IP addresses that could send spam. Based on the port scan evidence, the network firewall blocked the SMTP connections from suspicious IP addresses. Domain specific dynamic blacklists (DSDBLs) were used to blacklist spammers' IP addresses, which could attack the domain. However, there were some disadvantages to this technique. The first disadvantage was that at least one spam per IP address enters the network during the port scan investigation. The second disadvantage was that the network is not protected if the host, which runs this technique, is shut down.

Tran and Armitage (2006) proposed an anti-spam tool based on the TCP-layer algorithm, which statistically accepts or rejects the inbound TCP connection using the past history of spam or spammers' IP addresses. This tool aimed to reduce the operational cost of creating and using blacklists for mail server operators. This tool has many advantages. First, the degree of human intervention required by the mail server and blacklist operators is reduced because of the automatic rehabilitation of legitimate senders. Second, it reduces the problem of misclassifying legitimate emails (FPs).

Li and Hsieh (2006) developed a group-based filter based on the number of members in the group and the number of groups that a spammer is associated with. This filter can block from 70% to 90% of email spam, depending on the implementation parameters. The authors reported that group-based anti-spam method may not be highly effective as a standalone approach as some groups may have only one member, but it can be used as a complementary tool for other existing anti-spam tools, such as SpamAssasin.

Xie, Yin and Wang (2006) produced a simple and effective system, called DBSpam, to detect and break proxy-based email spam activities inside a customer network and to trace the corresponding spam sources outside the network. This system leverages the protocol semantics and timing causality of proxy-based spamming to identify spam proxies and the real spam sources behind them. DBSpam can be tuned to detect spam proxies and sources with low false positives and false negatives and effectively block spam traffic.

Lam and Yeung (2007) proposed a method to filter spam based on extracting features from email social networks (directed graph) for senders. These features were in-count and out-count, in-degree and out-degree, communication reciprocity (CR), communication interaction average (CIA), and clustering coefficient (CC). The social networks included nodes and edges. The nodes represented senders and the edges represented email transactions. The proposed method assigned a legitimacy score for each sender and this score was used to analyse batches of logs. This method classified senders as legitimate if the legitimacy score was high. A database of scores was built which could be used by online mitigating methods to query the score of a particular sender. The performance of the classification was increased by increasing the weight of the CC, which indicated that the CC feature was better than other features. The results showed that using only three features (CR, CIA and CC) resulted in low accuracies.

Ramachandran, Feamster and Vempala (2007) have presented a spam-filtering system called SpamTracker which classified email senders based on their behaviour instead of their IPs. The SpamTracker system complements blacklists and depends on a new technique called behavioural blacklisting, which classifies senders based on sending patterns. This system uses clustering algorithms to categorise email senders by how they send spam. The design of the system is easy to replicate and distribute. The results revealed that this system distinguished legitimate emails from spam, and it detected spammers who were missed by other filtering methods or IP-based blacklists.

Lieven et al. (2007) proposed a spam-filtering approach based on the connection-oriented analysis of email source retry patterns. This approach was very effective in detecting spam at an early stage in the mail delivery process and avoided receiving a large volume of spam at the email server.

Yih, McCann and Kolcz (2007) developed four simple methods to detect grey email and compared their performance in detecting grey email by using recall-precision curves. Grey email was defined as email that could be considered as spam or legitimate (Chang, Yih & McCann 2008). The four proposed methods were as follows: leveraging the output of a spam filter, comparing an ensemble of spam filters, creating approximate data using sender IP information, and identifying email campaigns with mixed labels. The results revealed that identifying email campaigns with mixed labels was more reliable than the other three methods.

Polz and Gansterer (2009) proposed a new system architecture that included two classes of messages: trusted and untrusted emails. The trusted emails were signed with a secure/multipurpose internet mail extensions (S/MIME) signature (Dusse et al. 1998). The S/MIME signature is an encryption and signing standard for sending and receiving secure email messages. To avoid the usability issue that faced the S/MIME signature in other methods, the signature in this method was executed on the email server without any user interaction. However, a major disadvantage of the trustnet architecture was a potential security risk when the attacker got access to the architecture. The results revealed that the trustnet architecture reduced the processing time and increased the amount of data transferred when compared to other methods such as SpamAssassin/ClamAV.

Engelberth et al. (2009) have developed a method called Mail-Shake that depended on the use of public and private emails addresses. This method aimed to hide valid email addresses from being public. The mechanism of this method is as follows: senders email public email address and receive an automatic reply containing a challenge that requires them to answer it to get valid private email address that they can send to it. The challenge could be the Completely Automated Public Turing Test to Tell Computers and Humans Apart (CAPTCHA) (Von Ahn et al. 2003), picture or simple text, the purpose of which is to make sure that the private email address delivers to real humans. After senders reply to a challenge, they receive the private email address and a random identification number (ID), which is used to assure that the challenge is received, and then can add senders' email address to whitelists. This method has many advantages. First, the spammer's activities and the volume of email spam is reduced because of the length of time required obtain the valid private email addresses. Second, this method makes it harder for automated harvesting software from getting valid email addresses due to the challenge that requires human interaction to solve it.

Saraubon and Limthanmaphon (2009) proposed a fast and effective botnet detection filter to detect text and image spam. This filter has achieved an accuracy rate of 96% in filtering text and image spam, with no false positive.

Esquivel, Akella and Mori (2010) divided SMTP senders into three main categories: legitimate servers, end-hosts, and spam gangs and then proposed a filter that builds custom IP reputation lists for each category, significantly improving the performance of existing IP reputation lists. The proposed filters can construct IP reputation lists that can cover 90% of all spam and legitimate, but the lists of spam gangs must be updated constantly to maintain accuracy.

7.3.2 The Effectiveness of Content-based Filters in Detecting Email Spam

Content-based methods depend on the content of emails to detect spam (Cook et al. 2006). These methods can classify email spam based on the content included in the headers or bodies of spam messages (Goweder, Rashed & Alhamammi 2008; Zhang, Zhu & Yao 2004). In the research literature, most of content-based methods have been proposed to detect email spam mostly in English, although a few methods have been developed to detect it in other languages. This section reviews previous studies proposed to develop email spam filters based on the content-based techniques provided in the previous section (the taxonomy of email spam filters), and discusses their effectiveness in detecting email spam in different languages.

Pantel and Lin (1998) have presented a spam-filtering system called SpamCop. The authors used NB algorithms in this system to classify emails as legitimate or spam. The results showed that the SpamCop is very accurate, as the system detected about 92% of spam, with only 1.16% false positives. The results also showed the system to be more accurate in detecting spam than other methods, such as RIPPER and keyword-detection methods (Cohen 1996).

Provost (1999) conducted three experiments to compare the performance of the NB algorithm against the RIPPER rule-based algorithm (Drucker, Wu & Vapnik 1999) on the following tasks: learning a user's mail-sorting preferences (hand-sorted mail), reconstructing the policy of automated sorting (automatically-sorted mail) and detecting spam. The results showed the NB algorithm to be more effective at classifying email spam than the RIPPER rule-based algorithm.

Androutsopoulos et al. (2000) have compared the performance of Naïve Bayesian to the performance of keyword patterns in the context of cost-sensitive evaluation measures. This study was conducted on an English email spam corpus. To select the appropriate attributes, spam recall and spam precision measures as well as mutual information were computed for each attribute in messages. The accuracy and error rates were calculated to measure the performance of different methods. The results revealed that the NB method outperformed the keywords patterns method in detecting email spam.

Based on the methodology and cost-sensitive evaluation measures used in the previous study, Androutsopoulos et al. (2000) conducted another study to compare the performance of two ML methods: Naïve Bayesian and MBL. The results indicated that the classification accuracy for both NB and MBL approaches was very high, and they outperformed the keywords patterns method.

Carreras and Marquez (2001) presented a spam filter based on boosting trees and compared its performance with other filters, such as NB and DT. This study was conducted on an English email spam corpus, and found that the performance of the boosting based filters in detecting spam was better than the performance of Naïve Bayesian and DT.

O'Brien and Vogel (2003) have presented a method for English spam identification based on authorship identification (the chi by degrees of freedom or $\frac{x}{\text{d.f.}}$) and compared it with the Bayesian method. The results showed that both methods were effective at filtering emails. The Bayesian method was effective at the character level tokenisation and this performance was reduced at word level. The chi by degrees of freedom was effective at the word level and the performance was reduced at the character level.

Sakkis et al. (2003) presented an empirical evaluation of MBL in filtering spam for mailing lists. This study investigated different attributes and distance weighting schemes, and the effect of the neighbourhood size, the size of the attribute and the training corpus. It identified three different cost scenarios and used the appropriate cost-sensitive evaluation. The memory-based method was shown to be feasible for combating email spam when combined with safety nets such as filters embedded in mailing lists. This method was more effective than the NB filter in detecting false positive (i.e., legitimate messages were mistakenly classified as spam).

Clark et al. (2003b) developed a new system to classify English email spam, called LINGER, which was based on a neural network. The effects of feature selection, such as Information Gain (IG) and Variance (V), were investigated. The results revealed that LINGER was better than NB, KNN, keywords, DT and boosting methods in classifying emails. The feature V was more effective than IG in a multiclass task, and LINGER-IG obtained a perfect performance in binary spam mail filtering, superior to LINGER-V.

Zhang, Zhu and Yao (2004) provided a comprehensive evaluation of five supervised ML methods in the context of cost-sensitive spam-filtering: NB, SVM, Boosting (AdaBoost), MEM and MBL. This study was conducted on English and Chinese email spam corpora. Document frequency (DF), IG and chi-square were used to select features. Spam precision and spam recall were computed to measure the performance of methods. Weighted accuracy (W_{ACC}) was calculated to evaluate the accuracy of the system for false positives and false negatives. The results showed that the SVM, boosting and MEM classifiers gave top performances in filtering email spam, and that the email header information was as reliable and effective in filtering spam as the email body.

Garcia, Hoepman and Nieuwenhuizen (2004) investigated the effectiveness of several spam detection methods such as Distributed Checksum Clearinghouse (DCC) (Haiyan, Runsheng & Yi 2009), genetic algorithms (Goweder, Rashed & Alhamammi 2008), and NB such as Bogofilter (Raymond 2005) and BMF. The results revealed that the genetic algorithm was the best way to filter spam at the ISP or server level; while NB was the most effective way to filtering spam at the user level.

Yoshida et al. (2004) proposed a new method of blocking Japanese and English email spam. The proposed method depended on the analysis of the document space density information to detect spam. In designing this method the authors considered the following three characteristics: ease of maintenance, high accuracy and privacy protection. The results showed that the method achieved a high processing speed and that it was suitable for use by ISPs to support a large mail server using small computers. However, this method was not useful for client terminals because it needed extensive volumes of email traffic to calculate the information density. This method automatically updated data and unlike other traditional spam filters that need a lot of operator time to update the data, did not require human maintenance. This method achieved 98% recall and 100% precision and were better than the rates of recall and precision of some ML methods, such as SVM, NB, C4.5 and ANN.

Özgür, Güngör and Gürgen (2004) presented an anti-spam method to detect Turkish email spam, based on ANN and Bayesian networks algorithms. The proposed method contained two modules: Morphology Module (MM) and Learning Module (LM). MM was used for the morphology of the word and LM was used to classify emails by the root of the word, which was executed in the morphological analysis. Two ANN structures were used: single layer perceptron and multi-layer perceptron. Three different approaches of the Bayesian networks were employed: binary, probabilistic and advanced probabilistic. The results showed a 90% success rate for this method in filtering Turkish spam.

Chuan et al. (2005) proposed a spam filter based on LVQ neural networks and compared its performance with two methods: Bayes methods and back propagation (BP). They used Term Frequency-Inverse Document Frequency (TF-IDF) to obtain feature weight, and calculated MI to select the appropriate features. The LVQ network model contained two layers. The first layer was the competitive layer, and the second was the output layer. The corpus used for the experiments included English emails, HTML tags, and message headers (except the subject line, which was removed). The results showed that the performance of a spam filter based on the LVQ network was better than that of one based on Bayes and various other forms of ANNs, such as BP.

Metsis, Androutsopoulos and Paliouras (2006) evaluated the performance of five different versions of NB in filtering English email spam. These methods included Multivariate Bernoulli Naïve Bayes (MV Bernoulli NB), Multinomial Term Frequency Naïve Bayes (MN TF NB), MN Boolean NB, MV Gauss NB and Flexible Bayes (FB) (Almeida & Yamakami 2010). The experiments were conducted on six non-encoded datasets called Enron (Klimt & Yang 2004). The experiments found that the spam-filtering performance of the FB and the MN with Boolean attributes was better than the performance of other versions of the NB method.

Krasser et al. (2007) developed a filter based on the C4.5 decision tree and SVMs to detect image spam. The results showed that this filter blocked about 60% of image spam.

Youn and McLeod (2007a) tested four different classifiers, ANN, SVM, NB and J48, and compared their performance in detecting spam. The results showed that J48 and NB classifiers were better than ANN and SVM in detecting spam.

Goweder et al. (2008) proposed an email spam filter based on multi-layer ANN as a classifier and a genetic algorithm (GA) as a training algorithm. They used advanced techniques of GA to train the multi-layer perceptron (MLP) to filter Arabic and English email spam. The Subject, From and Body fields in email headers were selected for the experiment and other irrelevant parts of email were deleted. The results of this study clearly indicate that there is enough information in the Subject and Body fields to classify email as spam or non-spam. The spam filter that they proposed achieved an accuracy of about 94% in detecting spam emails and 89% in detecting non-spam. On the other hand, it took a long time to train the MLP by using GA.

Kim and Hwang (2008) have proposed a detection method based on NB ML to detect English email spam. This method aimed to distinguish between spammers and non-spammers through their posing of information such as dates, URLs, tags, and descriptions. This study used the number of bookmark and bibtex postings and tags in the classification as feature variables. The mutual information between a tag and the target was calculated. The results revealed that the proposed method was better than mutual information-based methods in detecting spam.

Abu-Nimeh et al. (2008) proposed a method of classify English email spam using Bayesian additive regression trees (BART). Because of the problem of binary classification in the original form of BART, the authors modified BART to be suitable for Classification (CBART). Then they compared the performance of the CBART with six classification methods: logistic regression classifier (LRC), SVM, CART, ANN, RFs and NB. The results showed that the performance of the CBART in classifying spam was better than the performance of the six other methods.

El-Halees (2009) presented a comparison of six supervised ML spam filters in filtering spam for Arabic, English, and mixed (include Arabic and English texts) emails: MEMs, DTs, ANNs, NB, SVM, and K-NN. To achieve this, a system was built to filter spam. This system contains two subsystems: training and testing. To evaluate the performance of the system, SR and SP were used. The W_{ACC} was computed to evaluate the accuracy of the system for false positives and false negatives. The results showed that all classifiers performed much better in detecting English email spam than Arabic email spam. The author suggested that this might be because Arabic is a highly inflected language. The best filter for English spam was the SVM, while the ANN was the most effective filter for Arabic spam before stemming words. After stemming words, MEM and NB was better than other filters in detecting Arabic spam. The results found that the performance of most methods to detect Arabic spam was improved by stemming words.

Abdoh, Musa and Salman (2009) developed a new automated spam filter to detect English email spam based on the NB method. This filter used Bayesian rules to calculate the probability of spam words and use these probabilities to find a weight for the word based on its frequency in both spam and legitimate emails. The results showed that this filter achieved an accuracy rate of 95% in detecting English spam. By comparing the performance of this filter with similar methods, such as the Quick

Spam Filter and SPATIC (FreeCode 2013), this filter performed better than the other two methods.

Wang et al. (2010) proposed a filter based on the SVM method to detect English image spam. The experiments it to be effective for detecting English image spam.

Ergina et al. (2011) proposed a filter to detect Turkish email spam based on two models of Bayesian method: binary and probabilistic. The filter achieved a success rate of 89% for probabilistic Bayesian model, while the binary model achieved a success rate of 93% in detecting Turkish email spam.

It can be seen that many reputation- and content-based filters have been proposed to detect email spam in different languages, mostly in English. The proposed reputation-based filters were not completely effective in detecting email spam, and previous studies have suggested that they may not be highly effective as a standalone method, but can be used as a complementary tool for other existing content-based filters (Li & Hsieh 2006). The content-based filters, based on different languages, had achieved better performance in detecting email spam. These filters can detect email spam based on the content involved in the headers and bodies of email spam, and their effectiveness was higher in detecting English email spam than other non-English emails spam (Çıltık & Güngör 2008; El-Halees 2009; Nguyen, Tran & Nguyen 2008).

Of the content-based filters, ML filters have the best performance in detecting email spam (rule and fingerprinting methods) (El-Halees 2009; Zhang, Zhu & Yao 2004). El-Halees (2009) found SVM to be the more effective filter (depending on the content) for detecting English email spam, although maximum entropy and NB had were more effective in detecting Arabic email spam than other ML methods. Garcia, Hoepman and Nieuwenhuizen (2004) found genetic methods to be the best filter for blocking email spam at the ISP or server level, and NB more effective for filtering email spam at the user level.

In spite of the large number of email spam filters that have been developed to detect email spam and their high level of effectiveness, especially for English spam, only a few of these filters have been designed to filter Arabic email spam. Therefore, it is likely that filters that have achieved a high performance and accuracy in detecting

English email spam can be used to develop filters to combat Arabic email spam and the Arabic spammers' tricks described in the next chapter (Chapter 8).

7.4 Conclusions

This chapter described and discussed the taxonomy of email spam filters proposed in this study. The phenetics method or numerical taxonomy was used to classify objects based on their similarities (Nickerson, Varshney & Muntermann 2013). Consequently, this study employed the phenetics approach to develop a taxonomy of email spam filters that classifies anti-spam filters on the basis of the similarity between the methods filters use to detect email spam. The taxonomy requirements and the development of constructs of the email spam filter taxonomy were discussed in this chapter.

The proposed taxonomy was classified into two main techniques: reputation and content-based techniques. Reputation-based filters rely on information outside of the content of email messages to detect spam, whereas content-based filters depend on the content of emails to classify them as spam or non-spam (Cook et al. 2006; Garcia, Hoepman & Nieuwenhuizen 2004). These two major techniques have been classified by other researchers into subgroups and these subgroups were organised into different types. All of these techniques were explained and discussed in this chapter. However, the proposed taxonomy is not exhaustive, and a clear task could be to expand it with information about additional email spam filters and techniques.

The effectiveness of reputation and content-based filters, presented in the taxonomy, was provided and discussed. Neither reputation nor content-based filters were completely effective in detecting email spam. When installed alone, the performance of reputation-based filters in detecting email spam was not high, so they are best used as a complementary tool for other existing content-based filters (Li & Hsieh 2006). The effectiveness of content-based filters differed from one language to another. These filters perform better at detecting English email spam than non-English spam (Çıltık & Güngör 2008; El-Halees 2009; Nguyen, Tran & Nguyen 2008). Machine-learning methods had better performance in detecting email spam than other content-based filters such as rule and fingerprinting-based filters. Compared to other content-based filters, the best method for detecting English email spam was SVM. Compared with other ML methods, maximum entropy and NB were the most effective filters for detecting Arabic email spam.

As described and discussed in Chapter 6 (ISPs' results), Saudi ISPs found that anti-spam filters were not completely effective in detecting Arabic and English email spam, and these filters performed better in detecting English spam than Arabic spam. Previous studies have claimed that, as spammers continuously develop new tricks or methods to bypass these filters, this can reduce the effectiveness of anti-spam filters (Hayati & Potdar 2009; Wang et al. 2007). This encouraged the researcher to investigate the different tricks used by spammers in Arabic and English email spam. Understanding spammers' tricks can help developers to refine existing filters to be more effective in detecting email spam, especially Arabic email spam.

In the research literature, many filters based on different methods (e.g. reputation or content) have been proposed to detect email spam, mostly in English. This study clustered most of these filters into a taxonomy, which was presented in this chapter. It is hoped that this taxonomy can help in the selection of appropriate filters for the spammers' tricks observed in the headers and bodies of Arabic email spam. These tricks will be presented in the next chapter. This taxonomy could also help improve current filters or create new filters, especially for Arabic email spam. The proposed taxonomy indicated that many methods have been proposed to detect email spam, mostly in English, based on the content of header and body of the email, and they achieve a high level of effectiveness and accuracy in detecting English email spam. These methods include SVMs, boosting, maximum entropy (Zhang, Zhu & Yao 2004), LVQ (Chuan et al. 2005), and decision tree (C4.5) (Krasser et al. 2007).These methods can be used to develop effective filters against tricks included in the header and body of Arabic email spam. The implementation of such filters targeting Arabic spam should be pursued.

Chapter 8: Analysis of the Headers and Bodies of Arabic, English and Mixed Email Spam

Spammers have different reasons for sending email spam. Some of these reasons are malicious (e.g. phishing and viruses) and some of them are for the purpose of commercial businesses, such as product advertisements (Hayati & Potdar 2008). This necessitated the development of filters to combat email spam and the tricks used in sending it. However, anti-spam filters are not effective in detecting all email spam as spammers are constantly developing their methods and tricks to bypass the filters (Hayati & Potdar 2009; Wang et al. 2007). As reported in Chapter 6 (ISPs' results), Saudi ISPs found that the anti-spam filters used were not completely effective in detecting English and Arabic spam, and these filters were more effective in detecting English spam than Arabic spam. This result implies that further investigation of the tricks spammers used in Arabic and English email spam is needed to be able to develop more effective filters, especially for Arabic spam. From the findings of the taxonomy presented in Chapter 7, appropriate filters to identify Arabic spam have been suggested.

This chapter analyses the headers and bodies of a collection of Arabic, English and mixed language (Arabic and English) email spam received from Saudi public users, businesses and ISPs to identify the tricks spammers used to bypass filters. Spam in these languages was investigated because the researcher understands both Arabic and English, which assists the analysis process. Arabic is relevant as it is the official language of Saudi Arabia; English is a good comparison language because both spam and filters in English are advanced. The methodology followed in the analysis of Arabic, English and mixed email spam corpora is described in detail in Chapter 3 under Section 3.10.

This chapter is divided into the following sections:

- Section 8.1: presents the results of the analysis of the spammers' tricks used to bypass anti-spam filters.
- Section 8.2: discusses the results of the analysis of the spammers' tricks used in sending Arabic, English and mixed language email spam.

- Section 8.3: presents the conclusions of this chapter.

8.1 Results

This section describes the results of the analysis of headers and bodies of Arabic, English and mixed language email spam corpora, to determine spammers' tricks used to bypass anti-spam filters and to trick recipients. The chi-square test (X^2) was used in this chapter to analyse the data. It was used to test the category data relating to independent variables (spammers' tricks).

The data in Table 8.1 reveal that attractive words used in the subject line of emails appeared more often in Arabic email spam than in English and mixed language email spam (45%, p=0.006), whereas English email spam included more false statements in the subject lines than Arabic and mixed language email spam did (72%, p=0.006).

Compared with Arabic and mixed language email spam, most of the content of English email spam (92%) appeared as text (p<0.001), whereas the percentage of email spam that appeared as text embedded in images was larger in Arabic than in English and mixed language email spam (45%, p<0.001).

The percentage of links included in email spam was greater in English than in Arabic and mixed language email spam (84%, p=0.007). The results revealed that some links included in email spam were malicious, and more so in English email spam than Arabic and mixed language email spam (73%, p<0.001). Two types of malicious links were found in the content of all email spam. There were more links related to fake bank websites in English email spam than in Arabic and mixed language email spam (88%, p<0.001), and more false or forged unsubscribe links in Arabic email spam than in English and mixed language email spam (72%, p<0.001).

Arabic email spam contained more attachments than English and mixed email spam (28%, p=0.007), and the types of attachments varied in the three groups. Arabic email spam had more PDF files than English and mixed email spam (17%, p<0.001), and exe files were more common in English than in Arabic and mixed language email spam (75%, p<0.001). Text (txt) files and images in different formats such as GIF and JPEG were found more often in mixed language email spam (50%, p<0.001). The results indicated that some attachments were malicious, and English email spam included more malicious attachments than Arabic and mixed language email spam (89%, p<0.001).

The percentage of spam sent from fake or obfuscated email addresses was higher in Arabic and mixed language email spam than in English email spam (100% and 99% respectively, p<0.001).

Table 8.1: The percentages of spammers' tricks used in Arabic, English and mixed language email spam

Spammers' Tricks	Arabic (%) n=1035	English (%) n=179	Mixed (%) n=56	P*
Using attractive words or false statements in the subject line				
Attractive words	45	28	38	**0.006**
False statements	55	72	62	
Using different formats in writing content				
Text	55	92	61	**<0.001**
Text embedded in an image	45	8	39	
Adding links or attachment into the content				
Links	72	84	79	**0.007**
Attachments	28	16	21	
Types of attachments				
Images (e.g. GIF and JPEG)	45	14	50	**<0.001**
Pdf files	17	4	0	
Text (txt) files	20	7	50	
Executable (exe) files	8	75	0	
The percentages of malicious links and attachments				
Malicious links (%YES)	4	73	0	**<0.001**
Malicious attachments (%YES)	15	89	0	**<0.001**
Types of malicious links				
Fake bank's website link	28	88	0	**<0.001**
Forged unsubscribe link	72	12	0	
Using fake or obfuscated email addresses (%YES)	99	58	100	**<0.001**

*P values are based on chi-square test between types of Arabic, English and mixed language email spam; P values <0.05 were considered statistically significant.

8.2 Discussion

Many tricks have been used by spammers to achieve their objectives. This section discusses the results of the analysis of spammers' tricks observed in an Arabic, English and mixed language email spam corpora received in Saudi Arabia in this study. These tricks included using attractive words or false statements in the subject line, using different formats in writing the content, adding links or attachments into the content, and using fake or obfuscated email addresses.

8.2.1 Using Attractive Words or False Statements in the Subject Line of Email Spam

The results of the analysis indicated that spammers used attractive words more in Arabic email spam than in English and mixed language email spam. Words and phrases used in the subject lines of Arabic email spam were focused on entertainment advertisements, while business advertisements, and phishing and fraud words were more common in the subject lines of English email spam. Examples of the words and phrases observed in the subject lines of Arabic and English email spam are attached in the Appendix H. This finding is similar to the finding of a study conducted on an English and Japanese email spam corpora by Yamakawa and Yoshiura (2010), which found that most words used in English spam were related to commercial and business advertisements, whereas sexuality related words were the most used in Japanese email spam. The findings of both studies indicated that there were differences between words and phrases used in Japanese and Arabic email spam. Japanese spam words were related to sexuality, whereas Arabic email spam words focused on entertainment advertisements. The reason for this could be because of cultural differences between the two countries. According to Abdoh, Musa and Salman (2009), the nature of email spam is different from one country to another with the spammers' motivations and cultures. Some spammers send commercial emails, some send pornographic emails, and others send malicious programs.

Previous studies have indicated that the reason for using attractive words in the subject line of email spam is to convince the recipients to open the email, or to make them think that it is legitimate and they should read it (Attar, Rad & Atani 2013; Dhinakaran, Jae Kwang & Nagamalai 2009). The authors reported examples of words or phrases in the subject line that are designed to lure the victim, such as "account confirmation", "message from the bank", "security warning", and "update details" (Dhinakaran, Jae Kwang & Nagamalai 2009). Smith (2008) stated that some emails used subject lines such as a fake or real news event, inexpensive products, or easy ways to make money, to encourage the recipients to open the emails and the attachments. Wang and Chen (2007) noted that "sex", "for sale", "get rich" and "best deal" were the keywords most frequently used in the subject of email spam. "Hi", "Hello", "Was this from you?", "Alert", or "Thank you" are common words and phrases in subject lines (Wei et al. 2008). "Re" is another example of a frequently used attractive word (Chen, Zhan & Li 2010), implying that the spammer is answering an email from the recipient, or replying to a request from the recipient. This tricks recipients into thinking that the email is important, encouraging them to open it and read the content (Chigona et al. 2005).

Although a few methods have been developed to detect Arabic email spam based on the header (El-Halees 2009; Goweder et al. 2008), these methods are still not effective (Hayati & Potdar 2009). This suggests the need to develop anti-spam filters to block Arabic spam, and this could be achieved by creating a list of the common words observed in the subject line of Arabic email spam (entertainment advertisements words, as observed in this study) and producing filters to block spam emails that contain these words in the subject line. Anti-spam filters can block spam by scanning the words in the email subject line and body. Christina, Karpagavalli and Suganya (2010) proposed adding keywords observed in the subject line and body of email spam to the lists of one of the existing anti-spam filters. They suggested that "using combinations of keywords is a good solution to enhance filtering efficiency".

The previous chapter, which presented the taxonomy of email spam filters and the effectiveness of these filters in detecting email spam, described several methods that have been used to develop filters based on the content of the header, most of which were developed to detect English spam. These methods included support vector machines, boosting, maximum entropy (Zhang, Zhu & Yao 2004), and learning vector quantization (LVQ) (Chuan et al. 2005). The taxonomy developed in the researcher's study of email spam filters found that these methods were more effective than other methods in detecting English email spam. It would appear that these methods could be also used to produce more effective filters against Arabic email spam. Developing filters based on the combination of reputation and content focused methods could be another way to combat Arabic email spam (Li & Hsieh 2006).

Another trick used by spammers to lure the victims is to send emails with subject line that do not indicate the content of the email, called false statements (or misleading subject lines) (Hamel 2004; Simon 2004). For example, spammers added greetings or thank words or phrases in the subject lines while the content included phishing attachments or product advertisements. This could make it difficult for the recipients to determine the content of the email before they open it (Chigona et al. 2005). The results of this study revealed that most of email spam received in Saudi Arabia contained false statements in the subject lines of email spam, and the percentage of false statements was higher in English email spam than Arabic and mixed language email spam. A previous study conducted by the US Federal Trade Commission (FTC) (2003) revealed that 40% of the subjects of spam emails sent to American recipients did not indicate emails' content. On comparing the results of both studies, it was found that the percentages of false statement or misleading subject lines was higher in the email spam received in Saudi Arabia than in the US.

8.2.2 Using Different Formats in Writing the Content of Email Spam (Text or Text Embedded in an Image)

The percentage of email spam that appeared as text was higher in English than in Arabic and mixed language spam. This may be because many of the English email spam, as described in the results section, contained more malicious content, such as phishing and scams, than the Arabic and mixed language spam. Previous studies have indicated that malicious email spam mostly appeared as text. According to Dhinakaran, Lee and Nagamalai (2007a), "email spam which contains only text messages are mostly related to scam". As attackers or phishers constantly develop their methods, it is necessary to constantly develop and improve effectiveness of existing Internet security software in detecting malicious embedded contents of email spam. This is particularly true for English, as studies such as that by Ramanathan and Wechsler (2012) have found phishing to be more common in English than in non-English spam.

Email spam or image spam, which contains text embedded in an image (Soranamageswari & Meena 2010; Xu, Wang & Shao 2009), was found more often in Arabic spam than in English and mixed language email spam. Previous research has suggested possible reasons. Attar, Rad and Atani (2013) stated:

> Image spam is a new threat which is the most sophisticated kind of spam email up to now, because it makes the message

> interesting for the user and hard to detect by text based anti-spam filters.

Image spam was created to circumvent the anti-spam filters that classify spam based on texts included in the body of messages (Nielson, Aycock & de Castro 2008; Zuo et al. 2009). Gargiulo and Sansone (2008) claimed that image spam is a new trick used to attract users without detection by text-based filters.

Previous studies have found that the volume of image spam has increased in the past few years. Image spam first appeared in 2004 (Kelly 2007), its volume reaching 1% of all email spam around the world in late 2005 (Soranamageswari & Meena 2010) and growing to 55% of all email spam in 2010 (Attar, Rad & Atani 2013). There are two possible explanations for the increase of the volume of image spam. First, that spammers found image spam an attractive way to lure recipients and a more effective way to achieve their purposes than traditional containing only texts (Gargiulo & Sansone 2008; Yamakawa & Yoshiura 2010). Then they develop methods to bypass text-based filters (Wittel & Wu 2004). Second, the existing anti-spam filters may not be as effective in detecting image spam, as detecting texts embedded in images is more complicated than detecting texts included in the body of messages. According to Mehta et al. (2008), filtering images included in emails spam is more complicated than filtering texts, as images have different complex data formats that require a deep, comprehensive analysis of content to identify the images' properties. Yamakawa and Yoshiura (2010) claimed that:

> … spam filter has to recognize image in mail in order to detect image spam mail; however, many spam filters do not have such a function or it makes heavy load that spam filter checks images in emails.

It can be concluded that the volume of image spam has increased significantly worldwide (Attar, Rad & Atani 2013), especially in Saudi Arabia, where the results of this study found that the percentage of image spam was nearly similar to its percentage worldwide, although the percentage of email that it was higher in Arabic spam than English and mixed language spam. As described in the taxonomy proposed in this study in Chapter 7, many image spam filters were developed as a result of previous studies such as Krasser et al. (2007), Saraubon and Limthanmaphon (2009) and Wang et al. (2010). The methods used to produce these filters, support vector machines and decision tree (C4.5), achieved a high degree of effectiveness and accuracy in detecting English image spam compared to other methods. Therefore, these methods could be used again to create new effective filters against image spam in Arabic emails.

8.2.3 Adding Links or Attachments into the Content of Email Spam

The research for this thesis found that most email spam received in Saudi Arabia included links, and English spam had more links than Arabic and mixed language spam. This might be to evade the text-based anti-spam filters. These links direct users to webpages that promote products or commercial services, and include commercial advertisements for products, as text in the body is easy for text filters to detect (Attar, Rad & Atani 2013). Previous studies have indicated that email spam can include different forms of links, such as URLs; clickable links to social websites such as Facebook and YouTube; clickable links to spammers' targets, such as fake bank webpages, counterfeit business websites; and forged unsubscribe links. Kumar (2009) stated that spammers used social network websites such as Facebook to trick users and their friends for the purpose of obtaining personal information. Users' email addresses are then spammed, and this can help spread malicious programs, such as malware and worms, to their computers. Smith (2008) indicated that some attractive links, such as fake YouTube links, have been used to download malware onto users' computers when clicked. Leavitt (2005) and Barroso (2007) described spoofed links that open fake webpages of banks, or counterfeit websites of popular businesses, to steal important user information such as credit card details.

The results of this study revealed that some links in email spam received in Saudi Arabia were malicious, and more common in English spam than in Arabic and mixed language spam. A study conducted by John et al. (2009) found that 1% of 100,000 links included in email spam received in the US were used to run or download malicious executables and programs. This would indicate that the percentage of malicious links involved in email spam received in Saudi Arabia was higher than that received in the US. Links related to fake bank websites were the most malicious links observed in English email spam, and more common in that language than observed in Arabic and mixed language email spam. This finding supports the findings of studies conducted by Hinde (2003), IBM X-Force® (2011), and Ramanathan and Wechsler (2012), which indicated that financial institutions or organisations such as banks were the sector targeted most often by attackers or phishers. This finding agrees with the results of a study conducted by Cova, Kruegel and Vigna (2008) in the US, which revealed that most malicious links included in email spam were related to banks and auction sites. Although other researchers such as Fette, Sadeh and Tomasic (2006), Basnet, Mukkamala and Sung (2008) and John et al. (2009) proposed different systems to detect malicious links contained in email spam, more development is required, particularly to detect malicious links in English spam linked to fake bank websites.

The results of this study indicated that Arabic email spam received in Saudi Arabia included more forged unsubscribe links than English and mixed language email spam. The unsubscribe link is an option that enables users to remove their email addresses from mailing lists that they have subscribed to, or to unsubscribe from receiving more emails in the future (Allman 2003; Malcolm 2004; Vaile 2004). However, spammers have exploited the unsubscribe link by adding a so-called false or spoofed unsubscribe link into email spam (McCusker 2004). Previous studies mentioned possible reasons spammers use false or spoofed unsubscribe links. Clicking onto the false unsubscribe link indicate to spammers that the email address is valid, and trigger more spam (Chigona et al. 2005; Lambert 2003; Simpson 2003). It can be a way to add the victims' addresses to spammers' mailing lists, increasing the volume of spam received in the future. According to Allman (2003), "the unsubscribe link removes you from the list in question, but it also adds your address to another list". Andaker et al. (2006) stated that the spammers collecting the addresses can further annoy the recipients by distributing their email addresses to other spammers. False unsubscribe links can be added to open advertisements for businesses and products (Andaker et al. 2006; Lambert 2003). Some spammers add a deceptive or inoperative unsubscribe link to spam emails to evade the strict spam laws in some countries such as the US and South Africa.

Spammers use different types of attachments to advertise products or services, such as images or pdf files (Dhinakaran, Lee & Nagamalai 2007a), evading text-based filters (Attar, Rad & Atani 2013). Dhinakaran, Lee and Nagamalai (2007b) reported that when sending email spam with attachments, spammers use sophisticated tools that have not been used to send email spam without attachments. The sophisticated software can hide the sender's identity, select text messages randomly, identify open-relay machines, have mass-mailing capability, and define the spamming time and duration.

The study found that Arabic email spam received in Saudi Arabia included more attachments than English and mixed language email spam. These included images, PDF files, text files and executable files. Images were the most common type of attachment in Arabic email spam. This finding is in line with a study of worldwide email spam traffic by Dhinakaran, Lee and Nagamalai (2007b). The authors analysed the contents of 400,000 email spam collected from worldwide spam traffic in a period of 14 months by setting up a spam trap. A spam trap (also called a spam honeypot) is a decoy email address that is used for the purpose of collecting email spam (Boneh 2004; Pallas & Patrikakis 2005). The results of their study revealed that more than 50% of the email spam collection included attachments in the form of images and executable files, and most of the images were in the format of GIF and JPEG files. However, this finding differs from the results of a study conducted by Yamakawa and Yoshiura (2010) on a collection of Japanese email spam. The authors found that different attachments were included in Japanese email spam (e.g. PDF files, compressed files, Microsoft Word documents, Microsoft PowerPoint slides, Microsoft Excel sheets, binary data), and that most of the attachments contained in Japanese email spam were PDF files. It can be concluded that attachments in Arabic emails spam were similar to those that appeared in email spam worldwide, but different from those in Japanese email spam.

The attachments could be malicious to achieve the purposes of the spammers. Previous studies indicated that malicious attachments may be an important way for spammers to achieve malicious aims such as infection of users' computers by malicious programs such as viruses or malware (Cournane & Hunt 2004; Dantin & Paynter 2005; Hershkop & Stolfo 2004; Moustakas, Ranganathan & Duquenoy 2005; Pallas & Patrikakis 2005; Pfleeger & Bloom 2005; Sorkin 2001). One type of the dangerous attachments included in email spam is executable (exe) files. The exe files might be a way to transfer malicious programs such as worms and viruses. According to Nagamalai, Dhinakaran and Lee (2010), email spam with the attached (exe) files are mostly malicious programs such as viruses, worms and trojans. The results of this study found that English email spam received in Saudi Arabia had more executable (exe) files than Arabic and mixed language emails spam. This could make English email spam being more dangerous than Arabic and mixed language email spam, as the results of this study found that English email spam had more malicious attachments (89% of the total of English spam that contains attachments) than Arabic and mixed language email spam. A study conducted in Australia by Alazab et al. (2013) found that about 44% of emails spam included malicious attachments.

It can be concluded that the percentage of malicious attachments was higher in email spam received in Saudi Arabia than in Australia. Also, it can be observed that English email spam contained more malicious attachments than non-English email spam, such as Arabic. This is supported by Ramanathan and Wechsler (2012), who found that malicious attempts such as phishing, viruses, worms or trojans are more frequent in English than in non-English languages. In spite of several methods being developed by other researchers such as Kartaltepe and Xu (2006), Stolfo et al. (2010) and Amin (2011) to detect malicious attachments in email spam, these methods need to be further improved, particularly for detecting English spam. As the extent of malicious attachments in Arabic email spam was lower, it is likely to increase, which indicates the need to develop methods to block malicious attachments embedded in Arabic spam.

8.2.4 Using Fake or Obfuscated Email Addresses to Hide Spammers' Identities

The percentage of fake or obfuscated email addresses was higher in mixed language and Arabic email spam than in English email spam. This could hide spammers' identities, bypass anti-spam filters, and trick the recipients (Hayati & Potdar 2009). A previous study conducted by Dhinakaran, Jae Kwant and Nagamalai (2009) identified many characteristics in fake or obfuscated email addresses. The authors found that the length of the spammers' email addresses is more than 21 characters and up to 28 characters. The fake email address has three parts before the "@" symbol: (Word1)(numericvalue)(word1)@forged domain.com. Word 1 included sensitive words such as customer service, support, operator, service-number, operator-id, Client service, ref, reference number, and customers. The length of the second part (numericvalue) ranged from 5 to 12 characters, and the length of third part was only two characters. Examples of fake or obfuscated email addresses can be seen in Chapter 3 (Figure 3.2).

Obfuscating or generating fake email addresses of spam senders can be achieved in many ways. The first is by using spam software. According to Dhinakaran, Jae Kwant and Nagamalai (2009), the spam software can hide, obfuscate, or generate fake email addresses with different formats so that they appear similar to the legitimate email address. Examples of software used by spammers to produce fake email addresses were Phasma Email Spoofer, Bulk Mailer, Aneima 2.0, Avalanche 3.5, and Euthanasia (Dhinakaran, Lee & Nagamalai 2007a). Based on this, spam software is used more often to produce fake or obfuscated email addresses in mixed language and Arabic spam than in English spam.

Spammers can also hide their identities by using spoofed IP addresses (Nagamalai, D, Dhinakaran, BC & Lee, JK 2010). Hu and Mao (2007) stated that spammers sometimes use spoofed IP addresses to conduct malicious activities such as spamming and DoS attacks without worrying about revealing their identities. Li and Hsieh (2006) claimed that “the spammer can use unused IP addresses on the same Local Area Network (LAN) to spoof its source IP address”. A study conducted by Krishnamurthy (2006) found that 20% of the IP addresses blocked by anti-spam filters were spoofed. Open proxy servers can be used by spammers to send email spam without revealing their identities. The open relay or proxy is a SMTP server that allows connection between the user and server without the need for authentication (Ramachandran & Feamster 2006). When email spam is forwarded from the proxy to the recipient the email spam contains the proxy address, not the spammers’ address (DeBarr & Wechsler 2010; Levy 2004; Xu et al. 2009). Boneh (2004) found that more than 60% of all email spam was sent through open proxy servers. Hoanca (2006) claimed that most of the open relays or proxies were in the US, with a small number of them in China, Korea and other countries. To detect open proxy servers, tools such as Send-Safe have been used by spammers to search for open proxies on the Internet (Gansterer et al. 2005).

Renting botnets was another method spammers used to hide their identities. Using this method, spammers send email spam from multiple computers to their mailing lists to avoid blacklist updates and to hide their identities (Eggendorfer 2008). Boneh (2004) found that some spammers used false names and untraceable payment methods to buy ISP roaming access. Researchers have developed many methods to address spammers' tricks related to spoofed IPs and open proxies. Li and Hsieh (2006) and Esquivel, Akella and Mori (2010) proposed methods to block spoofed IP addresses. Xie, Yin and Wang (2006) developed systems to detect open relays (proxies) that are used to send email spam. However, these methods have not been completely effective in blocking spammers' IP addresses and open proxy or relay servers used to send spam (Li & Hsieh 2006). This leads to the suggestion that the existing origin-based filters, which depend on network information such as IP and email addresses to classify spam (Cook et al. 2006; Garcia, Hoepman & Nieuwenhuizen 2004), need to be improved to block email spam before it arrives in end users' inboxes. This can be achieved by identifying fake or obfuscated email addresses generated by spam software, spoofed IPs, and open proxy servers used by spammers, and blocking them before reaching email inboxes. As the results of this study indicated that the percentage of fake or obfuscated email addresses was larger in mixed language and Arabic email spam than English, it is possible that a combination of origin- and content-based filters could block it effectively (Li & Hsieh 2006).

8.3 Conclusions

This chapter analysed the header and bodies of a collection of 1,270 Arabic, English and mixed language (Arabic and English text) email spam received from public users, businesses and ISPs over a period of two years. This analysis aimed to investigate the differences between spammers' tricks used in the three different language groups to enable email spam to bypass anti-spam filters and lure victims.

These tricks were different in Arabic and English spam. Examples of these tricks were the use of attractive words and false statements in the subject lines of emails. Attractive words were used more often in the subject line of Arabic than in English and mixed language email spam, whereas more English spam than Arabic and mixed language email spam used false statements in the subject line more often. This suggests the need to develop more effective filters to detect Arabic spam based on the information in the headers of emails, such as subject lines. This could be achieved by producing filters to classify spam based on the attractive words most frequently used in the subject lines (e.g. entertainment words in Arabic spam) (Christina, Karpagavalli & Suganya 2010). The taxonomy presented in the previous chapter (Chapter 7) described methods to develop filters to combat English spam based on the content of the header of email. These include support vector machines, boosting, maximum entropy (Zhang, Zhu & Yao 2004), and LVQ (Chuan et al. 2005). These methods, which the taxonomy identified as more effective than others in detecting English email spam, can also be employed to produce filters to combat spammers' tricks observed in Arabic email spam.

Another trick used by spammers was to use different formats in writing the content of email spam, such as image spam. Image spam was more common in Arabic email spam than in English and mixed language email spam. The proposed taxonomy, presented in the previous chapter, described many image spam filters, such as support vector machines and decision tree (C4.5), that were developed by studies such as Krasser et al. (2007), Saraubon and Limthanmaphon (2009) and Wang et al. (2010) to detect English image spam. Because these have been identified as more effective and accurate than other methods, for English image spam, and it is reasonable to assume that they can be used to develop new effective filters to combat Arabic image spam.

Adding malicious links (such as fake bank website link or forged unsubscribe links), or attachments (e.g. PDF, exe and GIF) in the content of email spam was another trick used by spammers. The percentage of malicious links and attachments was significantly higher in English email spam than for Arabic and mixed language email spam. Although many methods have been proposed to detect malicious links and attachments in spam, these methods are still not effective, which means developers should pay attention to developing more effective methods to block email spam that include malicious content such as phishing, viruses and trojans, which are embedded in links and attachments.

To hide their identities, spammers used spam software to generate fake or obfuscated email addresses. In this study, more mixed language and Arabic email spam than

English email spam was sent from obfuscated or fake email addresses, which indicates the greater use of spam software in generating fake email addresses in those languages. Although several methods have been proposed to classify email spam, based on network information such as IP addresses and email addresses, these methods are not completely effective in detecting spoofed IPs or fake email addresses. This suggests the need to further develop existing filters to more effectively blocking email spam from unknown senders before the spam arrives in end users' inboxes. Another suggestion is to create a combination of origin- and content-based filters (Li & Hsieh 2006).

No significant results were found about the tricks used in mixed language email spam to bypass anti-spam filters, and this due to the small number of collected mixed language email spam. Further investigation of a large corpus of mixed language email spam is needed.

Chapters 5, 6, and 7 described and discussed the results of public users, businesses and ISP questionnaires about the nature of Arabic and English email spam, its effects on their performance, how they dealt with it, and the effectiveness of anti-spam filters in detecting Arabic and English email spam. As the results indicated that anti-spam filters were not completely effective in detecting spam in either language, the spammers tricks were investigated. This could help developers to improve the effectiveness of existing anti-spam filters. This investigation was explained in this chapter (Chapter 8), and it was achieved by analysing the headers and bodies of a collection of Arabic, English and mixed language email spam received from public users, businesses and ISPs. The next chapter will describe and discuss the main findings of the questions of this research, and suggest possible approaches to combat email spam in Saudi Arabia.

Chapter 9: General Discussion and Possible Suggestions to Combat Email Spam in Saudi Arabia

This chapter discusses the answers to the research questions as revealed by the questionnaire responses of public users, businesses, and ISPs (Chapters 4, 5 and 6) about the nature of email spam, their awareness of it, their dealing with it and its effects on their performances. It also discusses the spammers' tricks revealed by the analysis of headers and bodies of Arabic and English email spam corpora (Chapter 8). Possible ways government, businesses and ISPs can combat spam in Saudi Arabia are discussed.

This chapter is divided into the following sections:

- Section 9.1: revisits the research questions.
- Section 9.2: discusses the major findings of the research.
- Section 9.3: provides possible suggestions to combat email spam in Saudi Arabia.
- Section 9.4: concludes this chapter.

9.1 Revisiting the Research Questions

Several research questions were developed in order to achieve the research objectives. These questions are:

Awareness of, filters for, and efforts to combat email spam

Q1: Are public users and businesses aware of email spam and anti-spam filters, what are the sources of their knowledge and how do they define email spam?

Q2: Are public users and businesses aware of government and ISPs efforts to combat spam in Saudi Arabia?

The nature of email spam

Q3: What is the volume of email spam received by public users and businesses and blocked by ISPs in Saudi Arabia; in which languages does it occur; and what are the sources or origins of Arabic and English email spam?

Q4: What are the differences between Arabic and English email spam?

Dealing with email spam

Q5: How do public users, businesses and ISPs deal with email spam?

The effects of email spam

Q6: What are the effects of email spam on the performance of public users, businesses and ISPs?

Anti-spam filters and their effectiveness in detecting Arabic and English spam

Q7: What anti-spam filters are used by Saudi ISPs to block email spam, and how effective are they in detecting Arabic and English email spam?

Spammers’ tricks used in the headers and bodies of Arabic and English email spam

Q8: What is the extent of the following spammers’ tricks used in the headers and bodies of Arabic and English email spam, respectively:

- attractive words or false statements in the subject line
- texts or texts embedded in images in the content
- malicious links and attachments, by type
- fake or obfuscated email addresses.

9.2 Discussion of the Major Research Finding

This section discusses the main findings of this research: the awareness of email users about email spam, anti-spam filters and efforts to combat it; the nature of email spam; how public users, businesses and ISPs dealt with it; its effects on their performances; and the effectiveness of anti-spam filters in detecting Arabic and English email spam. It also discusses the tricks spammers used in the headers and bodies of Arabic and English email spam to bypass anti-spam filters; the anti-spam filters in the proposed taxonomy, and their effectiveness in detecting email spam, mostly English. It then suggests which of these filters could be chosen to produce new filters for Arabic email spam.

9.2.1 The Awareness of Email Spam, Anti-spam Filters and the Efforts to Combat it

The awareness of public users and businesses in Saudi Arabia about email spam and anti-spam filters was low compared to that in other countries, such as Malaysia and Australia. As described in Chapter 4, about two-thirds of public users (62%, 95%CI: 59%-64.9%) were aware of email spam, while one-third of public users (37.9%, 95%CI: 35%-40.9%) were aware of anti-spam filters. This was a smaller percentage than in Malaysia. Bujang and Hussin (2010) demonstrated that about 86.5% of Malaysian email users were aware of email spam, and 66.9% were aware of anti-spam filters. One reason might be that public users in Malaysia were more experienced than users in Saudi Arabia in using the Internet and email, which could increase their knowledge of email spam and anti-spam filters (Sait et al. 2008).

The percentage of Saudi businesses (90.2%, 95%CI: 82.9%-95%) that were aware of email spam and anti-spam filters was lower than in Australia. All Australian organisations were aware of email spam and methods of combatting it (ACMA 2011). For public users and businesses in Saudi Arabia the most common source of information about email spam and anti-spam filters was the Internet. Public users and businesses were aware of few efforts by government and ISPs to disseminate this information. This finding was supported by the responses of Saudi ISPs, which demonstrated that slightly over half of ISPs provided awareness programs about email spam for their customers, and only a few ISPs conducted workshops or training about it for their employees.

There were a wide variety of definitions of email spam in Saudi Arabia. Public users most commonly defined email spam as “an email was sent randomly and contains malicious programs such as viruses”, whereas Saudi businesses and ISPs defined email spam as UCE. It can be seen that the definition of businesses and ISPs for email spam agreed with the international definition of email spam as UCE (Boykin & Roychowdhury 2004; Cheng 2004; Sakkis et al. 2003), although public users defined it differently than in previous studies. This suggests that it is important for government to arrive at an agreed definition of email spam so that it can be used in designing strategies and enacting law to combat email it (Everett 2004).

Only a few public users and businesses were aware of the efforts of government and ISPs to combat email spam in Saudi Arabia. The most common government effort that they were aware of was technical, and these technical measures were initiated by CITC and KACST. Of the efforts Saudi ISPs to combat email spam, public users and businesses were most aware of the use of anti-spam filters. This is supported by the responses of Saudi ISPs, which reported that setting and updating Internet security software and hardware (including anti-spam) (55.6%, 95%CI: 25.4%-82.7%) was their main way approach to the problem. Most of the strategies provided by the government and ISPs were technical, and there was a lack of educational and legal strategies. This suggests that a combination of legal, educational and technical measures, and cooperation between government and ISPs would be a more effective way to reduce email spam (Cheng 2004).

Clearly, then, it would be worthwhile for the government of Saudi Arabia to provide awareness programs to increase email users' understanding of spam, anti-spam filters, and the services and work being done to tackle it. Previous studies have revealed that governments in other countries, such as the USA (Pfleeger & Bloom 2005), Denmark (Frost & Udsen 2006) and India (Jidiga & Sammulal 2013), have conducted awareness programs to increase the awareness of public users and companies of online threats such as spam. This approach would help to reduce the volume and effects of email spam in Saudi Arabia (Dantin & Paynter 2005). Similarly, it would be useful for Saudi ISPs also to provide awareness programs and strategies to educate their email users about spam, anti-spam filters, their services and efforts to counter email spam (Pallas & Patrikakis 2005). Refai and Nyanchama (2007) suggested that establishing awareness programs about spam through workshops, seminars and training for employees and customers, can help to raise their awareness and fight email spam.

9.2.2 The Nature of Email Spam

This section discusses the main findings about the nature of email spam in Saudi Arabia in terms of its volume, its languages, types of Arabic and English email spam and their sources.

9.2.2.1 The volume of email spam

The average number of spam emails received by public users and businesses in Saudi Arabia was lower than the average number received in other countries such as the US (Grimes, Hough & Signorella 2007), Finland (Siponen & Stucke 2006) and UK (Computer Fraud and security 2004). This is a surprising result. Although the developed countries have enacted laws against spam, the participants in these countries received more email spam than the participants in Saudi Arabia. One reason for this result may be that participants in Saudi Arabia are less likely than participants in the developed countries to notice that they have received email spam. Another reason could be that public users and businesses in the three developed countries (US, UK and Finland) used the Internet and email more for online shopping or in dealing with customers (for businesses) than those in Saudi Arabia (Hermanson 2003). According to Hassanein and Head (2007):

> In an online shopping context, consumers are vulnerable and likely to expose themselves to loss if they provide their email address (making themselves vulnerable to receiving spam email or other annoyances).

This could result in public users and businesses in Saudi Arabia receiving less email spam than public users and businesses in the US, Finland and UK.

ISPs are one of sectors which are responsible for combatting email spam (Sorkin 2001). Although Saudi ISPs blocked millions of email spam (an average of 1,500,000 weekly), public users and businesses still receive it. This could be because the anti-spam filters used by Saudi ISPs were not effective in detecting email spam, which could let more through. The results of this study support this explanation, as Saudi ISPs reported that the anti-spam filters were not completely effective in detecting email spam. This finding suggests that Saudi ISPs need to develop their anti-spam filters to be more effective in detecting email spam. Another suggestion is be to apply anti-spam filters at two levels: email user and ISP to combat email spam more effectively (Khong 2001).

9.2.2.2 The languages of email spam

All three groups of participants agreed that most of the email spam received in Saudi

Arabia was written in English, and the highest percentage of English email spam was sent from non-Arabic countries. This finding is in line with previous studies conducted in other countries, such as Greece (Pallas & Patrikakis 2005), Bahrain (Al-A'ali 2007) and Malaysia (Bujang & Hussin 2010), which found that most of email spam received in those countries were written in English. It also corresponds with the results of Shrivastava and Bindu (2012), who found that English was the most popular language of email spam around the world, and with the results of Pfleeger's and Bloom's study (2005), which demonstrated that most of email spam received in the EU were written in English, even though there are about 12 official languages in the EU .

But Saudi Arabia has its own spammers. In this study, Arabic was the second most used language of email spam, other than English (Altbach 2004; Huddleston & Pullum 2002; Kirkpatrick 2007), and most of the Arabic email spam was sent from Saudi Arabia. Previous studies conducted in other countries such as Greece (Pallas & Patrikakis 2005), Bahrain (Al-A'ali 2007) and Malaysia (Bujang & Hussin 2010) have revealed that spammers aim to reaching more recipients by using English as the first language for writing email spam, followed by the native language of the countries studied, such as Arabic in Bahrain, Malay in Malaysia and Greek in Greece.

A lower percentage of email spam written in other languages (e.g. Chinese, Japanese, Russian, Turkish, French, Brazilian, Spanish, Persian, German, Italian, Hindi, Urdu, Hebrew) was received in this study. One reason for receiving email spam in different languages could be that public users and businesses have published their email addresses on the Internet, where they were harvested by spam software, resulting in the receipt of email spam in these languages (Andreolini et al. 2005). Other reasons could be that some public users subscribed to websites that were designed in these languages, or businesses dealt with customers who speak these languages, leading to receiving email spam in these languages (Wood 2013).

9.2.2.3 The differences between types of Arabic and English email spam

All three groups of participants agreed that more spam emails related to forums, as well as religious and political emails, were received in Arabic than in English.

Electronic forums are a popular way for Saudi society to pursue education, purchase and sell, or communicate with each other (Al-Saggaf 2004). (Arabic, of course, is the official language of Saudi Arabia) (Chejne 2009). When they subscribe to Arabic forums, their email addresses are added to the forum's mailing lists, which can result in receiving more Arabic spam. Another possibility is that owners of Arabic forums wanting to increase the number of subscribers, collect email addresses from the Internet by using automated or harvesting software (Andreolini et al. 2005), or buying email addresses from other Arabic forums (Cook et al. 2006). This can also result in receiving more Arabic emails related to forums.

The percentage of religious and political emails was higher in Arabic than English. This could be because some Arab countries used the Internet-based media for political campaigns (Grossman 2004; Sweet 2003), or for religious purposes (Martinkova 2008). Email users who follow up these political and religious issues, may then receive more religious and political email spam, with the result that they receive more spam emails of this nature in Arabic than in English.

Public users, businesses and ISPs reported that there were more pornographic and phishing and fraud emails in English than in Arabic. In Saudi Arabia, the official religion is Islam, which prohibits pornography (Al-A'ali 2007), but pornography is also forbidden in Arabic culture (Abdoh, Musa & Salman 2009). These reasons can also contribute to a lower percentage of pornographic emails in Arabic than in English.

Phishing and fraud emails were more common in English than in Arabic. There are several possible explanations for this. It is possible that the organised criminal elements behind most email phishing and fraud attempts are not yet established to the same extent in Arabic-speaking countries as elsewhere, such as English-speaking countries (Ramanathan & Wechsler 2012). It is also possible that some organisations designed online payment portals with English language interfaces (AlGhamdi & Drew 2012). Spammers would then design fake portals comparable to the original ones, and attach the link in email spam.

Other studies have investigated the types of email spam in other languages (e.g. Chinese and Russian), in different countries, which showed different types of email spam to those noticed in English and Arabic spam. Yamakawa and Yoshiura (2010) demonstrated that the most common type of Japanese email spam was related with sexuality. Lev and Goldin (2006) described different types of spam for four countries: Russia, China, Germany and Korea. Russian email spam targeted food, accessories, education and construction. In China, the most common type of Chinese email spam was the sale of fake invoices designed to reduce the tax burdens of different businesses, and anti-government spam. The subjects of German email spam included racist and white supremacist spam. Typical types of Korean email spam included finance- or mortgage-related emails (Lev & Goldin 2006). The reasons for the differences between types of email spam in different languages and in different countries are likely to be the spammers' culture, motivation, religion and country (Abdoh, Musa & Salman 2009).

9.2.3 How Public Users, Businesses and ISPs Deal with Email Spam

How email users deal with email spam differs from one email user to another based upon different factors, such as experience in using email, and knowledge of spam, its effects, and the filters used against it. Approximately one-fifth of public users in the Saudi Arabian study (20.9%, 95%CI: 18.5%-23.5%) responded to offers made in email spam. A study conducted in Malaysia by Bujang and Hussins (2010) revealed that only 8.1% of Malaysian email users responded to email spam. About a quarter of public users in Saudi Arabia (27.6%, 95%CI: 25%-30.5%) always deleted email spam without reading it, and just a few public users (3.1%, 95%CI: 2.2%-4.3%) reported that they always contacted the ISPs and notified them about spam. This might be because public users were unaware of the ISPs' services to combat email spam. The results showed that only two Saudi ISPs had received email users' queries and complaints about email spam issues, and provided support in this regard. Another study conducted in the USA by Grimes, Hough and Signorella (2007) demonstrated that 66% of 205 American users deleted email spam, and 11.7% contacted their ISPs when they received spam. It is clear that American and Malaysian email users were more aware than Saudi email users of ways of dealing with email spam. Further effort is needed by the government and relevant agencies to educate email users about effective methods of dealing with spam.

Saudi businesses and ISPs had different ways of dealing with email spam. ISPs and businesses established business units or created teams to manage network security issues, including email spam (81.8%, 95%CI: 53%-96%; 58.7%, 95%CI: 48.5%-68.5% respectively), and the most task of these units were in setting and updating Internet security software (55.6%, 95%CI: 25.4%-82.7%; 48.6%, 95%CI: 32.7%-64.7% respectively). It is important to establishing business units or create teams to manage information security in organisations to assure the safety of the organisation's information (Vroom & von Solms 2004). A study by Johnson and Koch (2006) demonstrated that approximately 12% of the information technology department budgets of the American organisations were spent on network security.

All Saudi ISPs used anti-spam filters to block email spam, but only about 80.4% (95%CI: 71.5%-87.5%) of businesses applied these filters. Some businesses might not have installed anti-spam filters because they relied on ISPs to filter it for them, as it is one of the responsibilities of ISPs to protect their customers (public users or businesses) from security attacks (Sorkin 2001). But it is also possible that some businesses outsource their security issues (including email spam), to technical companies to manage (Frost & Udsen 2006; Ridzuan, Potdar & Talevski 2010).

Approximately half of Saudi ISPs (45.5%, 95%CI: 20%-73%) had employees with the specific responsibility to combat email spam, and about one-fifth of businesses (18.5%, 95%CI: 11.6%-27.3%) assigned employees to manage the spam issues. This could be because Saudi businesses did not have an IT department or business units to manage network security issues. This study found that less than half of businesses did not have business units to manage network security. Another possibility might be that businesses outsource the work, for example to ISPs, or hire external employees to combat email spam (Frost & Udsen 2006; Ridzuan, Potdar & Talevski 2010). Qualified people are needed to fix email spam problems. Arutyunov (2013) stated that in one American company, one full-time IT person was allocated to every 690 employees to address email spam issues. Ridzuan, Potdar and Talevski (2010) emphasised that companies need to spend money to buy the necessary anti-spam filters, recruit employees to deal with spam issues, and provide the required training for those employees to improve their understanding of email spam.

9.2.4 The Effects of Email Spam on the Performance of Public Email Users, Businesses and ISPs

Email spam had negative impact on the performance of public emails users, businesses and ISPs in Saudi Arabia. About half of public users (45.1%, 95%CI: 42.1%-48.2%) were negatively affected by email spam. Its major impact on the performance of public users was email inboxes filling with spam (28.1%, 95%CI: 25.4%-31%), which can consume the available email capacity and result in the loss of important emails (Zhang, Zhu & Yao 2004). The large volume of spam in email inboxes can also waste email users' time with the need to isolate important email from spam emails (Chigona et al. 2005; Hinde 2002; Özgür, Güngör & Gürgen 2004), and reduce their productivity (Leng 2006). About one-fifth of public users in this study (19.4%, 95%CI: 17.1%-21.9%) were affected by email spam through lost time and reduced productivity. Another major effect of email spam on the performance of public users was infection of computers by malicious programs such as viruses and trojans (24.5%, 95%CI: 21.9%-27.2%). This can be a way to steal the important information, such as credit card numbers (Cournane & Hunt 2004; Hermanson 2003). This suggests the need for educating public users about the malicious effects of email spam (Dantin & Paynter 2005).

The main impact of email spam on the performance of Saudi businesses was reduced efficiency of the organisation's email server due to the receipt of a large volume of spam (82.6%, 95%CI: 73.9%-89.3%), which can be a burden on the email server (Mo et al. 2006). Some spams have attachments, which when downloaded consumes the organisation's bandwidth (Cook et al. 2006). This suggests that by developing anti-spam filters to block email spam before it arrives, companies' networks and email servers could be more efficient. The second impact of email spam on the performance of Saudi businesses was loss of time and reduced productivity (71.7%, 95%CI: 62%-80.2%). Previous studies indicated that employees or workers spent more time isolating spam emails from important emails and fixing email spam problems (Bujang & Hussin 2013; Pérez-Díaz et al. 2012). Siponen and Stucke (2006) stated that employees spent an average of 13 minutes a day fixing problems related to email spam. Another study by Caliendo et al. (2008) demonstrated that employee spend about 1,200 minutes per each year identifying and deleting email spam. This time spent in fixing email spam problems can cost companies a lot of money in reduced productivity (Takemura & Ebara 2008).

Email spam also had a negative impact on the performance of Saudi ISPs. ISPs used anti-spam filters to block email spam, but applying and updating these filters is expensive (Ridzuan, Potdar & Talevski 2010). The greatest impact of email spam on the performance of Saudi ISPs was the expense of buying and updating anti-spam filters (90.9%, 95%CI: 64.7%-99%). Previous studies have indicated that ISPs in some countries spend billions of dollars to block spam. The US FTC forum reported that American ISPs spent billions of dollars to block spam (Allman 2003). A study conducted on the European community demonstrated that the ISPs paid about 10 billion euros a year to combat spam (Garcia, Hoepman & Nieuwenhuizen 2004).

Saudi ISPs were affected by email spam through the consumption of the bandwidth by excessive email spam (63.6%, 95%CI: 34.8%-86.3%). This effect cost ISPs more money to buy extra bandwidth. Cournane and Hunt (2004) described the impact of a large volume of email spam as consuming bandwidth; the extra spam traffic results in a slower Internet service to subscribers, the need to spend more money to increase the bandwidth, and increasing charges to subscribers due to the large bandwidth usage. This can affect the reputation of the ISPs and result in loss of customers or subscribers (Khong 2001; Moustakas, Ranganathan & Duquenoy 2005; Potashman 2006; Smith 2004). In this study, about one-third of Saudi ISPs (36.4%, 95%CI: 13.7%-65.2%) reported losing customers due to receipt of a large volume of email spam. A US study found that 7% of customers switched their ISPs because of email spam issue (Gartner Group 1999).

About half of Saudi ISPs (45.5%, 95%CI: 20%-73%) reported that fixing problems related to email spam wasted their time and reduced productivity, with an average 4 hours per week spent fixing these problems. A study conducted in Germany by Caliendo et al. (2012) revealed that ISPs spent an average of 25 minutes per week to solve email spam issues (Caliendo et al. 2012). Another study found that the time lost in fixing email spam problems was 40 minutes each week (Brod 2004). It can be clearly seen from these results that the average number of hours lost to fixing problems related to spam was higher in Saudi Arabia than other countries such as Germany and US.

Email spam in Saudi Arabia not only results in loss of time and reduced productivity of public users, businesses and ISPs, but can potentially affect the economic growth of the country, costing the government and companies billions of dollars. (Cook et al. 2006) found that about email spam cost US companies $10 billion in lost productivity. The Singapore IDA indicated the total cost of spam to consumers at about S$23 million in lost productivity each year (Leng 2006). In Japan the cost to GDP due to processing email spam was reported to be about 500 billion yen (Takemura & Ebara 2008). This suggests the importance of the government or relevant agencies in Saudi Arabia conducting further efforts (legal, educational and technical) to mitigate email spam and its effects.

9.2.5 The Anti-spam Filters Used by Saudi ISPs, and Their Effectiveness in Detecting Arabic and English Email Spam

The most common filters used by Saudi ISPs to combat email spam were Iron Port (content-based) and blacklists (origin-based). Previous studies have shown that different filters have been used by ISPs in other countries such as the US, South Africa and Greece to block email spam. In the US, Brightmail was the most common filter used by American ISPs (Gartner Group 1999). The ISPs in South Africa applied many filters to detect email spam, such as Postfix, Sender Policy Framework (SPF), SpamAssassin, Bayesian filters, distributed blacklists, heuristic engines, and statistical classification filters (Chigona et al. 2005). Greek ISPs deployed some filters such as Domain Name System Blacklists (DNSBLs), heuristic techniques, and custom techniques (Pallas & Patrikakis 2005). A study conducted by the European Network and Information Security Agency (ENISA) found that the most common filters used by the ISPs in 19 different countries of the EU were blacklists (Rossow 2007).

Despite Saudi ISPs updating their anti-spam filters regularly, the Saudi ISPs did not believe they were completely effective in detecting Arabic and English email spam. This finding is in line with other studies conducted by Chigona et al. (2005), Pallas and Patrikakis (2005) and Rossow (2007). This suggests the need to improve the existing anti-spam filters to be more effective in detecting Arabic and English email spam. The anti-spam filters used by Saudi ISPs had better performance in detecting English email spam than Arabic email spam. This finding corresponds to the results of other studies such as El-Halees (2009), Çıltık and Güngör (2008) and Nguyen, Tran & Nguyen (2008), which found that anti-spam filters were more effective in detecting English email spam than non-English email spam.

There is a clear need to refine the existing anti-spam filters to be more effective in detecting Arabic spam, or to develop new filters. To support this suggestion, the author has proposed a taxonomy that contains most of the email spam filters that have been developed to detect email spam (presented in Chapter 7). The taxonomy found that these filters depended on different methods to detect email spam based on the content of the header and body; and most of these filters have been were proposed to detect English spam. These filters included SVMs, boosting, maximum entropy (Zhang, Zhu & Yao 2004), LVQ (Chuan et al. 2005), and DT (C4.5) (Krasser et al. 2007). These filters were more effective than other filters in detecting English email spam, and so it is reasonable to assume these methods could be used to produce more effective filters against Arabic email spam. Another suggestion is to develop filters that combine reputation-and content-based methods to detect Arabic email spam (Li & Hsieh 2006). A list of keywords observed in Arabic email spam, as well as Arabic email spam corpora, were used in this research (described in Chapter 8). These materials be used to improve the effectiveness of current filters in detecting Arabic email spam.

Many studies such as Wang et al. (2007) and Hayati and Potdar (2009) have claimed that one constraint on the effectiveness of anti-spam filters in detecting email spam is that spammers are constantly developing their methods and tricks of bypassing these filters. For this reason, spammers' tricks used in Arabic and English email spam were investigated in this study and suggestions made for screening these tricks with anti-spam filters, especially for Arabic language spam. A taxonomy of most of the methods used in previous studies to combat email spam, mostly English-based, has helped in suggesting appropriate filters to combat spammers' tricks used in the headers and bodies of Arabic email spam. The next section discusses these tricks.

9.2.6 Spammers' Tricks Used in the Headers and Bodies of Arabic and English Email Spam

This study demonstrated that different tricks have been used by spammers in email spam, and that these tricks were different for Arabic and English spam. Spammers used more 'attractive' words in Arabic spam than in English and mixed (contains

Arabic and English text) spam. The probable motivation for using attractive words in the subject line of spam is to convince the recipients to open the email (Attar, Rad & Atani 2013; Dhinakaran, Jae Kwang & Nagamalai 2009). The attractive words used in the subject line of Arabic spam were related to entertainment. This finding differs from the results of a study conducted on Japanese email spam, which found that the attractive words used in the subject line were related to sexuality (Yamakawa & Yoshiura 2010). The reason for this difference could be cultural. According to Abdoh, Musa and Salman (2009), the nature of email spam is different from one country to another because of factors such as the motivations and cultures of spammers. Some spammers send commercial advertisements, some send pornographic emails, and others send malicious programs.

Although a few methods have been developed to detect Arabic email spam based on the header (El-Halees 2009; Goweder et al. 2008), these methods are still not effective (Hayati & Potdar 2009). This suggests the need to develop anti-spam filters to block Arabic spam. This can be achieved by developing filters to detect spam, based on the words observed in the subject line (entertainment advertisement words, as observed in this study). Anti-spam filters can block spam by analysing words in the subject line and body of email. Christina, Karpagavalli and Suganya (2010) developed a filter to block email spam by adding keywords observed in the subject line and body of spam to the lists of one of the existing anti-spam filters: "[U]sing combinations of keywords is a good solution to enhance filtering efficiency".

The taxonomy proposed in this study revealed the different methods that have been employed to develop filters to block email spam, based on the content of the header. Most of these methods have been developed to detect English spam. These methods included SVMs, boosting, maximum entropy (Zhang, Zhu & Yao 2004), and LVQ (Chuan et al. 2005), which the taxonomy revealed to be the most effective in detecting English email spam. These methods could be applied to create new, more effective filters to detect Arabic email spam.

English spam had more false statements (misleading subject line) in the subject line than Arabic and mixed spam. A false statement was defined by other researchers such as Hamel (2004) and Simon (2004) as a subject line that does not indicate the content of the email. This can make it difficult for the recipients to determine the content of the email before they open it (Chigona et al. 2005). The percentages of false statements observed in the subject line of Arabic and English spam (55% and 72% respectively) received in Saudi Arabia was higher than the percentage observed in the subject line of email spam received in the US. A study conducted by the US FTC (2003) demonstrated that 40% of the subjects of spam emails sent to the American recipients had false statements or misleading subject lines.

Image spam was used more often in Arabic spam than in English and mixed spam. Image spam was defined by other researchers such as Xu et al. (2009) and Soranamageswari and Meena (2010) as a type of email spam in which content of email appeared in the body of message as an image instead of text. Previous studies have indicated that image spam was probably developed to circumvent text-based filters (Attar, Rad & Atani 2013; Gargiulo & Sansone 2008; Nielson, Aycock & de Castro 2008; Zuo et al. 2009).

The proposed taxonomy has categorised different image spam filters and their effectiveness in detecting image spam. These filters (e.g. vector machines and C4.5 decision tree) were developed by previous studies such as Krasser et al. (2007), Saraubon and Limthanmaphon (2009) and Wang et al. (2010) to detect English image spam. Compared to other methods used to filtering spam, these filters were effective and accurate in detecting English image spam. It is suggested that these methods can be used to produce new, effective filters to combat Arabic image spam.

Links and attachments have been used in email spam for malicious purposes. This kind of spam appeared more often in English spam than in Arabic and mixed spam. There can be different motivations for including links in the body of spam, including redirecting email users to web pages that promote products or commercial services (Attar, Rad & Atani 2013), downloading malicious programs onto computers (Kumar 2009; Smith 2008), and opening fake bank web pages to steal important information such as credit card numbers (Barroso 2007; Leavitt 2005). The percentage of malicious links that redirect to spammers' or phisher websites, or download malicious programs (Dhinakaran, Lee & Nagamalai 2007a; Nagamalai, D, Dhinakaran, C & Lee, J-K 2010) has been reported to be higher in English spam than in Arabic and mixed spam. The most common malicious links observed in English spam in Saudi Arabia were links that redirect to fake bank web pages. This finding supports the findings of previous studies conducted by Hinde (2003), IBM X-Force® (2011) and Ramanathan and Wechsler (2012), which found that banks were the sector most targeted by phishers.

Another type of malicious link used by spammers was the forged unsubscribe link. An unsubscribe link is as an option that enables email users to remove their email addresses from mailing lists that they have subscribed to, or to unsubscribe from receiving more emails in the future (Allman 2003; Malcolm 2004; Vaile 2004). Forged unsubscribe links are used by spammers to test whether or not the email address is valid, which can result in the user receiving more email spam (Chigona et al. 2005; Lambert 2003; Simpson 2003); as a way to add victims' addresses to spammer lists, which also increases the probability of receiving a large volume of email spam in the future; as a way to advertise products, redirecting the user to business web sites (Andaker et al. 2006; Lambert 2003). The results of this study indicated that Arabic spam contained more forged unsubscribe links than English and mixed spam.

Spammers uploaded different types of attachments such as image, word or pdf files in email spam. This may be to evade text-based anti-spam filters. Attar, Rad and Atani (2013) reasoned that, as commercial or product advertisements, appearing as texts in the body of email (the traditional way of spamming) can be detected by the text-based anti-spam filters, spammers add their advertisements in attachments such as images or pdf files. The results of this study demonstrated that the percentage of attachments in Arabic spam was higher than the percentage in English and mixed spam, and images were the most common attachment observed in Arabic spam. However, some of the attachments included in Arabic and English spam had vicious objectives, such as to infect users' computers with malicious programs such as viruses or malware (Cournane & Hunt 2004; Dantin & Paynter 2005; Hershkop & Stolfo 2004; Moustakas, Ranganathan & Duquenoy 2005; Pallas & Patrikakis 2005; Pfleeger & Bloom 2005; Sorkin 2001).

In the researcher's Saudi Arabian study, English spam had more malicious attachments than Arabic and mixed spam. This is supported by Ramanathan and Wechsler (2012), whose study revealed that the malicious spam, such as phishing, viruses, worms and trojans, appear more often in English than in non-English languages. Although many methods for detecting malicious attachments in email spam have been developed by other researchers such as Stolfo et al. (2010) and Amin (2011), further development of these methods is required, especially for detecting malicious attachments in English email spam.

The percentage of mixed and Arabic spam sent from obfuscated or fake email addresses was higher the percentage in English spam. The reason for obfuscating email addresses can be to hide spammers' identities, bypass anti-spam filters, or to trick email users (Hayati & Potdar 2009). Previous studies such as Dhinakaran, Jae Kwang and Nagamalai (2009) have revealed that generating fake or obfuscated email addresses of spam senders can be achieved with spam software. Examples of spam software include Phasma Email Spoofer, Bulk Mailer, Aneima 2.0, Avalanche 3.5, and Euthanasia (Dhinakaran, Lee & Nagamalai 2007a). This indicated that spam software has been used to produce fake email addresses for mixed and Arabic email spam more than for English email spam. This discussion suggests that the origin-based filters, which depend on network information such as IP and email addresses to classify spam (Cook et al. 2006; Garcia, Hoepman & Nieuwenhuizen 2004), need to be developed to block email spam before it arrives in end users' inboxes. This might be achieved by identifying fake or obfuscated email addresses generated by spam software and blocking them before they reach email inboxes.

9.3 Research Suggestions to Combat Email Spam in Saudi Arabia

On the basis of the results of this study, legal, educational and technical solutions have been suggested for the government (governmental authorities or decision-makers), ISPs and businesses to mitigate email spam in Saudi Arabia. The literature review found that some of these suggestions had been effective in combatting email spam in different countries.

Overall, this study suggests that it is important for the government to design new strategies or policies, including new laws against spam in Saudi Arabia, awareness programs to educate email users about it, and technical measures to block it. This could help in reducing email spam and its effects in the country. Previous studies have found that different countries such as the UK ('ISPs get tougher on spam' 2004), Australia (Moustakas, Ranganathan & Duquenoy 2005), Denmark (Frost &

Udsen 2006) and the US (Sorkin 2009) have designed strategies that were effective in reducing email spam. A combination of measures, such as technical, legal, educational, and through international collaboration could be an important solution to the problem of spam. According to Cheng (2004), one of the important solutions to combat spam was “a combination of self-help preventive measures such as anti-spam filtering tools, robust regulation, international cooperation, and education and awareness of users”. Frost and Udsen (2006) recommended a combination of a number of different effective efforts including legislation, technological improvements such as advanced filters, education of users and companies about spam, and self-regulation by businesses and ISPs. Potential legal, educational and technical efforts are described in the following sections. Figure 9.1 summarises these suggestions.

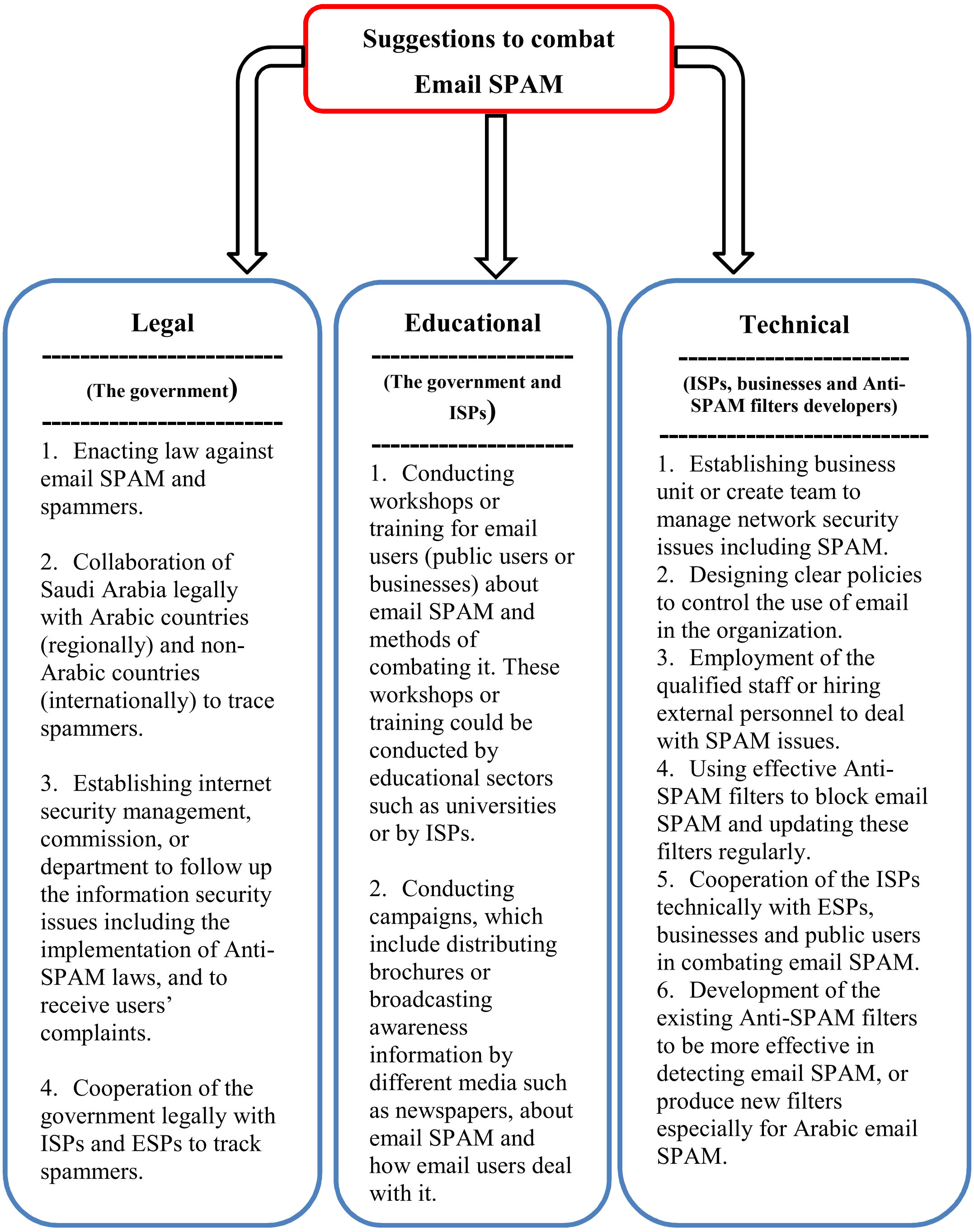
9.3.1 Legal Suggestions

This section provides legislative suggestions for the government, decision-makers or relevant agencies in Saudi Arabia. Anti-spam laws have contributed to a drop in the volume of email spam in some countries. According to Lev and Goldin (2006), “the spam to legitimate email ratio in Japan is much lower than average due to the strict attitude towards law enforcement”. In the USA, the volume of email spam decreased and spammers outside of the USA to escape the strict spam laws and send spam from countries that do not enact laws against spam (Yamakawa & Yoshiura 2010).

As spammers use many tricks in the headers (e.g. using attractive words or misleading subject lines) or in the bodies (malicious links or attachments) of email spam to bypass anti-spam filters or to lure the recipients, anti-spam laws in the USA have addressed these tricks: “Some laws are very tough and strict like United States CAN-spam Act of 2003, and one of the main provisions of the Act is that deceptive subject lines are prohibited” (Xu 2010). Some states in the US, such as Virginia, have enacted laws to prohibit false statements and misleading subject lines:

> The Virginia law enacted under that state’s Computer Crimes Act addresses the use of misleading subject lines, forged email headers, and criminal trespass when a spammer illegally uses a

computer to send out email messages and help disguise the origin of the email (Grimes (2004).



Suggestions to combat Email SPAM

Legal

(The government)

1. Enacting law against email SPAM and spammers.

2. Collaboration of Saudi Arabia legally with Arabic countries (regionally) and non-Arabic countries (internationally) to trace spammers.

3. Establishing internet security management, commission, or department to follow up the information security issues including the implementation of Anti-SPAM laws, and to receive users' complaints.

4. Cooperation of the government legally with ISPs and ESPs to track spammers.

Educational

(The government and ISPs)

1. Conducting workshops or training for email users (public users or businesses) about email SPAM and methods of combating it. These workshops or training could be conducted by educational sectors such as universities or by ISPs.

2. Conducting campaigns, which include distributing brochures or broadcasting awareness information by different media such as newspapers, about email SPAM and how email users deal with it.

Technical

(ISPs, businesses and Anti-SPAM filters developers)

1. Establishing business unit or create team to manage network security issues including SPAM.
2. Designing clear policies to control the use of email in the organization.
3. Employment of the qualified staff or hiring external personnel to deal with SPAM issues.
4. Using effective Anti-SPAM filters to block email SPAM and updating these filters regularly.
5. Cooperation of the ISPs technically with ESPs, businesses and public users in combating email SPAM.
6. Development of the existing Anti-SPAM filters to be more effective in detecting email SPAM, or produce new filters especially for Arabic email SPAM.

Figure 9.1: Possible suggestions for the government, ISPs, Businesses and Anti-SPAM filters developers to combat SPAM in Saudi Arabia

Virginia's legislation that includes criminal penalties for fraudulent and high-volume spamming (Butler 2003). In Washington and California states, one of the provisions of the anti-spam law was that the word "ADV" must be included in the subject line of commercial email to warn the recipient that the email is an advertisement (Fogo

2000). This supports the need to enact law to combat spam in Saudi Arabia. The experiences of countries that have enacted laws to combat spam can help in enacting a balanced and culturally fit law against spam and spammers in Saudi Arabia.

To address the issue of moving spammers between countries to send email spam, some countries have cooperated legally with each other to trace the origin or source of spammers. Examples of this cooperation are: the tripartite Memorandum of Understanding on Spam Enforcement Cooperation (an agreement between the UK, US, and Australia in combatting spam) and London Action Plan cooperation (collaboration between 19 bodies from 15 countries) (Moustakas, Ranganathan & Duquenoy 2005). As a percentage of email spam received by public users and businesses was sent from outside of Saudi Arabia, this suggests that the collaboration of Saudi Arabia legally with other Arabic countries (regionally) and non-Arabic countries (internationally) could be an efficient strategy to mitigate it. This could be achieved by tracing the origin of spammers in the cooperating countries (regionally or internationally) and bringing them to justice.

One of the issues in enacting laws to combat spam is the definition of email spam. The literature review demonstrated that there were a wide variety of definitions. Even in Saudi Arabia, this study revealed that there was no specific definition of email spam by Saudi society. This suggests that it is worthwhile for the government to specify an agreed definition of email spam that can be used for enacting laws to combat spam in Saudi Arabia (Everett 2004).

Another suggestion is to establish an Internet security management commission, or create a cyber-crimes department to enforce information security, including implementation of anti-spam laws and dealing with users' complaints about security attacks. Different countries have created agencies or units to address information security issues. In 2005, ENISA was established in Greece. The agency's mission was to achieve a high level of information security within the EU institutions and member states. "The ENISA seeks to develop a culture of network and information security for the benefit of citizens, consumers as well as business and public sector organisations in the European Union" (Rossow 2007). In the USA, laws to combat spam are regulated by the FTC (Hinde 2002; Khong 2001; Rogers 2006; Sorkin 2009). In Australia, ACMA is responsible for implementation of anti-spam laws

(Australian Communications & Media Authority 2006; Cheng 2004; Moustakas, Ranganathan & Duquenoy 2005). Two mailboxes in Denmark were created by the Danish Consumer Ombudsman office to receive complaints about spam (one for Danish spam and one for international spam) (Frost & Udsen 2006).

Cooperation between the government, ISPs, and the public and private sectors to enforce anti-spam legislation could be another effective way to combat spam in Saudi Arabia. This could be achieved by ISPs receiving subscribers' complaints about email spam, warning spammers about their misuse of email, and reporting spammers' activities and IDs by to the legal agencies in Saudi Arabia. A study by Leng (2006) indicated that cooperation between the ISPs, ESPs, public users, businesses and the government is important to trace the sources of email spam and then combat them legally. Butler (2003) claimed that AOL, Microsoft and Yahoo collaborate actively with the US law enforcement agencies to combat spam. They have developed a mechanism that includes preserving evidence of spammers' activities and coordinating enforcement efforts with industry, such as by referring spammers to the police or government agencies.

9.3.2 Suggestions for Educating Email Users about Spam

This section provides suggestions for the government and ISPs about how to educate email users about email spam. The awareness of public users and businesses in Saudi Arabia about email spam, anti-spam filters and effort to combat it, was low. There was also a perceived deficiency in the government's efforts to increase awareness of email spam and anti-spam filters: only a few public users and businesses were aware of government programs around email spam and spam filters. This suggests that it is necessary for the government to educate public email users and businesses about email spam and ways to combat it. Educating email users is one of the most effective ways to combat email spam: "an important step towards minimising unsolicited and unwanted emails is raising the awareness of users" (Dantin & Paynter 2005). The Confederation of Danish Industries and the Danish Consumer Ombudsman office provided awareness programs for private individuals and companies (Frost & Udsen 2006). According to Jidiga and Sammulal (2013), "the Indian government organizations like Ministry of Communication and Information Technology set up separate divisions to conduct security awareness programs to the people, employs, students about spam".

Awareness programs in Saudi Arabia could be conducted by educational sectors, such as universities, through workshops or training for email users. Refai and Nyanchama (2007) described spam awareness programs that provide workshops, seminars and training for employees and customers, as valuable ways to increase awareness and fight email spam.

Conducting campaigns about email spam and how to deal with it, such as distributing brochures or broadcasting awareness information through different media, could be another way to educated email users about email spam. The member states of the EU have taken many actions to increase the awareness of email users about spam, including campaigns (Pfleeger & Bloom 2005).

Saudi ISPs can take a part in education and awareness of email users (public users and businesses) about email spam and the effective ways in how they deal with it. This can be achieved by conducting workshops or training for email users (employees and customers) (Refai & Nyanchama 2007). This study found that a few Saudi ISPs conducted workshops for their employees and provided awareness programs for their customers about email spam. These require additional effort. Pallas and Patrikakis (2005) suggest that "the ISPs should take actions by providing assistance to enforcement agencies along with undertaking of users' education about spam". A previous study conducted on Singapore email users suggested that one of the responsibilities of ISPs was to advise email users about using anti-spam filtering software and effective steps they can take to combat email spam. This might be achieved through workshops and newsletters (Leng 2006). In Denmark, the ISP Security Forum, established by nine Danish ISPs, was responsible, amongst other tasks, for providing common guidelines for customers about email spam and filters used to combat it (Frost & Udsen 2006).

9.3.3 Technical Suggestions

This section provides technical suggestions for ISPs, businesses and anti-spam developers, which may help in combatting email spam. Some Saudi ISPs and businesses did not have business units or teams to manage network security; however, this is an important function that warrants a specialised task force. A previous study indicated that establishing business units or management, or creating teams to manage information security in organisations, is important (Vroom & von Solms 2004). Johnson and Koch (2006) reported that about 12% of the budget of the IT department was spent on network security in the USA. A specialised unit or team could conduct the following tasks: apply and update Internet security software or hardware to block security attacks, design security policies for the organisation and provide technical support when customers need it (von Solms 2005). This could help reduce email spam and its effects in Saudi Arabia (Vroom & von Solms 2004).

Using effective anti-spam filters and updating them regularly is another suggestion to combat email spam in Saudi Arabia. Anti-spam filters can save ISPs and businesses millions of dollars. A study by Osterman Research Inc. (2008) found that the cost of email spam to a company with an average number of 1,200 employees could be US$2.4 million, but by using anti-spam filters, they could save US$1.2 million. Other researchers have pointed to different effective approaches in using anti-spam filters in organisations. A technology consultant at Mirapoint, a company that sells email security products, recommended that network managers use an email firewall with anti-spam and anti-virus software to monitor and clean machines, and they should update software regularly. The consultant also recommended that network managers implement intrusion detection software to prevent spammers' activities from taking place within the firewall (Everett 2004). According to Clayton (2004), some ISPs installed email "smarthost" servers for their customers. The customers used an email client to transfer outgoing emails to the smarthost. The smarthost servers then arranged emails for delivery to remote sites. Some ISPs redirect all outgoing port 25 traffic, the port used by the SMTP, to the smarthost and make its use compulsory (Clayton 2004). Sorkin (2001) suggested that "Filtering by ISPs and third-party proxy filtering services like Brightmail can be more effective than end user filtering, requiring less effort and expertise on the part of the users".

As a few Saudi ISPs and businesses had specific employees to combat email spam, this suggests that the employment of qualified staff with a specific responsibility to combat email spam in the ISPs and businesses, or hire external employees could be effective in combatting email spam. Those employees can deploy and update anti-spam filters, and add email spam and spammers into their blacklists. In the USA, UUNET, one of the largest ISPs in the world, has created a group of six employees with a budget of one million dollars, and with a specific responsibility to combat spam (Khorsi 2007). Arutyunov (2013) stated that one business in the USA allocated one IT person for every 690 workers, just to fix spam problems.

Another effective technical suggestion for combatting email spam is to design clear policies and standards to control the use of email in the organisation, as the results indicated that only a few Saudi businesses had designed security policies to combat malicious attacks. These policies could contribute in reducing the volume of email spam and its effects. According to Sorkin (2001), some ISPs and organisations have applied clear policies that do not allow using their facilities to send email spam. Spammers and spam-friendly providers were blacklisted and were of by the ISPs and boycotted. A study conducted by Sunner (2005) on 182 IT security professionals in the UK revealed that 51% had formal policies regarding security attacks. Pfleeger and Bloom (2005) reported that some companies developed standards and policies to combat spam. An example of these companies is the ePrivace Group, which has developed the Trusted Email Open Standard (TEOS) to reduce the volume of spam. Some industry groups representing marketers and ISPs have tackled email spam by applying self-regulatory policies. An example of these policies is that developed by the Direct Marketing Association (DMA), which prohibits members from sending email spam to email addresses that appear in the DMA database (Leng 2006; Sorkin 2001).

Technical collaboration between Saudi ISPs, or cooperation between ISPs and ESPs, businesses, and public users could reduce email spam and its effects in Saudi Arabia. According to Leng (2006), it is necessary for ISPs to collaborate with network service providers to trace the origin of email spam. Collaboration could be achieved by creating a special group or forum in Saudi Arabia for ISPs, ESPs and businesses to discuss the best technical practices to combat email spam. Previous studies have indicated the importance of such these forums. In the UK, about 150 ISPs at the LINX forum tackled spammers who hosted their websites on reputable ISPs but sent spam from other networks. The LINX forum recommended shutting down websites that sell spamming accessories such as stolen email addresses. Malcolm Hutty, a LINX regulation officer, said that LINX was the best current practice to stop spam.

The officer said that the number of open relay mail servers that sent spam was about 20% of the UK mail servers in 1999 and this number decreased to less than 1% in 2003, for which LINX was responsible ('ISPs get tougher on spam' 2004). In Denmark, nine ISPs created an organisation called, the ISP Security Forum, to combat security attacks, including email spam, technically. This organisation aimed to provide a central spam filter for customers and to take technical actions against spammers who send spam from their Internet connections (Frost & Udsen 2006).

Public email users can cooperate with ISPs to reduce the volume of email spam and its effects. One way is by paying money for spam-filtering services. A study conducted by the Gartner Group (1999) revealed that 24% of the American participants were willing to pay money to ISPs that provide a spam-filtering service. The study found that 70% of the participants would pay ISPs $1 or more each month to filter spam. Another way public users can cooperate with the ISPs is to pay additional money for exceeding an email limit in a certain period (Leng 2006).

Anti-spam filters were not effective in detecting English and Arabic email spam, although these filters performed better in detecting English spam than Arabic spam. This result was based on Saudi ISPs' evaluation of the effectiveness of these filters. Previous studies such as Chigona et al. (2005), Pallas and Patrikakis (2005), Rossow (2007), Çıltık and Güngör (2008), Nguyen, Tran and Nguyen (2008) and El-Halees (2009) have supported this finding. This suggests that further development of the existing anti-spam filters is needed. This could be achieved by the cooperation of ISPs with companies that develop anti-spam filters, and discussion about the strengths and weakness of filters. This could lead to producing filters with a higher performance in detecting email spam (Potashman 2006).

Another suggestion is to produce new filters to detect Arabic email spam. There are two ways to do this. The first way is to use effective content-based filters (extracting keywords from the content of the email header and body) (Cook et al. 2006) that were effective in detecting email spam in other languages such as English, and apply them in Arabic spam. Christina, Karpagavalli and Suganya (2010) suggested that "using combinations of keywords is a good solution to enhance filtering efficiency". The second way is to develop filters based on a combination of reputation- and content-based methods. This could yield filters with high effectiveness and accuracy in detecting Arabic email spam (Li & Hsieh 2006). To develop current filters or produce new, more effective filters for Arabic email spam this study produced:

- a taxonomy of email spam filters containing methods proposed by other researchers to detect email spam, mostly in English
- an Arabic email spam corpus
- a list of keywords and phrases used in Arabic spam.

9.4 Conclusions

This chapter began by revisiting the research questions and discussing the major findings to these questions. It also provided legal, education and technical suggestions for combatting spam in Saudi Arabia.

The awareness of email users about spam, anti-spam filters and the efforts being made to combat it was low compared to other countries. Most email users did not know how to deal with spam. There was little awareness of the efforts provided by the government and ISPs to educate email users about spam, which suggests that it would be fruitful for government or relevant agencies to provide awareness programs for email users. This could reduce the effects of email spam in Saudi Arabia (Dantin & Paynter 2005). There was also a deficiency in the technical efforts of businesses and ISPs to combat email spam, which suggests the need for additional work. This could reduce the volume of email spam (Lam & Yeung 2007). There is no specific definition of email spam in Saudi Arabia, which indicates the need for an agreed definition that can be used in designing strategies and policies to combat it (Everett 2004).

Most of the email spam in Saudi Arabia was written in English and the greatest percentage of English spam was sent from non-Arabic countries. Arabic was the second most language used in email spam and most Arabic spam was sent from Saudi Arabia. This indicated that Saudi Arabia has its own spammers, which requires the implementation of anti-spam laws in Saudi Arabia. This could reduce the volume of email spam and spammers' activity. Cooperation with other countries, legally or technically, to trace spammers' sources and their activities, is another suggestion (Leng 2006).

Emails related to forums, religion and politics were more common in Arabic than English, while English had more pornographic, and phishing and fraud emails than Arabic. The differences between types of Arabic and English spam can be explained by the differences in the culture and motivations of spammers (Abdoh, Musa & Salman 2009). Email spam had negative impacts on the performance of email users and ISPs. It wastes employees' time in isolating spam from non-spam, and the infection of computers by malicious programs attached to spam such as fraud and phishing emails. In reducing the productivity of employees, such spam can in turn affect the economic activity in Saudi Arabia, which suggests the need for anti-spam laws to combat email spam.

Saudi ISPs thought that anti-spam filters were not completely effective in detecting email spam, and performed better in detecting English spam than non-English spam. Tricks used by spammers in the header and body of email spam can reduce effectiveness of these filters (i.e. using these tricks to bypass anti-spam filters) (Hayati & Potdar 2009; Wang et al. 2007). Attractive words have been used more often in the subject line of Arabic spam than in English and mixed (contains Arabic and English texts) spam. English spam had more false statements (misleading subject line) in the subject line than Arabic and mixed spam. Image spam (text embedded in an image) was used more frequently in the Arabic spam than English and mixed spam. English spam contained more links than Arabic and mixed spam. The percentage of forged unsubscribe links was higher in Arabic spam than in English and mixed spam. English spam had more malicious attachments than Arabic and mixed spam. The percentage of mixed and Arabic spam sent from obfuscated or fake email addresses was higher than the percentage of English spam. These results suggest that a further development of the current anti-spam filters is needed, especially for non-English languages such as Arabic, to detect new tricks used in email spam.

The next chapter will conclude the research by revisiting the research aim and objectives and presenting the main findings. The research limitations, research implications and recommendations for future work for other spam issues in Saudi Arabia are also discussed in the next chapter.

Chapter 10: Research Conclusions, Limitations, Implications and Recommendations for Future Work

The final chapter presents the conclusions of this research. This chapter is organised as follows:

- Section 10.1: revisits the research aim and objectives.
- Section 10.2: provides the main findings and conclusions of this research.
- Section 10.3: provides the novelty of the research.
- Section 10.4: describes the research limitations.
- Section 10.5: discusses the research implications.
- Section 10.6: provides recommendations for future research work.

10.1 Revisiting the Research Aim and Objectives

The research aim was to understand the nature of email spam in Saudi Arabia, to investigate the awareness of email users about it and the efforts to combat it, and to provide possible suggestions to mitigate it. In order to meet the aim of the research, the following objectives were addressed:

- To investigate the awareness of public users and businesses about email spam, anti-spam filters and the efforts to combat it in Saudi Arabia.
- To investigate the nature of email spam (volume, languages and types) received by public users and businesses, and blocked by ISPs.
- To investigate the differences between Arabic and English email spam.
- To investigate how public users, businesses and ISPs deal with email spam.
- To investigate the effects of email spam on the performance of public users, businesses and ISPs.
- To investigate the anti-spam filters used by Saudi ISPs, and their evaluation of the effectiveness of these filters in detecting Arabic and English email spam.

- To propose a taxonomy, which includes most of anti-spam filters used to detect email spam, mostly in English; and then suggest which of these filters could be selected to produce new filters for Arabic email spam.
- To investigate the differences between spammers' tricks used in Arabic and English email spam to bypass anti-spam filters.

10.2 Main Research Findings and Conclusions

This section presents the main findings and conclusions to the research questions. The awareness of email users about spam, methods of combatting it, and efforts of government and ISPs to combat it was low. Email users revealed that there was a deficiency in the efforts provided by the Saudi Arabian Government and ISPs to educate them about spam. A high percentage of email users did not know what email spam was, how to deal with it or how to protect their computer from potential malicious attachments. Some users had been negatively affected by malicious programs such as viruses or trojans attached to spam. Further effort is needed by government or relevant agencies (e.g. ISPs) to educate people about email spam, effective ways to deal with it and the services ISPs provide to combat it. This could reduce the negative impact on employees (Dantin & Paynter 2005). The main impact of email spam on the performance of Saudi society was wasting employees' time in reading, deleting or isolating it from legitimate emails. This can cost companies a lot of money in reduced productivity and in turn affect the economic progress of the country.

There was a deficiency in the technical efforts provided by Saudi organisations, such as creating business units or teams to manage security, or employing particular employees to deal with spam problems. Additional efforts are required. A wide variety of email spam definitions have been used in Saudi Arabia and there was no specific definition that could be used in designing strategies and enacting laws against spammers. This suggests the need for an agreed definition of email spam in Saudi Arabia (Everett 2004). The volume of email spam in Saudi Arabia was lower than that revealed in developed countries such as the US and UK; however, this might be because users in Saudi Arabia received email spam without recognising it. This suggests that it is important for government to design strategies and policies to combat email spam at the early stages in order to be able to minimise its negative impact in Saudi Arabia.

Surprisingly, Saudi Arabia has its own spammers. The results showed that the Arabic was the second language most frequently used by spammers, after English, and the highest percentage of Arabic spam was sent from Saudi Arabia. The most frequent Arabic email spam received from Arabic spammers was related to forums, politics and religion; whereas pornography, phishing and fraud were observed mostly in English spam. Consequently, there is a need for anti-spam laws in Saudi Arabia. This could reduce the incidence of email spam by greatly reducing the spammers' ability to generate it without fear of penalty (Xu 2010; Yamakawa & Yoshiura 2010). Regional or international cooperation with other countries to trace spammers' sources and their activities also help reduce email spam in the country (Moustakas, Ranganathan & Duquenoy 2005).

Spammers used different tricks in the headers and bodies of Arabic, English and mixed email spam. Most Arabic spam (compared to mixed and English spam), appeared as text embedded in an image (image spam), and included attractive words in the subject line. More English had malicious content such as viruses, trojans or malware. The percentage of mixed and Arabic spam sent from obfuscated or fake email addresses was higher than the percentage of English spam. This could affect the effectiveness of anti-spam filters in detecting Arabic and English email spam, and indicates the need to develop the current filters to detect the tricks used by spammers. Several of the filters described in the taxonomy produced in this study were designed to combat email spam based on its header and body, mostly in English. Some of these filters (e.g. SVMs, boosting, maximum entropy, DT C4.5, and LVQ), had better performance in detecting English spam than other methods. These filters could be used by anti-spam developers to create new, more effective filters for detecting Arabic spam.

Saudi ISPs deployed different types of email spam filters, such as Iron Port (content-based) and blacklists (origin-based). They reported that these filters were not completely effective in detecting Arabic and English spam. This suggests that anti-spam developers need to spend further effort improving effectiveness with which existing spam filters detect Arabic and English spam.

Anti-spam filters were reported to be more effective in detecting English spam than Arabic spam. This suggests that efficient and effective filters for Arabic email spam could be built from SVMs, boosting, maximum entropy, DT C4.5 and LVQ methods. The taxonomy reveals that these methods have achieved a high level of effectiveness and accuracy in detecting English spam compared to other methods.

10.3 Novelty of the Research

Previous studies have investigated different aspects of email spam in different countries, such as Greece (Pallas & Patrikakis 2005), Bahrain (Al-A'ali 2007), the USA (Grimes, Hough & Signorella 2007) and Malaysia (Bujang & Hussin 2010). These aspects have included the volume of email spam, its languages, its types, the awareness of email users about it, and how they deal with it. This research, however, studies these same aspects of email spam in Saudi Arabia, a country in which it had not previously been investigated. In addition, this research investigates the differences between Arabic and English email spam, something that had not previously been investigated.

Another significant novelty of this research is the use of three validated questionnaires: one for public email users, one for businesses and one for ISPs. Some items in the three groups of questionnaires have not been used in previous studies and were developed for this study to cover the gap in the knowledge. The validity of these questionnaires was checked by academic faculty members who are experts in the field of information security.

Another novelty introduced in this research was investigating the differences between spammers’ tricks used in Arabic and English email spam to bypass anti-spam filters. This process was conducted by analysing the headers and bodies of a collection of Arabic, English and mixed language (Arabic and English) email spam received from Saudi public users, businesses and ISPs.

This was the first study to investigate the awareness of public users and businesses about email spam: how they define it, how they deal with it, and their awareness of the efforts of government and ISPs to combat it. This is another significant novelty of this research.

A proposed taxonomy that includes most of the anti-spam filters used to detect email spam is another novelty introduced in this research. This taxonomy was created and designed especially for this study and can help the researchers or other future developers to improve or produce new filters for email spam, or suggest ways to combat spammers' tricks.

10.4 Research Limitations

As with any research, certain limitations were unavoidable due to the constraints under which the research was conducted. This research has two limitations. First, the questionnaires for the three groups (public users, businesses and ISPs) were valid, but the reliability of the questionnaires was not examined due to the limited time frame of the study period. Second, the small sample size, especially for the ISP participants, might affect the results of this group.

10.5 Research Implications

The results of this research provide implications to the literature, practice and society. These implications are described in the following sections.

10.5.1 Implications of the Research for Literature

In the literature, a number of studies were found that have investigated the nature of email spam in different languages, the awareness of email users about spam, their attitude, dealings or experiences with it, its effects, and efforts to combat it in different countries. However, no previous studies could be found that investigated email spam issues in Saudi Arabia. Hence, this study provides the following valuable knowledge to the literature.

This is one of few studies that have investigated email spam in Arab countries, and it provides the literature with valuable knowledge about the nature of email spam in Saudi Arabia, including its volume, its languages, and how Saudi society defines it. It presents insights into the difference between types and sources of Arabic and English email spam, and the difference between spammers’ tricks used in Arabic and English email spam.

This study provides the literature with research findings about the awareness of public users and businesses about email spam, anti-spam filters, and efforts of the government and ISPs to combat it. This research contributes to the literature findings on how public email users, businesses and ISPs in Saudi Arabia deal with spam, and the effects of email spam on their performance.

The study findings add to the knowledge about anti-spam filters used by Saudi ISPs to block email spam in Saudi Arabia, the types of these filters, and their effectiveness in detecting Arabic and English email spam. This research work also provides a taxonomy of email spam filters that includes most of the filters used to combat email spam and their effectiveness in detecting email spam, mostly in English.

10.5.2 Implications of the Research for Practice

This section provides practical implications for government, decision-makers, ISPs and businesses in Saudi Arabia. The results of this study can be used by government or decision-makers in Saudi Arabia to design strategies or policies to combat email spam. These could include anti-spam laws, awareness programs to educate email users about email spam, and technical measures to block it. This study highlights the importance of anti-spam laws in reducing the volume of email spam and its effects in Saudi Arabia. Therefore, the government could do well to enact a culturally fit law against spam in Saudi Arabia, and to establish a management or department to follow up the implementation of anti-spam laws and to receive email users' complaints about spam (Rossow 2007). The government also needs to cooperate legally with ISPs to track spammers and their activities, and to collaborate with other regional or international countries to trace spammers' sources or origins (Butler 2003).

This research work highlights the significance of the awareness and education of public users and businesses about email spam, anti-spam filters and efforts to combat it in Saudi Arabia, in reducing its volume and its effects. The awareness programs could be achieved through workshops or training for public users and businesses about email spam and methods of combatting it (Refai & Nyanchama 2007). These programs could be conducted by ISPs or governmental sectors (e.g. educational sectors such as universities). The government or ISPs also need to conduct awareness campaigns about email spam and methods of combatting it by distributing brochures, or broadcasting information about email spam in different media such as newspapers and TV (Pfleeger & Bloom 2005).

The study findings highlight the importance of establishing business units (i.e. IT centres), or creating teams, to combat security threats, including email spam. Hence, businesses and ISPs need to create business units or teams to manage network security attacks (Vroom & von Solms 2004). It is necessary for businesses and ISPs to design clear policies or standards to control the use of email in the organisation, apply effective anti-spam filters, and update these filters regularly (Sorkin 2001). Saudi ISPs need to provide their customers (i.e. subscribers) and employees with free anti-spam filters to block email spam. These efforts could reduce the volume of email spam and its effects in Saudi Arabia.

The study outcomes show that technical measures are an important and effective way to combat email spam. Saudi ISPs believed that the existing anti-spam filters were not effective in detecting English and Arabic email spam, but these filters had better performance in detecting English spam than Arabic spam. Therefore, anti-spam developers need to improve these filters to be more effective in detecting English and Arabic email spam, or produce new filters for Arabic email spam. This study provides an English and Arabic email spam corpora, and a list of keywords observed in Arabic and English email spam that can help in the development of filters. This study presents a taxonomy of email spam filters, which includes many filters that have been proposed to detect spam, mostly in English. This could help future researchers or anti-spam filter developers in choosing or suggesting the appropriate filters for classifying Arabic email spam.

10.5.3 Implications of the Research for Saudi Society

This section presents implications for Saudi society. This study highlights the significance of the awareness of public users in combatting email spam. Hence, Saudi ISPs and businesses need to educate their employees and customers about email spam and methods of combatting it. In addition, public users need to increase their knowledge about email spam and effective methods to combat it by searching the internet or registering at workshops or training sessions about it. Examples of effective ways to dealing with it include contacting ISPs and notifying them about email spam that they receive, not publishing or adding email addresses to untrusted websites, not reading or responding to spam but directly deleting it, and not clicking onto unsubscribe links in the body of email spam.

This research highlights the importance of using anti-spam filters in combatting email spam. Therefore, public email users need to use anti-spam filters on their computers to protect from potential malicious attacks attached to email spam, and they need to know how to use them.

10.6 Recommendations for Future Research

This section describes some areas for future research. More work is required to investigate other types of spam in Saudi Arabia, such as web spam, image spam and SMS spam, and effective ways to combat them. In the absence of law against spam in Saudi Arabia, further research to bring a balanced and culturally fit anti-spam law is needed. Monitoring of spam levels and composition, in advance of any anti-spam law in Saudi Arabia, and then after its enactment, so that the impact of the law can be measured quantitatively.

Future work could include investigating effective ways to educate email users in public and private sectors, and educational sectors, about email spam, anti-spam filters and the efforts of government or relevant agencies to combat email spam in Saudi Arabia.

Three validated questionnaires were used to collect data from public users, businesses and ISPs. Future work is recommended to examine the reliability of these questionnaires. On the other hand, to be able to generalise the results of the study to the whole population, access to a bigger sample size is needed. It is recommended that future researchers use a large sample size which is representative of the population of Saudi email users.

Anti-spam filters were not completely effective in detecting Arabic and English email spam. Further research is needed to improve the performance of anti-spam filters in detecting Arabic and English spam. This could be achieved by testing the effectiveness of anti-spam filters in detecting Arabic and English spam, and then developing more effective anti-spam filters.

The existing anti-spam filters were more effective in detecting English email spam than Arabic email spam. Future work is required to produce more effective anti-spam filters for Arabic email spam. Part of the research would involve developing a taxonomy of email spam filters, which includes different methods of detecting email spam, mostly in English; and a listing of keywords and phrases used in Arabic spam. This could help in creating specific anti-spam filters for Arabic spam. The taxonomy of anti-spam filters presented in this research is clearly preliminary in nature, and non-exhaustive. A clear task for the future is to expand it with information about additional email spam filters and techniques, and to address any refinements that become apparent during that process.

References

Abdoh, M, Musa, M & Salman, N 2009, 'Detecting Spam by Weighting Message Words', *Journal of Arts and Sciences*, pp. 1-14.

Abdul-Muhmin, AG & Al-Abdali, O 2011, 'Adoption of online purchase by consumers in Saudi Arabia: an exploratory study', in *2nd Conference on Administrative Sciences, Dhahran, Saudi Arabia*, pp. 19-21.

Abu-Nimeh, S, Nappa, D, Xinlei, W & Nair, S 2008, 'Bayesian Additive Regression Trees-Based Spam Detection for Enhanced Email Privacy', in *Third International Conference on Availability, Reliability and Security, 2008, ARES 08*, pp. 1044-1051.

ACMA 2011, *Education and awareness*, acma, viewed 05 June 2012, <http://www.acma.gov.au/WEB/STANDARD/pc=PC_310310>.

Adam, JOD 2007, 'The Evolutionary Microcosm of Stock Spam', *Security & Privacy, IEEE*, vol. 5, no. 1, pp. 70-75.

Ahmed, T & Oppenheim, C 2006, 'Experiments to identify the causes of spam', in *Aslib proceedings*, vol. 58, pp. 156-178.

Al-A'ali, M 2007, *A Study of Email Spam and How to Effectively Combat It*, Webology, viewed 05 June 2012, <http://www.webology.org/2007/v4n1/a37.html>.

Al-Ghamdi, SM 2010, 'Mobily of UAE: penetrating Saudi Arabia–a global case study', *Journal for Global Business Advancement*, vol. 3, no. 4, pp. 295-312.

Al-Saggaf, Y 2004, 'The effect of online community on offline community in Saudi Arabia', *The Electronic Journal of Information Systems in Developing Countries*, vol. 16.

Al-Saggaf, Y & Williamson, K 2004, 'Online communities in Saudi Arabia: Evaluating the impact on culture through online semi-structured interviews', in *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, vol. 5.

Al-Somali, SA, Gholami, R & Clegg, B 2008, 'Internet banking acceptance in the context of developing countries: an extension of the technology acceptance model', in *European Conference on Management of Technology*.

Al-Tawil, KM 2001, 'The Internet in Saudi Arabia', *Telecommunications Policy*, vol. 25, no. 8, pp. 625-632.

Al-Majed, I, Murray, JJ & Maguire, A 2001, 'Prevalence of dental trauma in 5–6-and 12–14-year-old boys in Riyadh, Saudi Arabia', *Dental Traumatology*, vol. 17, no. 4, pp. 153-158.

Alazab, M, Layton, R, Broadhurst, R & Bouhours, B 2013, 'Malicious Spam Emails Developments and Authorship Attribution', in *Cybercrime and Trustworthy*

Computing Workshop (CTC), 2013 Fourth, pp. 58-68.

Aldossary, A, While, A & Barriball, L 2008, 'Health care and nursing in Saudi Arabia', *International nursing review*, vol. 55, no. 1, pp. 125-128.

Alepin, D-C 2004, 'Opting-Out: A Technical, Legal and Practical Look at the CAN-Spam Act of 2003', *Colum. JL & Arts*, vol. 28, p. 41.

AlGhamdi, R & Drew, S 2012, 'Seven Key Drivers to Online Retailing Growth in KSA', *arXiv preprint arXiv:1211.3148*.

Alhazmi, A & Nyland, B 2010, 'Saudi international students in Australia and intercultural engagement: A study of transitioning from a gender segregated culture to a mixed gender environment', *RMIT University. Melbourne, Australia*.

Ali, ABMS & Yang, X 2007, 'Spam Classification Using Adaptive Boosting Algorithm', in *6th IEEE/ACIS International Conference on Computer and Information Science, 2007, ICIS 2007*, pp. 972-976.

Aliaga, M & Gunderson, B 2000, *Interactive statistics*, Prentice Hall.

Alkahtani, H, Gardner-Stephen, P & Goodwin, R 2012, 'Email SPAM in Saudi Arabia and how do end users deal with it?', in *The International Conference on Computing, Networking and Digital Technologies (ICCNDT2012)*, Sanad, Bahrain, pp. 194-206.

Alkahtani, H, Goodwin, R & Gardner-Stephen, P 2012a, 'Combating of email SPAM by different businesses in Saudi Arabia', in *The 13th International Arab Conference on Information Technology (ACIT2012)*, Zarq, Jordan.

Alkahtani, H, Goodwin, R & Gardner-Stephen, P 2012b, 'A Comparative Study of the Perceptions of End Users in the Eastern, Western, Central, Southern and Northern Regions of Saudi Arabia about Email SPAM and Dealing with it', *International Journal of Cyber-Security and Digital Forensics (IJCSDF)*, vol. 1, no. 4, pp. 297-310.

Alkahtani, H, Goodwin, R & Gardner-Stephen, P 2013, 'The Key Findings of Surveys Related to Email SPAM and Methods of Combating it in Saudi Arabia', *International Journal of Emerging Technology and Advanced Engineering (IJETAE)*, vol. 3, no. 10, pp. 1-11.

Alkahtani, HS, Gardner-Stephen, P & Goodwin, R 2011, 'A taxonomy of email SPAM filters', in *The 12th International Arab Conference on Information Technology (ACIT2011)*, Riyadh, Saudi Arabia, pp. 351-356.

Alkahtani, HS, Goodwin, R & Gardner-Stephen, P 2011, 'Email SPAM related issues and methods of controlling used by ISPs in Saudi Arabia', in *The 12th International Arab Conference on Information Technology (ACIT2011)*, Riyadh, Saudi Arabia, pp. 344-351.

Allman, E 2003, 'Spam, Spam, Spam, Spam, Spam, the FTC, and Spam', *Queue*, vol. 1, no. 6, pp. 62-69.

Almeida, TA & Yamakami, A 2010, 'Content-based spam filtering', in *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-7.

Almeida, TA, Yamakami, A & Almeida, J 2009, 'Evaluation of Approaches for Dimensionality Reduction Applied with Naive Bayes Anti-Spam Filters', in *International Conference on Machine Learning and Applications, 2009, ICMLA '09*, pp. 517-522.

Alnajim, A & Munro, M 2009, 'An Evaluation of Users' Anti-Phishing Knowledge Retention', in *Information Management and Engineering, 2009. ICIME '09. International Conference on*, pp. 210-214.

Alongi, EA 2004, 'Has the US canned spam', *Ariz. L. Rev.*, vol. 46, p. 263.

Alperovitch, D, Judge, P & Krasser, S 2007, 'Taxonomy of Email Reputation Systems', in *27th International Conference on Distributed Computing Systems Workshops, 2007, ICDCSW '07*, pp. 27-27.

Altbach, PG 2004, 'Globalisation and the university: Myths and realities in an unequal world', *Tertiary Education & Management*, vol. 10, no. 1, pp. 3-25.

Amin, RM 2011, 'Detecting targeted malicious email through supervised classification of persistent threat and recipient oriented features', The George Washington University.

Andaker, KL, Davis, M, Fulmer, DR & Gibbon, JL 2006, *SOURCE-SPECIFIC ELECTRONIC MESSAGE ADDRESSING*, Google Patents.

Andreolini, M, Bulgarelli, A, Colajanni, M & Mazzoni, F 2005, 'HoneySpam: honeypots fighting spam at the source', paper presented to Proceedings of the Steps to Reducing Unwanted Traffic on the Internet on Steps to Reducing Unwanted Traffic on the Internet Workshop, Cambridge, MA.

Androutsopoulos, I, Koutsias, J, Chandrinos, KV & Spyropoulos, CD 2000, 'An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with personal e-mail messages', paper presented to Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, Athens, Greece.

Androutsopoulos, I, Paliouras, G, Karkaletsis, V, Sakkis, G, Spyropoulos, CD & Stamatopoulos, P 2000, 'Learning to Filter Spam E-Mail: A Comparison of a Naive Bayesian and a Memory-Based Approach', pp. 1-13.

Arutyunov, V 2013, 'Spam: Its past, present, and future', *Scientific and Technical Information Processing*, vol. 40, no. 4, pp. 205-211.

Attar, A, Rad, RM & Atani, RE 2013, 'A survey of image spamming and filtering techniques', *Artificial Intelligence Review*, vol. 40, no. 1, pp. 71-105.

Attenberg, J, Weinberger, K, Dasgupta, A, Smola, A & Zinkevich, M 2009, 'Collaborative Email-Spam Filtering with the Hashing Trick', in.

Attorney General's Chamber 2007, *SPAM Control Act*, viewed 16 January 2013, <http://statutes.agc.gov.sg/aol/search/display/view.w3p;ident=78a8b6af-11fa-40a9-961d-a1b61008d377;page=0;query=Id%3A%22b81d86b6-20e4-4467-93ed-9c3901c55e73%22%20Status%3Ainforce;rec=0#legis>.

Australian Communications & Media Authority 2006, *Australian SPAM laws*, ACMA, viewed 14 March 2010, <http://www.efa.org.au/Issues/Privacy/spam.html#acts>.

Awawdeh, SA & Tubaishat, A 2014, 'An Information Security Awareness Program to Address Common Security Concerns in IT Unit', in *Information Technology: New Generations (ITNG), 2014 11th International Conference on*, pp. 273-278.

Babbie, E 2012, *The practice of social research*, Cengage Learning.

Bailey, KD 1994, *Typologies and taxonomies: an introduction to classification techniques*, vol. 102, Sage.

Barroso, D 2007, 'Botnets-the silent threat', *European Network and Information Security Agency (ENISA)*, vol. 15, p. 171.

Basnet, R, Mukkamala, S & Sung, AH 2008, 'Detection of phishing attacks: A machine learning approach', in *Soft Computing Applications in Industry*, Springer, pp. 373-383.

Bell, J 2010, *Doing your research project*, McGraw-Hill International.

Bernik, I 2013, 'Information Warfare Effects on Businesses in Slovenia', in *Proceedings of the 7th European Computing Conference (ECC '13)*, Dubrovnik, Croatia.

Bindu, V & Thomas, C 2012, 'Spam war: Battling ham against spam', in *Industrial Electronics and Applications (ICIEA), 2012 7th IEEE Conference on*, pp. 691-696.

Blanzieri, E & Bryl, A 2008, 'A survey of learning-based techniques of email spam filtering', *Artificial Intelligence Review*, vol. 29, no. 1, pp. 63-92.

Blessing, LT & Chakrabarti, A 2009, *DRM, a design research methodology*, Springer.

Boneh, D 2004, *The Difficulties of Tracing Spam Email*, U.S. FTC.

Boykin, O & Roychowdhury, V 2004, 'Personal Email networks: an effective anti-spam tool', *Condensed Matter cond-mat/0402143*, pp. 1-10.

Brod, AC 2004, 'The Logic of Email Stamps'.

Bujang, YR & Hussin, H 2010, 'Spam e-mail: How Malaysian e-mail Users Deal with It?', *Online Special International Journal Issues*, vol. 2010, no. 43, pp. 1007-1013.

Bujang, YR & Hussin, H 2013, 'Should we be concerned with spam emails? A look at its impacts and implications', in *Information and Communication Technology for the Muslim World (ICT4M), 2013 5th International Conference on*, pp. 1-6.

Burgess, S 2002, *Managing information technology in small business: challenges and solutions*, IGI Global.

Burke, RR 2002, 'Technology and the customer interface: what consumers want in the physical and virtual store', *Journal of the academy of Marketing Science*, vol. 30, no. 4, pp. 411-432.

Butler, M 2003, 'Spam -- the meat of the problem', *Computer Law & Security Report*, vol. 19, no. 5, pp. 388-391.

Caliendo, M, Clement, M, Papies, D & Scheel-Kopeinig, S 2008, *The cost impact of spam filters: Measuring the effect of information system technologies in organizations*, IZA discussion papers.

Caliendo, M, Clement, M, Papies, D & Scheel-Kopeinig, S 2012, 'Research Note-The Cost Impact of Spam Filters: Measuring the Effect of Information System Technologies in Organizations', *Information Systems Research*, vol. 23, no. 3-part-2, pp. 1068-1080.

Carreras, X & Marquez, L 2001, 'Boosting Trees for Anti-Spam Email Filtering', paper presented to Proceedings of RANLP, 4th International Conference on Recent Advances in Natural Language Processing, Tzigov Chark, BG.

Cavana, R, Delahaye, BL & Sekeran, U 2001, *Applied business research: Qualitative and quantitative methods*, John Wiley & Sons Australia.

CDSI 2013, *Latest statistical Releases*, <http://www.cdsi.gov.sa/english/>.

Center for Democracy & Technology 2003, *Why Am I Getting All This Spam?*, viewed 06 June 2012, <https://www.cdt.org/pr_statement/cdt-releases-new-report-origins-spam>.

Chang, M-w, Yih, W-t & McCann, R 2008, 'Personalized spam filtering for gray mail', in *Proceedings of the Fifth Conference on Email and Anti-Spam (CEAS)*.

Chaudhary, A, Kolhe, S & Kamal, R 2013, 'Machine Learning Classification Techniques: A Comparative Study'.

Chejne, AG 2009, *The Arabic language: Its role in history*, U of Minnesota Press.

Chen, H, Zhan, Y & Li, Y 2010, 'The application of decision tree in Chinese email classification', in *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on*, vol. 1, pp. 305-308.

Cheng, TSL 2004, 'Recent international attempts to can spam', *Computer Law & Security Report*, vol. 20, no. 6, pp. 472-479.

Chigona, W, Bheekun, A, Späth, M, Derakhashani, S & Belle, J-PV 2005, 'Perceptions on SPAM in a South African context', paper presented to 5th International Business Information Management Conference, Cairo Egypt, 13th – 15th December.

Christina, V, Karpagavalli, S & Suganya, G 2010, 'Email spam filtering using supervised machine learning techniques', *International Journal on Computer Science and Engineering (IJCSE) Vol*, vol. 2, pp. 3126-3129.

Chuan, Z, Xianliang, L, Mengshu, H & Xu, Z 2005, 'A LVQ-based neural network anti-spam email approach', *SIGOPS Oper. Syst. Rev.*, vol. 39, no. 1, pp. 34-39.

Çıltık, A & Güngör, T 2008, 'Time-efficient spam e-mail filtering using< i> n</i>-gram models', *Pattern Recognition Letters*, vol. 29, no. 1, pp. 19-33.

CITC 2012, *Internet in Saudi Arabia*, viewed 30 March 2012, <http://www.isu.net.sa/saudi-internet/local-information/All-isps.htm>.

CITC 2014, *Contact Us*, viewed 10 July 2014, <http://www.citc.gov.sa/English/Pages/Contactus.aspx>.

Clark, J, Koprinska, I & Poon, J 2003a, 'Linger-a smart personal assistant for e-mail classification', in *Proceedings of the 13th International Conference on Artificial Neural Networks (ICANN'03), Istanbul, Turkey*, pp. 26-29.

Clark, J, Koprinska, I & Poon, J 2003b, 'A neural network based approach to automated e-mail classification', in *IEEE/WIC International Conference on Web Intelligence, 2003, WI 2003*, pp. 702-705.

Clayton, R 2004, 'Stopping Spam by Extrusion Detection', in *Proceedings of the First Conference on Email and Anti-Spam (CEAS)*.

Cohen, WW 1995, 'Fast effective rule induction', in *ICML*, vol. 95, pp. 115-123.

Cohen, WW 1996, 'Learning rules that classify e-mail', in *AAAI Spring Symposium on Machine Learning in Information Access*, vol. 18, p. 25.

Collis, J & Hussey, R 2003, *Business research*, Citeseer.

Computer Fraud & Security 2008, 'Russia spam level output surges', *Computer Fraud & Security*, vol. 2008, no. 3, pp. 4-5.

Computer Fraud and security 2004, 'The Prudential gets smart with spam', *Computer Fraud & Security*, vol. 2004, no. 6, p. 20.

Cook, D, Hartnett, J, Manderson, K & Scanlan, J 2006, 'Catching spam before it arrives: domain specific dynamic blacklists', paper presented to Proceedings of the 2006 Australasian workshops on Grid computing and e-research - Volume 54, Hobart, Tasmania, Australia.

Cormack, G & Lynam, T 2005, 'Spam corpus creation for TREC', in *Proceedings of Second Conference on Email and Anti-Spam CEAS*, vol. 2005, pp. 1-2.

Cormack, GV & Kolcz, A 2009, 'Spam filter evaluation with imprecise ground truth', paper presented to Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, Boston, MA, USA.

Cournane, A & Hunt, R 2004, 'An analysis of the tools used for the generation and prevention of spam', *Computers & Security*, vol. 23, no. 2, pp. 154-166.

Cova, M, Kruegel, C & Vigna, G 2008, 'There Is No Free Phish: An Analysis of" Free" and Live Phishing Kits', *WOOT*, vol. 8, pp. 1-8.

Cranor, LF & LaMacchia, BA 1998, 'Spam!', *Communications of the ACM*, vol. 41, no. 8, pp. 74-83.

Creswell, JW 2013, *Research design: Qualitative, quantitative, and mixed methods approaches*, Sage.

Cristianini, N & Shawe-Taylor, J 2000, *An introduction to support vector machines and other kernel-based learning methods*, Cambridge university press.

Crystal, D 2001, *Language and the Internet*, Cambridge University Press.

D'Ambra, D 2007, 'Killer spam: clawing at your door - [tech focus: email security]', *Information Professional*, vol. 4, no. 2, pp. 28-31.

Damiani, E, Vimercati, SDCd, Paraboschi, S & Samarati, P 2004, 'An Open Digest-based Technique for Spam Detection', San Francisco, CA, USA, <citeseer.ist.psu.edu/758790.html>.

Dantin, U & Paynter, J 2005, 'Spam in Email Inboxes', in *18th Annual Conference of the National Advisory Committee on Computing Qualifications*, Tauranga, New Zealand, pp. 33-38.

de Vaus, DA 2001, *Research design in social research*, Sage.

DeBarr, D & Wechsler, H 2009, 'Spam detection using clustering, random forests, and active learning', in *Sixth Conference on Email and Anti-Spam. Mountain View, California*.

DeBarr, D & Wechsler, H 2010, 'Using social network analysis for spam detection', in *Advances in Social Computing*, Springer, pp. 62-69.

Denzin, KN & Lincoln, YS 2009, 'Qualitative research', *Yogyakarta: PustakaPelajar*.

Denzin, NK & Lincoln, YS 2000, 'The discipline and practice of qualitative research', *Handbook of qualitative research*, vol. 2, pp. 1-28.

Dhinakaran, C, Jae Kwang, L & Nagamalai, D 2009, '"Reminder: please update your details": Phishing Trends', in *Networks and Communications, 2009. NETCOM '09. First International Conference on*, pp. 295-300.

Dhinakaran, C, Lee, JK & Nagamalai, D 2007a, 'Characterizing spam traffic and spammers', in *International Conference on Convergence Information Technology, 2007*, pp. 831-836.

Dhinakaran, C, Lee, JK & Nagamalai, D 2007b, 'An empirical study of spam and spam vulnerable email accounts', in *Future Generation Communication and Networking (FGCN 2007)*, vol. 1, pp. 408-413.

Dojkovski, S, Lichtenstein, S & Warren, MJ 2007, 'Fostering information security culture in small and medium size enterprises: an interpretive study in Australia'.

Dong, J, Cao, H, Liu, P & Ren, L 2006, 'Bayesian Chinese Spam Filter Based on Crossed N-gram', in *Sixth International Conference on Intelligent Systems Design and Applications, 2006, ISDA '06.*, vol. 3, pp. 103-108.

Drucker, H, Wu, D & Vapnik, VN 1999, 'Support vector machines for spam categorization', *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1048-1054.

Duan, Z, Chen, P, Sanchez, F, Dong, Y, Stephenson, M & Barker, JM 2012, 'Detecting spam zombies by monitoring outgoing messages', *Dependable and Secure Computing, IEEE Transactions on*, vol. 9, no. 2, pp. 198-210.

Dudley, J, Barone, L & While, L 2008, 'Multi-objective spam filtering using an evolutionary algorithm', in *Evolutionary Computation, 2008. CEC 2008.(IEEE World Congress on Computational Intelligence). IEEE Congress on*, pp. 123-130.

Dusse, S, Hoffman, P, Ramsdell, B, Lundblade, L & Repka, L 1998, 'S/MIME version 2 message specification', *RFC2311.*

Eggendorfer, T 2008, 'Methods to identify spammers', in *Proceedings of the 1st international conference on Forensic applications and techniques in telecommunications, information, and multimedia and workshop*, p. 7.

El-Halees, A 2009, 'Filtering Spam E-Mail from Mixed Arabic and English Messages: A Comparison of Machine Learning Techniques', *The International Arab Journal of Information Technology*, vol. 6, no. 1, pp. 52-59.

Engelberth, M, Gobel, J, Gorecki, C & Trinius, P 2009, 'Mail-Shake', in *20th International Workshop on Database and Expert Systems Application, 2009, DEXA '09.* , pp. 43-47.

Ergina, S, Gunala, ES, Yigita, H & Aydina, R 2011, 'Turkish anti-spam filtering using binary and probabilistic models', in *2nd World Conference on Information Technology (WCIT-2011)*, pp. 1-6.

Ermakova, L 2010, 'Spam and Phishing Detection in Various Languages', *International Journal "Information Technologies and Knowledge*, vol. 4, no. 3, pp. 216-232.

Esquivel, H, Akella, A & Mori, T 2010, 'On the effectiveness of IP reputation for spam filtering', in *Communication Systems and Networks (COMSNETS), 2010*

Second International Conference on, pp. 1-10.

Europa 2007, *Fight against spam, spyware and malicious software*, viewed 18 March 2013, <http://europa.eu/legislation_summaries/information_society/internet/l24189a_en.htm>.

Everett, C 2004, 'Stronger laws needed to stem spam', *Computer Fraud & Security*, vol. 2004, no. 1, pp. 1-2.

Fawcett, T 2003, 'In vivo spam filtering: a challenge problem for KDD', *ACM SIGKDD Explorations Newsletter*, vol. 5, no. 2, pp. 140-148.

Federal Trade Commission 2003, *False Claims in Spam*.

Feild, L, Pruchno, RA, Bewley, J, Lemay, EP & Levinsky, NG 2006, 'Using Probability vs. Nonprobability Sampling to Identify Hard-to-Access Participants for Health-Related Research Costs and Contrasts', *Journal of Aging and Health*, vol. 18, no. 4, pp. 565-583.

Fette, I, Sadeh, N & Tomasic, A 2006, *Learning to detect phishing emails*, DTIC Document.

Fogel, J & Raghupathi, V 2013, 'Spam E-mail Advertisements for Cosmetics/Beauty Products and Consumer Behavior', *Journal of Business Theory and Practice*, vol. 1, no. 1, p. p28.

Fogo, CE 2000, 'The Postman Always Rings 4,000 Times: New Approaches to Curb Spam, 18 J. Marshall J. Computer & Info. L. 915 (2000)', *The John Marshall Journal of Information Technology & Privacy Law*, vol. 18, no. 4, p. 2.

FreeCode 2013, *SPASTIC*, viewed 30 March 2013, <http://freecode.com/projects/spastic>.

Freschi, V, Seraghiti, A & Bogliolo, A 2006, 'Filtering Obfuscated Email Spam by means of Phonetic String Matching', in *ECIR*, pp. 505-509.

Frost, K & Udsen, H 2006, 'Anti spam regulation in Denmark', *Computer Law & Security Report*, vol. 22, no. 3, pp. 241-249.

Furnell, S, Gennatou, M & Dowland, P 2000, 'Promoting security awareness and training within small organisations', in *Proceedings of the 1st Australian Information Security Management Workshop, Deakin University, Geelong*.

Gansterer, W, Ilger, M, Lechner, P, Neumayer, R & Strauß, J 2005, 'Anti-spam methods—state of the art', *Institute of Distributed and Multimedia Systems, University of Vienna*.

Garcia, FD, Hoepman, J-H & Nieuwenhuizen, Jv 2004, 'SPAM Filters Analysis', paper presented to SEC.

Gardner-Stephen, P 2009, 'A Biologically Inspired Method of SPAM Detection', in

20th International Workshop on Database and Expert Systems Application, 2009, DEXA '09, pp. 53-56.

Gargiulo, F & Sansone, C 2008, 'Combining visual and textual features for filtering spam emails', in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1-4.

Gartner Group 1999, *ISPs and Spam: The Impact of Spam on Customer Retention and Acquisition*.

Ghiam, S & Pour, AN 2012, 'A Survey on Web SPAM Detection Mathods: Taxonomy', *International Journal of Network Security & Its Applications (IJNSA)*, vol. 4, no. 5, pp. 119-134.

Glickman, H 2005, 'The Nigerian "419" advance fee scams: prank or peril?', *Canadian Journal of African Studies/La Revue canadienne des études africaines*, vol. 39, no. 3, pp. 460-489.

Gliner, JA & Morgan, GA 2000, *Research methods in applied settings: An integrated approach to design and analysis*, Psychology Press.

Golbeck, J & Hendler, J 2004, 'Reputation Network Analysis for Email Filtering', paper presented to CEAS, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.6.1674>.

Goodman, J, Heckerman, D & Rounthwaite, R 2005, 'Stopping spam', *Scientific American*, vol. 292, no. 4, pp. 42-49.

Goodman, JT & Rounthwaite, R 2004, 'Stopping outgoing spam', paper presented to Proceedings of the 5th ACM conference on Electronic commerce, New York, NY, USA.

Goodson, P, McCormick, D & Evans, A 2001, 'Searching for Sexually Explicit Materials on the Internet: An Exploratory Study of College Students' Behavior and Attitudes', *Archives of Sexual Behavior*, vol. 30, no. 2, pp. 101-118.

Gopalakrishnan, V, Ganchev, P, Ranganathan, S & Bowser, R 2006, 'Rule learning for disease-specific biomarker discovery from clinical proteomic mass spectra', in *Data Mining for Biomedical Applications*, Springer, pp. 93-105.

Goweder, AM, Rashed, T, Elbekaie, AS & Alhamammi, HA 2008, 'An Anti-SPAM system using Artificial Neural Networks and Genetic Algorithms', in *Proceedings of the 2008 International Arab Conference on Information Technology*, pp. 1-8.

Goweder, AM, Rashed, TE, Ali S. & Alhamammi, HA 2008, 'An Anti-SPAM system using Artificial Neural Networks and Genetic Algorithms', in *Proceedings of the 2008 International Arab Conference on Information Technology*, pp. 1-8.

Graham, P 2003, 'Better bayesian filtering', in *Proceedings of the 2003 Spam Conference*, vol. 11, pp. 15-17.

Grimes, GA 2004, 'Issues with spam', *Computer Fraud & Security*, vol. 2004, no. 5, pp. 12-16.

Grimes, GA, Hough, MG & Signorella, ML 2007, 'Email end users and spam: relations of gender and age group to attitudes and actions', *Computers in Human Behavior*, vol. 23, no. 1, pp. 318-332.

Grossman, S 2004, 'Keepign Unwanted Donkeys and Elephants out of Your Inbox: The Case for Regulating Political Spam', *Berkeley Tech. LJ*, vol. 19, p. 1533.

Guzella, TS & Caminhas, WM 2009, 'A review of machine learning approaches to Spam filtering', *Expert Systems with Applications*, vol. 36, no. 7, pp. 10206-10222.

Gyongyi, Z & Garcia-molina, H 2004, 'Web Spam Taxonomy', paper presented to Adversarial Information Retrieval on the Web.

Haiyan, W, Runsheng, Z & Yi, W 2009, 'An Anti-spam Filtering System Based on the Naive Bayesian Classifier and Distributed Checksum Clearinghouse', in *Third International Symposium on Intelligent Information Technology Application, 2009, IITA 2009.*, vol. 1, pp. 128-131.

Halligan, P 2006, 'Caring for patients of Islamic denomination: critical care nurses' experiences in Saudi Arabia', *Journal of clinical nursing*, vol. 15, no. 12, pp. 1565-1573.

Hamel, A 2004, 'Will the CAN-SPAM Act of 2003 Finally Put a Lid on Unsolicited E-Mail', *New Eng. L. Rev.*, vol. 39, p. 961.

Hamner, M & Al-Qahtani, F 2009, 'Enhancing the case for Electronic Government in developing nations: A people-centric study focused in Saudi Arabia', *Government Information Quarterly*, vol. 26, no. 1, pp. 137-143.

Hanke, M & Hauser, F 2008, 'On the effects of stock spam e-mails', *Journal of Financial Markets*, vol. 11, no. 1, pp. 57-83.

Hansman, S & Hunt, R 2003, 'A taxonomy of network and computer attack methodologies', *Department of Computer Science and Software Engineering, University of Canterbury, Christchurch, New Zealand*, vol. 7.

Harris, E 2003, *The next step in the spam control war: Greylisting.*

Hassanein, K & Head, M 2007, 'Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping', *International Journal of Human-Computer Studies*, vol. 65, no. 8, pp. 689-708.

Hayati, P & Potdar, V 2008, 'Evaluation of spam detection and prevention frameworks for email and image spam: a state of art', paper presented to Proceedings of the 10th International Conference on Information Integration and Web-based Applications \& Services, Linz, Austria.

Hayati, P & Potdar, V 2009, 'Spammer and hacker, two old friends', in *3rd IEEE*

International Conference on Digital Ecosystems and Technologies, 2009, DEST '09, pp. 290-294.

Hayati, P, Potdar, V, Talevski, A, Firoozeh, N, Sarenche, S & Yeganeh, EA 2010, 'Definition of spam 2.0: New spamming boom', in *Digital Ecosystems and Technologies (DEST), 2010 4th IEEE International Conference on*, pp. 580-584.

He, J & Thiesson, B 2007, 'Asymmetric Gradient Boosting with Application to Spam Filtering', in *CEAS*.

Hermanson, S 2003, *Unsolicited Commercial Email (SPAM) and older persons online*, Data Digest Number 94, AARP Public Policy Institute.

Heron, S 2009, 'Technologies for spam detection', *Network Security*, vol. 2009, no. 1, pp. 11-15.

Hershkop, S & Stolfo, SJ 2004, 'Identifying spam without peeking at the contents', *Crossroads*, vol. 11, no. 2, pp. 3-3.

Heymann, P, Koutrika, G & Garcia-Molina, H 2007, 'Fighting spam on social web sites: A survey of approaches and future challenges', *Internet Computing, IEEE*, vol. 11, no. 6, pp. 36-45.

Hinde, S 2002, 'Spam, scams, chains, hoaxes and other junk mail', *Computers & Security*, vol. 21, no. 7, pp. 592-606.

Hinde, S 2003, 'Spam: the evolution of a nuisance', *Computers & Security*, vol. 22, no. 6, pp. 474-478.

Hird, S 2002, 'Technical solutions for controlling spam', *proceedings of AUUG2002*.

Hoanca, B 2006, 'How good are our weapons in the spam wars?', *Technology and Society Magazine, IEEE*, vol. 25, no. 1, pp. 22-30.

Hoshaw-Woodard, S 2001, *Description and comparison of the methods of cluster sampling and lot quality assurance sampling to assess immunization coverage*, Department of Vaccines and Biologicals, World Health Organization Geneva.

Hovold, J 2005, 'Naive Bayes Spam Filtering Using Word-Position-Based Attributes', paper presented to Conference on Email and Anti-Spam.

Hu, X & Mao, ZM 2007, 'Accurate real-time identification of IP prefix hijacking', in *Security and Privacy, 2007. SP'07. IEEE Symposium on*, pp. 3-17.

Huang, Y-y & Chou, C 2010, 'An analysis of multiple factors of cyberbullying among junior high school students in Taiwan', *Computers in Human Behavior*, vol. 26, no. 6, pp. 1581-1590.

Huddleston, R & Pullum, GK 2002, 'The Cambridge Grammar of English', *Language. Cambridge: Cambridge University Press*, pp. 1-23.

Hulten, G, Goodman, J & Rounthwaite, R 2004, 'Filtering spam e-mail on a global scale', paper presented to Proceedings of the 13th international World Wide Web conference on Alternate track papers &amp; posters, New York, NY, USA.

Hwang, H-G, Chen, R-F & Lee, J-M 2007, 'Measuring customer satisfaction with internet banking: an exploratory study', *International Journal of Electronic Finance*, vol. 1, no. 3, pp. 321-335.

IBM X-Force® 2011, *IBM X-Force® 2010 Trend and Risk Report* IBM Corporation, viewed 01 August 2014, <http://public.dhe.ibm.com/common/ssi/ecm/en/wgl03007usen/WGL03007USEN.PDF>.

Internet Research Task Force 2013, *Anti-Spam Research Group*, viewed 09 April 2013, <http://irtf.org/concluded/asrg>.

Islam, MR, Zhou, W, Guo, M & Xiang, Y 2009, 'An innovative analyser for multi-classifier e-mail classification based on grey list analysis', *Journal of network and computer applications*, vol. 32, no. 2, pp. 357-366.

'ISPs get tougher on spam', 2004, *Computer Fraud & Security*, vol. 2004, no. 9, pp. 3-3.

Jidiga, GR & Sammulal, P 2013, 'The need of awareness in cyber security with a case study', in *Computing, Communications and Networking Technologies (ICCCNT),2013 Fourth International Conference on*, pp. 1-7.

Jin, R, Liu, Y, Si, L, Carbonell, JG & Hauptmann, A 2003, 'A new boosting algorithm using input-dependent regularizer'.

Jin, X, Xu, A, Bie, R & Guo, P 2006, 'Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles', in *Data Mining for Biomedical Applications*, Springer, pp. 106-115.

John, JP, Moshchuk, A, Gribble, SD & Krishnamurthy, A 2009, 'Studying Spamming Botnets Using Botlab', in *NSDI*, vol. 9, pp. 291-306.

Johnson, DW & Koch, H 2006, 'Computer Security Risks in the Internet Era: Are Small Business Owners Aware and Proactive?', in *System Sciences, 2006. HICSS '06. Proceedings of the 39th Annual Hawaii International Conference on*, vol. 6, pp. 130b-130b.

Joseph, AB 2008, 'The effect of using the internet by the youth', paper presented to The thirteenth scientific conference for media ethics: theory and application, Cairo, Media College.

KACST 2014, *Contact Us*, viewed 10 July 2014, <http://www.kacst.edu.sa/en/contact/Pages/default.aspx>.

Kartaltepe, EJ & Xu, S 2006, 'Towards blocking outgoing malicious impostor emails', in *Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks*, pp. 657-661.

Kaspersky Lab 2012, *Saudi Arabia leads the region as a source of spam*, viewed 10 February 2013, <http://me.kaspersky.com/en/about/news/spam/2011/Saudi_Arabia_leads_the_region_as_a_source_of_spam>.

Kelly, N 2007, 'Image spam: the new email scourge', *McAfee, Inc*, vol. 3965.

Khong, DWK 2004, 'The problem of spam law: a comment on the Malaysian communications and multimedia commission's discussion paper on regulating unsolicited commercial messages', *Computer Law & Security Report*, vol. 20, no. 3, pp. 206-212.

Khong, WK 2001, 'Spam Law for the Internet', *Journal of Information, Law and Technology (JILT)*, vol. 3, pp. 1-19.

Khorsi, A 2007, 'An Overview of Content-Based Spam Filtering Techniques', *Informatica (Slovenia)*, pp. 269-277.

Kim, C & Hwang, K-B 2008, 'Naive bayes classifier learning with feature selection for spam detection in social bookmarking', *ECML PKDD Discovery Challenge 2008*, p. 32.

Kirkpatrick, A 2007, *World Englishes Paperback with Audio CD: Implications for International Communication and English Language Teaching*, Cambridge University Press.

Kitchenham, B & Pfleeger, SL 2002, 'Principles of survey research: part 5: populations and samples', *ACM SIGSOFT Software Engineering Notes*, vol. 27, no. 5, pp. 17-20.

Klimt, B & Yang, Y 2004, 'The enron corpus: A new dataset for email classification research', *Machine Learning: ECML 2004*, pp. 217-226.

Kohn, D 2002, *Enforceable spam identification and reduction system, and method thereof*, Google Patents.

Kołcz, A, Chowdhury, A & Alspector, J 2004, 'The impact of feature selection on signature-driven spam detection', in *Proceedings of the 1st Conference on Email and Anti-Spam (CEAS-2004)*.

Kosik, P, Ostrihon, P & Rajabiun, R 2009, 'Ipv6 and spam', in *Proceedings of the 2009 MIT Spam Conference*.

Kraft, J 2008, 'The Influence of the Oligopolistic Fringe on Economies of New EU Countries on the Example of the Czech Republic', *Inzinerine Ekonomika-Engineering Economics (5)*, pp. 48-53.

Krasser, S, Tang, Y, Gould, J, Alperovitch, D & Judge, P 2007, 'Identifying image spam based on header and file properties using C4. 5 decision trees and support vector machine learning', in *Information Assurance and Security Workshop, 2007. IAW'07. IEEE SMC*, pp. 255-261.

Krishnamurthy, B 2006, 'Unwanted traffic: Important problems, research approaches', in *Internet Architecture Board Workshop*.

Krishnaswamy, K, Sivakumar, AI & Mathirajan, M 2009, *Management research methodology: integration of principles, methods and techniques*, Pearson Education India.

Kumar, A 2005, *Phishing-A new age weapon*, Technical report, Open Web Application Secuirtry Project (OWASP).

Kumar, M 2009, 'Spam Email: The Unwanted Guest at Your Doorstep!', *IETE Technical Review*, vol. 26, no. 2, pp. 83-84.

Kumar, R, Punetha, M, Sharma, I & Bhattacharya, M 2014, 'Consignor is a Spammer', in *Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on*, pp. 631-634.

Lam, H-Y & Yeung, D-Y 2007, 'A Learning Approach to Spam Detection based on Social Networks', paper presented to Conference on Email and Anti-Spam, CEAS 2007.

Lambert, A 2003, 'Analysis of SPAM', Master of Science in Computer Science thesis, University of Dublin, Trinity College.

LaRose, R & Rifon, NJ 2007, 'Promoting i-safety: effects of privacy warnings and privacy seals on risk assessment and online privacy behavior', *Journal of Consumer Affairs*, vol. 41, no. 1, pp. 127-149.

Leavitt, N 2005, 'Instant Messaging: A new target for hackers', *Computer*, vol. 38, no. 7, pp. 20-23.

Leng, TK 2006, 'Singapore's multi-pronged strategy against spam', *Computer Law & Security Report*, vol. 22, no. 5, pp. 402-408.

Lev, A & Goldin, J 2006, 'Containing SPAM - The Local Challenge', paper presented to Virus Bulletin Conference.

Levy, E 2004, 'Criminals become tech savvy', *Security & Privacy, IEEE*, vol. 2, no. 2, pp. 65-68.

Li, F & Hsieh, M-H 2006, 'An Empirical Study of Clustering Behavior of Spammers and Group-based Anti-Spam Strategies', in *CEAS*.

Lieven, P, Scheuermann, B, Stini, M & Mauve, M 2007, 'Filtering Spam Email Based on Retry Patterns', paper presented to IEEE International Conference on Communications.

Lueg, CP 2005, 'From spam filtering to information retrieval and back: seeking conceptual foundations for spam filtering', *Proceedings of the American Society for Information Science and Technology*, vol. 42, no. 1.

Lugaresi, N 2004, 'European union vs. spam: a legal response', in *Proceedings of the*

First Conference on E-mail and Anti-Spam, pp. 1-8.

Lynam, TR, Cormack, GV & Cheriton, DR 2006, 'On-line spam filter fusion', in *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 123-130.

Madleňák, R 2006, 'BASIC ASPECTS OF SPAM'.

Mahdi, MDH, Rezaul, KM & Rahman, MA 2010, 'Credit fraud detection in the banking sector in UK: a focus on e-business', in *Digital Society, 2010. ICDS'10. Fourth International Conference on*, pp. 232-237.

Malcolm, J 2004, 'Australia's Stand on Spam', *AUUGN*, p. 39.

Margariti, S, Meletiou, G, Stergiou, E, Vasiliadis, D & Rizos, G 2007, 'Security Systems Consideration: A Total Security Approach', in *COMPUTATION IN MODERN SCIENCE AND ENGINEERING: Proceedings of the International Conference on Computational Methods in Science and Engineering 2007 (ICCMSE 2007): VOLUME 2, PARTS A and B*, vol. 963, pp. 954-958.

Martinkova, L 2008, 'Unsolicited Religious E-mail Researching New Context of Religious Communication', *Masaryk UJL & Tech.*, vol. 2, p. 113.

Mazidah, S & Burairah, H 2014, 'Profile of iCT innovaTiveness in Malaysian sMes froM serviCes seCTor based on Core iCT indiCaTors', *The Journal of Technology Management and Technopreneurship (JTMT)*, vol. 2, no. 1.

McCusker, R 2004, 'Spam: Nuisance or menace, prevention or cure?', *Journal 2002, cited in OECD*, p. 9.

Mehta, B, Nangia, S, Gupta, M & Nejdl, W 2008, 'Detecting image spam using visual features and near duplicate detection', in *Proceedings of the 17th international conference on World Wide Web*, pp. 497-506.

Méndez, JR, Cid, I, Glez-Peña, D, Rocha, M & Fdez-Riverola, F 2008, 'A comparative impact study of attribute selection techniques on Naïve Bayes spam filters', in *Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects*, Springer, pp. 213-227.

Metsis, V, Androutsopoulos, I & Paliouras, G 2006, 'Spam filtering with naive bayes-which naive bayes', in *Third conference on email and anti-spam (CEAS)*, vol. 17, pp. 28-69.

Meyer, TA & Whateley, B 2004, 'SpamBayes: Effective open-source, Bayesian based, email classification system', in *CEAS*.

Milletary, J & Center, CC 2005, 'Technical trends in phishing attacks', *Retrieved December*, vol. 1, p. 2007.

Ministry of Internal Affairs and Communication 2007, *Information and Communications Policy Site*, viewed 16 January 2013, <http://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/laws_dt03.html>.

Mirowski, L, Hartnett, J & Williams, R 2009, 'An RFID attacker behavior taxonomy', *Pervasive Computing, IEEE*, vol. 8, no. 4, pp. 79-84.

Mo, G, Zhao, W, Cao, H & Dong, J 2006, 'Multi-agent interaction based collaborative p2p system for fighting spam', in *Intelligent Agent Technology, 2006. IAT'06. IEEE/WIC/ACM International Conference on*, pp. 428-431.

Mohamed, OSM 2011, 'Trust and security of electronic banking services in Saudi commercial banks: Saudis versus Non Saudis opinions', *African Journal of Business Management*, vol. 5, no. 14, pp. 5524-5535.

Moustakas, E, Ranganathan, C & Duquenoy, P 2005, 'Combating Spam through Legislation: A Comparative Analysis of US and European Approaches', in *CEAS*, pp. 1-8.

Nagamalai, D, Dhinakaran, BC & Lee, JK 2010, 'An In-depth analysis of spam and spammers', *arXiv preprint arXiv:1012.1665*.

Nagamalai, D, Dhinakaran, C & Lee, J-K 2010, 'Novel mechanism to defend DDoS attacks caused by spam', *arXiv preprint arXiv:1012.0610*.

'New Bugbear targets banks', 2003, *Computer Fraud & Security*, vol. 2003, no. 7, p. 2.

Nguyen, T, Tran, Q & Nguyen, N 2008, 'Vietnamese spam detection based on language classification', in *HUT-ICCE 2008-2nd International Conference on Communications and Electronics*, p. 74.

Nickerson, R, Muntermann, J, Varshney, U & Isaac, H 2009, 'Taxonomy development in information systems: Developing a taxonomy of mobile applications', in *European Conference in Information Systems*.

Nickerson, RC, Muntermann, J & Varshney, U 2010, 'Taxonomy development in information systems: a literature survey and problem statement'.

Nickerson, RC, Varshney, U & Muntermann, J 2013, 'A method for taxonomy development and its application in information systems', *European Journal of Information Systems*, vol. 22, no. 3, pp. 336-359.

Nicosia, G 2004, *Artificial Immune Systems: Third International Conference, ICARIS 2004, Catania, Sicily, Italy, September 13-16, 2004, Proceedings*, vol. 3239, Springer.

Nielson, J, Aycock, J & de Castro, D 2008, 'Image spam—ASCII to the rescue!', in *Malicious and Unwanted Software, 2008. MALWARE 2008. 3rd International Conference on*, pp. 65-68.

Nishioka, D, Murayama, Y & Fujihara, Y 2012, 'Producing a Questionnaire for a User Survey on Anshin with Information Security for Users without Technical Knowledge', in *System Science (HICSS), 2012 45th Hawaii International Conference on*, pp. 454-463.

O'Brien, C & Vogel, C 2003, 'Spam filters: bayes vs. chi-squared; letters vs. words', paper presented to Proceedings of the 1st international symposium on Information and communication technologies, Dublin, Ireland.

Oda, T 2005, *A Spam-Detecting Artificial Immune System*.

Oda, T & White, T 2005, 'Immunity from spam: An analysis of an artificial immune system for junk email detection', in *Artificial Immune Systems*, Springer, pp. 276-289.

Okolica, JS, Peterson, GL & Mills, RF 2008, 'Using PLSI-U to detect insider threats by datamining e-mail', *International Journal of Security and Networks*, vol. 3, no. 2, pp. 114-121.

Onwuegbuzie, AJ & Collins, KM 2007, 'A Typology of Mixed Methods Sampling Designs in Social Science Research', *Qualitative Report*, vol. 12, no. 2, pp. 281-316.

Osterman Research Inc. 2008, 'How Spamhaus Cost-Effectively Eliminates Spam', *An Osterman Research White Paper*.

Oudot, L 2003, 'Fighting Spammers With Honeypots: Part 1'.

Özdemir, RS, St. Louis, KO & Topbaş, S 2011, 'Public attitudes toward stuttering in Turkey: Probability versus convenience sampling', *Journal of Fluency Disorders*, vol. 36, no. 4, pp. 262-267.

Özgür, L, Güngör, T & Gürgen, F 2004, 'Adaptive anti-spam filtering for agglutinative languages: a special case for Turkish', *Pattern Recognition Letters*, vol. 25, no. 16, pp. 1819-1831.

Pallas, AA & Patrikakis, CZ 2005, 'An Overview of Spam Phenomenon; and the Key Findings of a Survey for Spam in Greece', pp. 1-5.

Pantel, P & Lin, D 1998, 'SpamCop: A Spam Classification and Organization Program', in *Workshop on Learning for Text Categorization*.

Passari, A, Radmand, L & Batoie, Y 2013, 'Influence of adopting Amazon Cloud computing services in e-commerce in small and medium organization, Malaysia', *Business Management Quarterly Review*, vol. 4, no. 3 & 4, pp. 52-61.

Pathak, A, Hu, YC & Mao, ZM 2008, 'Peeking into Spammer Behavior from a Unique Vantage Point', *LEET*, vol. 8, pp. 1-9.

Paulson, LD 2004, 'Spam hits instant messaging', *Computer*, vol. 37, no. 4, p. 18.

Pérez-Díaz, N, Ruano-Ordás, D, Fdez-Riverola, F & Méndez, J 2012, 'Developing Anti-spam Filters Using Automatically Generated Rough Sets Rules', in *Management Intelligent Systems*, Springer, pp. 325-334.

Pfleeger, SL & Bloom, G 2005, 'Canning Spam: Proposed Solutions to Unwanted Email', *IEEE Security and Privacy*, vol. 3, no. 2, pp. 40-47.

Phelps, JE, Lewis, R, Mobilio, L, Perry, D & Raman, N 2004, 'Viral Marketing or Electronic Word-of-Mouth Advertising: Examining Consumer Responses and Motivations to Pass Along Email', *Journal of Advertising Research*, vol. 44, no. 04, pp. 333-348.

Polanski, PP 2008, 'Spam, spamdexing and regulation of Internet advertising', *International Journal of Intellectual Property Management*, vol. 2, no. 2, pp. 139-152.

Polz, D & Gansterer, WN 2009, 'Trustnet Architecture for E-mail Communication', in *20th International Workshop on Database and Expert Systems Application, 2009, DEXA '09*, pp. 48-52.

Ponterotto, JG 2005, 'Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science', *Journal of counseling psychology*, vol. 52, no. 2, p. 126.

Potashman, M 2006, 'International Spam Regulations & (and) Enforcement: Recommendations following the World Summit on the Information Society', *BC Int'l & Comp. L. Rev.*, vol. 29, p. 323.

Prakash, VV & O'Donnell, A 2005, 'Fighting Spam with Reputation Systems', *Queue*, vol. 3, no. 9, pp. 36-41.

Pressey, AD, Winklhofer, HM & Tzokas, NX 2009, 'Purchasing practices in small-to medium-sized enterprises: An examination of strategic purchasing adoption, supplier evaluation and supplier capabilities', *Journal of purchasing and supply management*, vol. 15, no. 4, pp. 214-226.

Provost, J 1999, 'Naıve-Bayes vs. Rule-Learning in Classification of Email', *University of Texas at Austin*.

Punch, KF 2013, *Introduction to social research: Quantitative and qualitative approaches*, Sage.

Puniškis, D, Laurutis, R & Dirmeikis, R 2006, 'An artificial neural nets for spam e–mail recognition', *Elektronika ir Elektrotechnika (Electronics and Electrical Engineering)*, vol. 5, no. 69, pp. 73-76.

Ramachandran, A & Feamster, N 2006, 'Understanding the network-level behavior of spammers', *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 4, pp. 291-302.

Ramachandran, A, Feamster, N & Vempala, S 2007, 'Filtering spam with behavioral blacklisting', paper presented to Proceedings of the 14th ACM conference on Computer and communications security, Alexandria, Virginia, USA.

Ramady, MA & Sohail, MS 2010, 'Assessing the role of family business in promoting economic growth: perspectives from Saudi Arabia', *International Journal of Entrepreneurship and Small Business*, vol. 10, no. 4, pp. 447-459.

Ramanathan, V & Wechsler, H 2012, 'Phishing website detection using Latent Dirichlet Allocation and AdaBoost', in *Intelligence and Security Informatics*

(ISI), 2012 IEEE International Conference on, pp. 102-107.

Raymond, ES 2005, *Bogofilter: A fast open source bayesian spam filters*.

Refai, M & Nyanchama, M 2007, 'Fighting E-mail Spam Epidemic'.

Ridzuan, F, Potdar, V & Talevski, A 2010, 'Factors involved in estimating cost of email spam', in *Computational Science and Its Applications–ICCSA 2010*, Springer, pp. 383-399.

RIT 2013, *WHY AM I RECEIVING THIS MESSAGE?*, viewed 29 October 2013, <http://www.rit.edu/security/content/alert10-24>.

Rogers, KM 2006, 'Viagra, viruses and virgins: A pan-Atlantic comparative analysis on the vanquishing of spam', *Computer Law & Security Report*, vol. 22, no. 3, pp. 228-240.

Rossow, C 2007, *Anti-spam measures of European ISPs/ESPs*, August.

Saeedian, MF & Beigy, H 2009, 'Dynamic classifier selection using clustering for spam detection', in *Computational Intelligence and Data Mining, 2009. CIDM '09. IEEE Symposium on*, pp. 84-88.

Sahami, M, Dumais, S, Heckerman, D & Horvitz, E 1998, 'A Bayesian Approach to Filtering Junk E-Mail', paper presented to Learning for Text Categorization: Papers from the 1998 Workshop, Madison, Wisconsin, <citeseer.ist.psu.edu/sahami98bayesian.html>.

Sait, SM & Al-Tawil, KM 2007, 'Impact of Internet usage in Saudi Arabia: a social perspective', *International Journal of Information Technology and Web Engineering (IJITWE)*, vol. 2, no. 2, pp. 81-115.

Sait, SM, Al-Tawil, KM, Khan, SA & Faheemuddin, M 2008, 'The Use and Effect of Internet on General Education in Saudi Arabia'.

Sakkis, G, Androutsopoulos, I, Paliouras, G, Karkaletsis, V, Spyropoulos, CD & Stamatopoulos, P 2003, 'A Memory-Based Approach to Anti-Spam Filtering for Mailing Lists', *Information Retrieval*, vol. 6, no. 1, pp. 49-73.

Sanz, EP, Gómez Hidalgo, JM & Cortizo Pérez, JC 2008, 'Email spam filtering', *Advances in Computers*, vol. 74, pp. 45-114.

Saraubon, K & Limthanmaphon, B 2009, 'Fast effective botnet spam detection', in *Computer Sciences and Convergence Information Technology, 2009. ICCIT'09. Fourth International Conference on*, pp. 1066-1070.

Sathiyaseelan, S & Filmore, P 2011, 'Network Security and User Awareness in IT Organisations', *Advances in Communications, Computing, Networks and Security Volume 8*, p. 216.

Saunders, ML & Thornhill, P 2004, 'A.(2007), Research Methods for Business Students', *Financial Times/Prentice Hall*.

Saunders, MN, Saunders, M, Lewis, P & Thornhill, A 2011, *Research methods for business students, 5/e*, Pearson Education India.

Schaub, MY 2002, 'Unsolicited email: Does Europe allow SPAM? The state of the art of the European legislation with regard to unsolicited commercial communications', *Computer Law & Security Report*, vol. 18, no. 2, pp. 99-105.

Schneider, K-M 2003, 'A comparison of event models for Naive Bayes anti-spam e-mail filtering', paper presented to Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics - Volume 1, Budapest, Hungary.

Schryen, G 2007, 'The impact that placing email addresses on the Internet has on the receipt of spam: An empirical analysis', *Computers & Security*, vol. 26, no. 5, pp. 361-372.

Schwartz, PM & Janger, EJ 2007, 'Notification of data security breaches', *Michigan Law Review*, pp. 913-984.

Sculley, D, Wachman, G & Brodley, CE 2006, 'Spam Filtering Using Inexact String Matching in Explicit Feature Space with On-Line Linear Classifiers', paper presented to Text REtrieval Conference.

Secker, A, Freitas, AA & Timmis, J 2003, 'AISEC: an artificial immune system for e-mail classification', in *The 2003 Congress on Evolutionary Computation, 2003, CEC '03*, vol. 1, pp. 131-138 Vol.131.

Seigneur, J-M, Dimmock, N, Bryce, C & Jensen, CD 2004, 'Combating spam with TEA (trustworthy email addresses)', in *Proceedings of the Second Annual Conference on Privacy, Security and Trust (PST'04)*, pp. 47-58.

Sharma, AK & Sahni, S 2011, 'A Comparative Study of Classification Algorithms for Spam Email Data Analysis', *International Journal on Computer Science and Engineering*, vol. 3, no. 5, pp. 1890-1895.

Sheu, J-J 2009, 'An Efficient Two-phase Spam Filtering Method Based on E-mails Categorization', *IJ Network Security*, vol. 9, no. 1, pp. 34-43.

Shrivastava, JN & Bindu, MH 2012, 'Trends, issues and challenges concerning spam mails', *International Journal of Information Technology and Computer Science (IJITCS)*, vol. 4, no. 8, p. 10.

Simon, M 2004, 'The CAN-SPAM Act of 2003: Is Congressional Regulation of Unsolicited Commercial E-Mail Constitutional?', *J. High Tech. L.*, vol. 4, pp. 85-231.

Simpson, P 2003, 'Spoofed Identities: Virus, Spam or Scam?', *Computer Fraud & Security*, vol. 2003, no. 10, pp. 6-8.

Sinha, S, Bailey, M & Jahanian, F 2008, 'Shades of grey: On the effectiveness of reputation-based “blacklists”', in *Malicious and Unwanted Software, 2008. MALWARE 2008. 3rd International Conference on*, pp. 57-64.

Sipior, JC, Ward, BT & Bonner, PG 2004, 'Should spam be on the menu?', *Commun. ACM*, vol. 47, no. 6, pp. 59-63.

Siponen, M & Stucke, C 2006, 'Effective anti-spam strategies in companies: An international study', in *System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on*, vol. 6, pp. 127c-127c.

Sirisanyalak, B & Sornil, O 2007, 'An artificial immunity-based spam detection system', in *IEEE Congress on Evolutionary Computation, 2007, CEC 2007*, pp. 3392-3398.

Smith, B 2008, 'A Storm (Worm) Is Brewing', *Computer*, vol. 41, no. 2, pp. 20-22.

Smith, R 2004, 'Eliminating the Spam From Your Internet Diet: the Possible Effects of the Unsolicited Commercial Electronic Mail Act of 2001 on Junk E-Mail', *Tex. Tech L. Rev.*, vol. 35, p. 411.

Šolić, K, Šebo, D, Jović, F & Ilakovac, V 2011, 'Possible Decrease of Spam in the Email Communication', *MIPRO2011*.

Sophos 2013, *Information on malware known as Ransomware*, viewed 29 October 2013, <http://www.sophos.com/en-us/support/knowledgebase/119006.aspx>.

Sophos Ltd 2012, *India Spews More Spam Than Ever Before as UK Returns to Dirty Dozen*, viewed 10 February 2013, <http://www.sophos.com/en-us/press-office/press-releases/2012/10/india-spews-more-spam-than-ever-before-as-uk-returns-to-dirty-dozen.aspx>.

Soranamageswari, M & Meena, C 2010, 'Statistical feature extraction for classification of image spam using artificial neural networks', in *Second International Conference on Machine Learning and Computing (ICMLC)*, pp. 101-105.

Sorkin, DE 2001, 'Technical and Legal Approaches to Unsolicited Electronic Mail', *University of San Francisco Law Review*, vol. 35, no. 2, pp. 325-384.

Sorkin, DE 2009, *SPAM LAWS*, The Center for Information Technology and Privacy Law, viewed 01 March 2010, <http://www.spamlaws.com/>.

Stepanikova, I & Zheng, L 2004, 'TEN YEARS AFTER THE BIRTH OF THE INTERNET, HOW DO AMERICANS USE THE INTERENT IN THEIR DAILY LIVES?'.

Stevens, PF 2003, 'History of Taxonomy', *eLS*.

Stolfo, SJ, Eskin, E, Herskop, S & Bhattacharyya, M 2010, *System and methods for detecting malicious email transmission*, Google Patents.

Stone-Gross, B, Holz, T, Stringhini, G & Vigna, G 2011, 'The underground economy of spam: A botmaster’s perspective of coordinating large-scale spam campaigns', in *USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*.

Subramaniam, T, Jalab, HA & Taqa, AY 2010, 'Overview of textual anti-spam filtering techniques', *International Journal of Physical Sciences*, vol. 5, no. 12, pp. 1869-1882.

Sujatha, R & krishna Rao, BR 2011, 'Taxonomy construction techniques–issues and challenges', *Indian Journal of Computer Science and Engineering (IJCSE)*, vol. 2, no. 5.

Sunner, M 2005, 'Email security best practice', *Network Security*, vol. 2005, no. 12, pp. 4-7.

Sweet, M 2003, 'MEDIA & COMMUNICATIONS: Political E-Mail: Protected Speech or Unwelcome Spam?', *Duke L. & Tech. Rev.*, vol. 2003, pp. 1-32.

Symantec 2010, *Annual Security Report*.

Symantec Corporation 2012, *Symantec Report Finds Spammers Are Taking Advantage of New Year Holidays and Major Events*, viewed 05 June 2012, <http://investor.symantec.com/phoenix.zhtml?c=89422&p=irol-newsArticle_print&ID=1653068&highlight=>.

Takemura, T & Ebara, H 2008, 'Spam Mail Reduces Economic Effects', in *Digital Society, 2008 Second International Conference on the*, pp. 20-24.

Takesue, M 2009, 'Personalized filtering of polymorphic e-mail spam', in *Emerging Security Information, Systems and Technologies, 2009. SECURWARE'09. Third International Conference on*, pp. 249-254.

Tansey, O 2007, 'Process tracing and elite interviewing: a case for non-probability sampling', *PS: Political Science & Politics*, vol. 40, no. 04, pp. 765-772.

Teddlie, C & Tashakkori, A 2003, *Handbook of mixed methods in social & behavioral research*, Sage.

Teddlie, C & Yu, F 2007, 'Mixed methods sampling a typology with examples', *Journal of mixed methods research*, vol. 1, no. 1, pp. 77-100.

Templeton, B 2004, 'Proper principles for challenge/response anti-spam systems', *Web site*.

Thompson, C 1999, 'If you could just provide me with a sample: examining sampling in qualitative and quantitative research papers', *Evidence based nursing*, vol. 2, no. 3, pp. 68-70.

Tive, C 2006, *419 scam: Exploits of the Nigerian con man*, iUniverse.

Tran, M & Armitage, G 2006, 'Mitigating Email Spam by Statistical Rejection of TCP Connections Using Recent Sender History', in *Australian Telecommunication Networks and Application Conference*.

Tseng, C-Y, Huang, J-W & Chen, M-S 2007, 'ProMail: using progressive email social network for spam detection', in *Advances in Knowledge Discovery and*

Data Mining, Springer, pp. 833-840.

Turban, E, Lee, JK, King, D, Liang, TP & Turban, D 2009, *Electronic commerce 2010*, Prentice Hall Press.

Ungar, LH & Foster, DP 1998, 'Clustering methods for collaborative filtering', in *AAAI Workshop on Recommendation Systems*, vol. 1.

Vaile, D 2004, 'Spam canned—new laws for Australia', *Internet Law Bulletin*, vol. 6, no. 9, pp. 113-115.

van Teijlingen, E & Hundley, V 2002, 'The importance of pilot studies', *Nursing Standard*, vol. 16, no. 40, pp. 33-36.

Viudes, FAAL 2011, 'WHEN E--MAIL CROSSES ROLE BOUNDARIES–EXPOSURE TO SPAM AND PROTECTION STRATEGIES', GRENOBLE ECOLE DE MANAGEMENT.

Von Ahn, L, Blum, M, Hopper, NJ & Langford, J 2003, 'CAPTCHA: Using hard AI problems for security', in *Advances in Cryptology—EUROCRYPT 2003*, Springer, pp. 294-311.

von Solms, SH 2005, 'Information Security Governance – Compliance management vs operational management', *Computers & Security*, vol. 24, no. 6, pp. 443-447.

Vroom, C & von Solms, R 2004, 'Towards information security behavioural compliance', *Computers & Security*, vol. 23, no. 3, pp. 191-198.

Wahsheh, H, Alsmadi, I & Al-Kabi, M 2012, 'Analyzing the Popular Words to Evaluate Spam in Arabic Web Pages', *IJJ: The Research Bulletin of JORDAN ACM–ISWSA. v2*, no. 2, pp. 22-26.

Waltz, CF, Strickland, O & Lenz, ER 2010, *Measurement in nursing and health research*, Springer Pub.

Wang, C-C & Chen, S-Y 2007, 'Using header session messages to anti-spamming', *Computers & Security*, vol. 26, no. 5, pp. 381-390.

Wang, C, Zhang, F, Li, F & Liu, Q 2010, 'Image spam classification based on low-level image features', in *Communications, Circuits and Systems (ICCCAS), 2010 International Conference on*, pp. 290-293.

Wang, Z, Josephson, W, Lv, Q, Charikar, M & Li, K 2007, 'Filtering image spam with near-duplicate detection', in *Proceedings of CEAS*.

Weaver, N, Paxson, V, Staniford, S & Cunningham, R 2003, 'A taxonomy of computer worms', in *Proceedings of the 2003 ACM workshop on Rapid malcode*, pp. 11-18.

Wei, C, Sprague, A, Warner, G & Skjellum, A 2008, 'Mining spam email to identify common origins for forensic application', in *Proceedings of the 2008 ACM symposium on Applied computing*, pp. 1433-1437.

Wesley-Smith, I 2006, 'A parallel artificial neural network implementation', in *Proceedings of The National Conference On Undergraduate Research*.

Wittel, GL & Wu, SF 2004, 'On Attacking Statistical Spam Filters', paper presented to Proceedings of the Conference on Email and Anti-Spam (CEAS), Mountain View, CA, USA, <http://www.ceas.cc/papers-2004/index.html>.

Wood, SK 2013, *Methods, systems, and computer program products for mitigating email address harvest attacks by positively acknowledging email to invalid email addresses*, Google Patents.

Worthington, RL & Whittaker, TA 2006, 'Scale development research a content analysis and recommendations for best practices', *The Counseling Psychologist*, vol. 34, no. 6, pp. 806-838.

Xie, M, Yin, H & Wang, H 2006, 'An effective defense against email spam laundering', paper presented to Proceedings of the 13th ACM conference on Computer and communications security, Alexandria, Virginia, USA.

Xie, Y, Yu, F, Achan, K, Panigrahy, R, Hulten, G & Osipkov, I 2008, 'Spamming botnets: signatures and characteristics', paper presented to Proceedings of the ACM SIGCOMM 2008 conference on Data communication, Seattle, WA, USA.

Xu, D 2010, 'Solutions to Spam', Hochschule Furtwangen.

Xu, KS, Kliger, M, Chen, Y, Woolf, PJ & Hero, A 2009, 'Revealing social networks of spammers through spectral clustering', in *Communications, 2009. ICC'09. IEEE International Conference on*, pp. 1-6.

Xu, Z, Wang, H-g & Shao, Z-z 2009, 'Evaluation of Image Spam Classification System Based on AHP', in *Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on*, pp. 1-4.

Yamakawa, D & Yoshiura, N 2010, 'Analysis of spam mail sent to Japanese mail addresses in the long term', in *Network Operations and Management Symposium (NOMS), 2010 IEEE*, pp. 833-836.

Yeh, Q-J & Chang, AJ-T 2007, 'Threats and countermeasures for information system security: A cross-industry study', *Information & Management*, vol. 44, no. 5, pp. 480-491.

Yeniman Yildirim, E, Akalp, G, Aytac, S & Bayram, N 2011, 'Factors influencing information security management in small- and medium-sized enterprises: A case study from Turkey', *International Journal of Information Management*, vol. 31, no. 4, pp. 360-365.

Yih, W-t, McCann, R & Kolcz, A 2007, *Improving spam filtering by detecting gray mail*.

Yoshida, K, Adachi, F, Washio, T, Motoda, H, Homma, T, Nakashima, A, Fujikawa, H & Yamazaki, K 2004, 'Density-based spam detector', paper presented to Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA.

Youn, S & McLeod, D 2007a, 'A Comparative Study for Email Classification', *Advances and Innovations in Systems, Computing Sciences and Software Engineering*, pp. 387-391.

Youn, S & McLeod, D 2007b, 'Efficient Spam Email Filtering using Adaptive Ontology', in *Fourth International Conference on Information Technology, 2007, ITNG '07*, pp. 249-254.

Youn, S & McLeod, D 2007c, 'Spam Email Classification using an Adaptive Ontology', *Journal of Software*, vol. 2, no. 3, pp. 43-55.

Yue, X, Abraham, A, Chi, Z-X, Hao, Y-Y & Mo, H 2007, 'Artificial immune system inspired behavior-based anti-spam filter', *Soft Computing*, vol. 11, no. 8, pp. 729-740.

Zaidan, AA, Ahmed, NN, Karim, HA, Alam, GM & Zaidan, BB 2011, 'Spam influence on business and economy: Theoretical and experimental studies for textual anti-spam filtering using mature document processing and naive Bayesian classifier', *African Journal of Business Management*, vol. 5, no. 2, pp. 596-607.

Zhang, L & Yao, T-s 2003, 'Filtering junk mail with a maximum entropy model', in *Proceeding of 20th international conference on computer processing of oriental languages (ICCPOL03)*, pp. 446-453.

Zhang, L, Zhu, J & Yao, T 2004, 'An evaluation of statistical spam filtering techniques', vol. 3, no. 4, pp. 243-269.

Zhao, Y & Zhang, Y 2008, 'Comparison of decision tree methods for finding active objects', *Advances in Space Research*, vol. 41, no. 12, pp. 1955-1959.

Zheleva, E, Kolcz, A & Getoor, L 2008, 'Trusting spam reporters: A reporter-based reputation system for email filtering', *ACM Transactions on Information Systems (TOIS)*, vol. 27, no. 1, p. 3.

Zhuang, L, Dunagan, J, Simon, DR, Wang, HJ & Tygar, JD 2008, 'Characterizing botnets from email spam records', paper presented to Proceedings of the 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats, San Francisco, California.

Zinman, A & Donath, JS 2007, 'Is Britney Spears Spam?', paper presented to CEAS, California USA, <http://dblp.uni-trier.de/db/conf/ceas/ceas2007.html#ZinmanD07>.

Zuo, H, Hu, W, Wu, O, Chen, Y & Luo, G 2009, 'Detecting image spam using local invariant features and pyramid match kernel', in *Proceedings of the 18th international conference on World wide web*, pp. 1187-1188.

Appendices

Appendix A: A Final Approval Letter of Flinders University of this Research

Flinders University and Southern Adelaide Health Service

SOCIAL AND BEHAVIOURAL RESEARCH ETHICS COMMITTEE

Research Services Office, Union Building, Flinders University
GPO Box 2100, ADELAIDE SA 5001
Phone: (08) 8201 3116
Email: human.researchethics@flinders.edu.au

FINAL APPROVAL NOTICE

Principal Researcher: Mr Hasan Alkahtani

Email: alka0022@flinders.edu.au

Address: School of Computer Science, Engineering & Mathematics

Project Title: Exploration of Email SPAM, with a focus on its effects and mitigation in the Kingdom of Saudi Arabia

Project No.: **5074** Final Approval Date: 21 December 2010 Approval Expiry Date: **9 November 2013**

The above proposed project has been **approved** on the basis of the information contained in the application, its attachments and the information subsequently provided.

If you have any outstanding permission letters (item D8), that may have been previously requested, please ensure that they are forwarded to the Committee as soon as possible. Additionally, for projects where approval has also been sought from another Human Research Ethics Committee (item G1), please be reminded that a copy of the ethics approval notice will need to be sent to the Committee on receipt.

In accordance with the undertaking you provided in your application for ethics approval for the project, please inform the Social and Behavioural Research Ethics Committee, giving reasons, if the research project is discontinued before the expected date of completion.

You are also required to report anything which might warrant review of ethical approval of the protocol. Such matters include:

- serious or unexpected adverse effects on participants;
- proposed changes in the protocol (modifications);
- any changes to the research team; and
- unforeseen events that might affect continued ethical acceptability of the project.

To modify/amend a previously approved project please either mail or email a completed copy of the Modification Request Form to the Executive Officer, which is available for download from http://www.flinders.edu.au/research/info-for-researchers/ethics/committees/social-and-behavioural-research-ethics-committee/notification-of-committee-decision.cfm.

Please ensure that any new or amended participant documents are attached to the modification request.

In order to comply with monitoring requirements of the *National Statement on Ethical Conduct in Human Research (March 2007)* an annual progress and/or final report must be submitted. A copy of the pro forma is available from http://www.flinders.edu.au/research/ info-for-researchers/ethics/committees/social-behavioural.cfm.

Your first report is due on **21 December 2011** or on completion of the project, whichever is the earliest. *Please retain this notice for reference when completing annual progress or final reports.* If an extension of time is required, please email a request for an extension of time, to a date you specify, to human.researchethics@flinders.edu.au before the expiry date.

aMather

Andrea Mather (formerly Jacobs)
Executive Officer
Social and Behavioural Research Ethics Committee
22 December 2010

c.c Dr Robert Goodwin, robert.goodwin@flinders.edu.au
Dr Paul Gardner-Stephen, paul.gardner-stephen@flinders.edu.au

Appendix B: A Letter of Introduction + A Public User Questionnaire (Arabic Version)

Flinders
UNIVERSITY

**School of Computer Science,
Engineering and Mathematics**

Room (358), Information Science &
Technology Building

GPO Box 2100
Adelaide SA 5001

Tel: (+61 8) 8201 3113
Fax: (+61 8) 8201 2904
Email: Robert.goodwin@flinders.edu.au

http://csem.flinders.edu.au/

CRICOS Provider No. 00114A

خطاب تعريف

عزيزي/ عزيزتي المحترم/ة

السلام عليكم ورحمه الله وبركاته ... وبعد

هذا الخطاب لتعريف الطالب : حسن بن شجاع القحطاني والذي تم ترشيحه لدرجة الدكتوراه في قسم علوم الحاسب الآلي , الهندسة والرياضيات في جامعة فلندرز في استراليا .
يقوم المرشح حالياً بإجراء الدراسات المتعلقة ببحثه بغرض الحصول على رسالة درجة الدكتوراه أو بغرض نشر أبحاث أخرى حول موضوع "استكشاف رسائل البريد الإلكتروني الإختراقية , مع التركيز على آثارها والتخفيف منها في المملكة العربية السعودية".
هو سيكون ممتن جداً لو تفضلتم بالتطوع بالمساعدة في هذا المشروع البحثي من خلال إكمال هذا الاستبيان الذي يغطي جوانب معينة من هذا البحث . هذا الاستبيان سيحقق في مشكلة رسائل البريد الإلكتروني الإختراقية وآثارها على مستخدمي البريد الإلكتروني في السعودية. هذا الاستبيان أيضاً سيحقق في فهم مستخدمي البريد الإلكتروني للرسائل الإختراقية وكيفية تعاملهم معها وأيضاً سيحقق في الجهود المبذولة لمكافحته في السعودية. إجابة هذا الاستبيان لن تأخذ من وقتكم الثمين أكثر من 30 دقيقة .
سيتم إرسال ملخص نتائج البحث عن طريق البريد الإلكتروني إلى المشاركين المهتمين بهذا البحث .
كن متأكداً من أن أي معلومات مقدمة منكم ستُعامل في سرية تامة وأنه لن يتم تحديد أو تمييز أياً من المشاركين في البحث بشكل فردي في نتائج الرسالة أو التقارير أو غيرها من الأبحاث المنشورة . يحق لك بالطبع التوقف عن المشاركة في هذا البحث في أي وقت تشاء كما يحق لك أيضاً رفض الإجابة عن أسئلة معينة في الاستبيان .
أيه استفسارات لديكم بشأن هذا البحث يجب أن توجه إلى العنوان المذكور أعلاه أو عن طريق الهاتف على (+61 8) 8201 3113 أو عن طريق الفاكس على 2904 8201 (8 61+) أو عن طريق إرسال إيميل إلى العنوان (Robert.goodwin@flinders.edu.au) .
أشكركم على اهتمامكم ومساعدتكم .

تقبلوا فائق تحياتي واحترامي

د. روبرت قودوين
قسم علوم الحاسب , الهندسة والرياضيات

لقد تمت الموافقة على هذا المشروع البحثي من لجنة أخلاقيات البحث بجامعة فلندرز (رقم المشروع: 5074). لمزيد من التفاصيل بشأن الموافقة الأخلاقية للمشروع , يمكن الاتصال بالمسؤول التنفيذي للجنة على الهاتف 3116 8201 (8 61+) أو عن طريق الفاكس على 2035 8201 (8 61+) أو عن طريق إرسال إيميل إلى العنوان *human.researchethics@flinders.edu.au*

inspiring achievement

- الجزء الأول : معلومات عامة

1. الجنس :

O ذكر

O أنثى

2. كم عمرك ؟

□

3. الجنسية :

O سعودي

O آخر □

4. ماهي اللغة التي تجيد التحدث بها؟ **يمكن أن تختار أكثر من خيار**

□ العربية

□ الانجليزية

□ آخر □

5. المستوى الأعلى من التعليم :

O الابتدائية **انتقل إلى السؤال رقم 7**

O المتوسطة **انتقل إلى السؤال رقم 7**

O الثانوية **انتقل إلى السؤال رقم 7**

O الدبلوم

O البكالوريوس

O الماجستير

O الدكتوراه

6.إذا كان مستوى التعليم من الفئات الأربع الأخيرة في **السؤال 5** , ماهو تخصصك الدراسي؟ **اختر خيار واحد فقط**

O التربية والتعليم

O علوم الحاسب وتقنية المعلومات

O العلوم الاجتماعية

O العلوم الفيزيائية والحيوية

O العلوم الصحية والطب

O آخر □

7. ما هو وضعك المهني الحالي ؟

O طالب **انتقل إلى السؤال رقم 9**

O موظف

8. إذا أنت موظف , ماهي طبيعة عملك ؟ **اختر خيار واحد فقط**

O تربوية

O صحية

O تقنية

O إدارية

O آخر □

- الجزء الثاني : طبيعة رسائل البريد الإلكتروني الإختراقية في السعودية , أضرارها على أداء مستخدمي الإيميل , وكيفية تعاملهم معها

9. كل شخص يعرف رسائل البريد الإلكتروني بشكل مختلف, **بأسلوبك الخاص** , كيف تعرف رسائل البريد الإلكتروني الإختراقية؟

▪ تعريف رسائل البريد الإلكتروني الإختراقية:

يمكن تعريف الرسالة الإختراقية على أنها ***" البريد الإلكتروني الغير مرغوب به و الذي يشتمل على محتوى تجاري أو غير تجاري والذي يتم إرساله بشكل عشوائي أو بشكل مباشر أو غير مباشر إلى مجموعة كبيرة من المستلمين بدون أخذ موافقتهم على إرسال هذا الإيميل وكذلك لا تربطهم أي علاقة مع المرسل".***

تأخذ الرسائل الإختراقية العديد من الأشكال , فمن الأمثلة لهذه الرسائل : الإعلانات الترويجية من القطاعات التجارية المختلفة , الرسائل البريدية من المجموعات الإخبارية والسياسية والأحزاب الدينية والمنتديات , الإعلانات لمجموعة واسعة من المنتجات والخدمات والتي تتضمن المنتجات الصحية مثل الترويج لدواء معين أو الاستشارات الطبية مثل إتباع برنامج حمية معين , الرسائل الرياضية , الألعاب الإلكترونية على الإنترنت.

كما يمكن استخدام الرسائل الإختراقية للخداع والتصيد وسرقة المعلومات الشخصية مثل اسم المستخدم وكلمة المرور وكذلك معلومات الحساب البنكية مثل أرقام البطاقات الإئتمانية.

▪ أمثلة للكلمات والعبارات المستخدمة في الرسائل الإختراقية:

1) ***" اضغط واربح" , " لقد ربحت" , " لقد ربحت مليون ريال سعودي"*** : تعتبر هذه العبارات أمثلة للإعلانات التجارية والخدمات.

2) ***"فياقرا" ,"حمية"*** : تعتبر أمثلة للإعلانات والتي تروج لمنتجات صحية وطبية.

3) ***"بيانات حسابك البنكي تحتاج إلى تحديث" , "معلوماتك الشخصية غير مكتملة"*** : تعتبر أمثلة للعبارات التي تستخدم في الخداع وسرقة المعلومات الشخصية.

10. هل كنت تعرف عن رسائل البريد الإلكتروني الإختراقية قبل قراءة هذا الاستبيان ؟

O نعم

O لا **انتقل إلى السؤال رقم 12**

11. إذا نعم , كيف عرفت عن رسائل البريد الإلكتروني الإختراقية ؟ **يمكن أن تختار أكثر من خيار**

□ مزودي خدمة الإنترنت (ISPs)

□ الإنترنت والمنتديات

□ وسائل البث الإذاعي و الإعلامي مثل الراديو, التلفزيون , الصحف , المجلات

□ الوزارات والهيئات الحكومية

□ من خلال الدراسة في التعليم العام أو التعليم الجامعي

□ آخر

12. هل تلقيت رسائل بريد إلكتروني إختراقية (سبام) ؟

O نعم

O لا

13. إذا نعم, متى كانت آخر مرة استلمت فيها رسائل بريد إلكترونية إختراقية؟

O الثلاث الأيام الماضية

O الأسبوع الماضي

O الشهر الماضي

O الثلاث أشهر الماضية

O الست أشهر الماضية

O التسع أشهر الماضية

O الاثني عشر شهر الماضية

14. إذا نعم , كم عدد رسائل البريد الإلكتروني الإختراقية المستلمة أسبوعياً ؟

رجاءً , قدّر العدد :

ملاحظة : السؤال التالي سوف يسألك عن تقدير النسبة المئوية التقديرية لكل لغة رسالة بريد إلكترونية إختراقية تم استلامها. مجموع النسب المئوية لجميع لغات رسائل السبام يجب أن يكون **100%**.
على سبيل المثال, لو كانت لغات رسائل السبام التي قمت باستلامها هي اللغة الانجليزية و العربية و التركية , إذن النسب المئوية التقريبية لكل لغة رسالة سبام مستلمة يمكن أن تقدّر كما في الجدول التالي :

	لغة رسائل البريد الإلكتروني الإختراقية	**النسبة المئوية**
✓	الانجليزية	% 20
✓	العربية	% 50
✓	لغة أخرى , رجاءً اذكر : **التركية**	% 30
	لغة أخرى , رجاءً اذكر:	%0
	المجموع	**% 100**

نلاحظ من الجدول السابق:
السبام الانجليزي (20%) + السبام العربي (50%) + السبام التركي (30%) = المجموع هو (**%100**)

15. ماهي لغة رسائل البريد الإلكتروني الإختراقية المستلمة أسبوعياً؟ **يمكن أن تختار أكثر من خيار**

	لغة رسائل البريد الإلكتروني الإختراقية	**النسبة المئوية**
	الانجليزية	**%**
	العربية	**%**
	لغة أخرى , رجاءً اذكر:	**%**
	لغة أخرى , رجاءً اذكر:	**%**
	لغة أخرى , رجاءً اذكر:	**%**
	لغات لا أستطيع تمييزها	**%**
	المجموع	**% 100**

16. إذا كانت لغة الرسائل الإختراقية هي **اللغة العربية** , ماهي أنواع الرسائل الإختراقية التي تم استلامها أسبوعياً؟ **يمكن أن تختار أكثر من خيار**

ملاحظة قبل أن تجيب على هذا السؤال: الرجاء تقدير النسبة المئوية التقريبية لكل خيار تختاره , مجموع النسب المئوية يجب أن يكون **100%**. انظر المثال في **سؤال 15** لمزيد من التفاصيل عن تقدير النسبة المئوية التقديرية.

	أنواع الرسائل الإختراقية العربية	**النسبة المئوية**
	الإعلانات التجارية	%
	رسائل من المجموعات الدينية والأطراف السياسية	%
	رسائل من المواقع الإباحية	%
	رسائل من المنتديات	%
	الإعلانات للمنتجات والخدمات	%
	الخداع والاحتيال أوسرقة المعلومات	%
	آخر:	%
	المجموع	**% 100**

17. إذا كانت لغة الرسائل الإختراقية هي **اللغة الإنجليزية** , ماهي أنواع الرسائل الإختراقية التي تم استلامها أسبوعياً؟ **يمكن أن تختار أكثر من خيار**

ملاحظة قبل أن تجيب على هذا السؤال: الرجاء تقدير النسبة المئوية التقريبية لكل خيار تختاره , مجموع النسب المئوية يجب أن يكون **100%**. انظر المثال في **سؤال 15** لمزيد من التفاصيل عن تقدير النسبة المئوية التقديرية.

	أنواع الرسائل الإختراقية الانجليزية	النسبة المئوية
	الإعلانات التجارية	%
	رسائل من المجموعات الدينية والأطراف السياسية	%
	رسائل من المواقع الإباحية	%
	رسائل من المنتديات	%
	الإعلانات للمنتجات والخدمات	%
	الخداع والاحتيال أوسرقة المعلومات	%
	آخر:	%
	المجموع	**% 100**

18. ماهو مزود بريدك الإلكتروني الرئيسي ؟ **اختر خيار واحد فقط**

O هوتميل
O ياهو
O جي ميل
O آخر

19. تقريباً , منذ متى وأنت تستخدم بريدك الإلكتروني ؟

O يوم / أيام
O أسبوع / أسابيع
O شهر / أشهر
O سنة / سنوات

ملاحظة: السؤال التالي سوف يسألك أن تقوم بوضع دائرة على الخيار المناسب لتعاملك مع رسائل البريد الإلكتروني الإختراقية.
على سبيل المثال , عندما أنا لا أقرأ الرسائل الإختراقية مطلقاً , في هذه الحالة أنا سأضع دائرة على الخيار " **أبداً** " في المقياس الموجود في الجدول التالي. إذا أنا أحياناً أقوم بقراءة السبام , أنا في هذه الحالة أقوم باختيار الخيار " **أحياناً**".

أقرأ البريد الإلكتروني كاملاً	[✓ **أبداً**]	**أحياناً**	**دائماً□**

20. ماذا تفعل عندما تستلم رسائل البريد الإلكتروني الإختراقية ؟
رجاءً ضع دائرة على الخيار المناسب لتعاملك مع رسائل البريد الإلكتروني الإختراقية

أقرأ البريد الإلكتروني الإختراقي (السبام) كاملاً	**أبداً**	**أحياناً**	**دائما□**
احذف البريد الإلكتروني الإختراقي (السبام) بدون قراءاته	**أبداً**	**أحياناً**	**دائماً□**

أتواصل مع مزود خدمة الانترنت وأبلغه عن رسائل السبام	**أبداً**	**أحياناً**	□**دائماً**

21. هل سبق لك أن استجبت بشكل مقصود لبعض العروض المقدمة من رسائل البريد الإلكتروني الإختراقية؟

O نعم

O لا   **انتقل إلى السؤال رقم 23**

22. إذا نعم , ماهي الفوائد التي اكتسبتها من تفاعلك مع رسائل البريد الإلكتروني الإختراقية؟ **يمكن أن تختار أكثر من خيار**

□ البيع والشراء

□ التعلم

□ المرح والترفيه

□ آخر

23. هل تأثرت سلبياً من رسائل البريد الإلكتروني الإختراقية؟

O نعم

O لا   **انتقل إلى السؤال رقم 25**

24. إذا نعم , ماهي تأثيرات رسائل البريد الإلكتروني الإختراقية عليك؟ **يمكن أن تختار أكثر من خيار**

□ سرقة المعلومات الشخصية مثل اسم المستخدم , كلمة السر , أرقام البطاقات الائتمانية

□ ضياع الوقت و تقليل الإنتاجية

□ ثقة أقل في استخدام البريد الإلكتروني

□ امتلاء صندوق وارد البريد الإلكتروني

□ إصابة جهاز الكمبيوتر بالفيروسات والديدان والبرامج الخبيثة

□ تأثيرات أخر ى , اذكر

25. هل تعرف عن برامج مكافحة رسائل البريد الإلكتروني الإختراقية أو ما تسمى (Anti-SPAM Filters)؟

O نعم

O لا   **انتقل إلى السؤال رقم 28**

26. إذا نعم , كيف عرفت عن هذه البرامج ؟ **يمكن أن تختار أكثر من خيار**

□ مزودي خدمة الإنترنت في السعودية (ISPs)

□ الإنترنت والمنتديات

□ وسائل البث الإذاعي و الإعلامي مثل الراديو, التلفزيون , الصحف , المجلات

□ الوزارات والهيئات الحكومية

□ من خلال الدراسة في التعليم العام أو التعليم الجامعي

□ آخر

27. إذا استخدمت برامج مكافحة رسائل البريد الإلكتروني الإختراقية , الرجاء تقدير فعاليتها في اكتشاف رسائل البريد الإلكتروني الإختراقية الانجليزية والعربية ؟

ملاحظة: الرجاء اختيار النسب المئوية المناسبة لفعالية التقنيات في اكتشاف الرسائل الإختراقية العربية والإنجليزية من الخيارات الموجودة أسفل أو تقدير نسب مئوية تقريبية أخرى بناءاً على رأيك

البرامج الحالية / النسبة المئوية	**0 %**	**25 %**	**50 %**	**75 %**	**100 %**
فعالية البرامج الحالية في اكتشاف الرسائل الإختراقية العربية	○	○	○	○	○
فعالية البرامج الحالية في اكتشاف الرسائل الإختراقية الإنجليزية	○	○	○	○	○

نسب مئوية أخرى , رجاءً قدر

- الجزء الثالث : وعي مستخدمي الإيميل عن الجهود المبذولة لمكافحة رسائل البريد الإلكتروني الإختراقية في السعودية

28. هل أنت مدرك وعلى علم بجهود حكومة المملكة العربية السعودية لمكافحة رسائل البريد الإلكتروني الإختراقية؟

O نعم

O لا **انتقل إلى السؤال رقم 30**

29. إذا نعم , ماهذه الجهود التي أنت على علم بها؟

30. هل أنت مدرك و على علم بجهود مزودي خدمة الإنترنت في المملكة العربية السعودية لمكافحة رسائل البريد الإلكتروني الإختراقية ؟

O نعم

O لا **انتقل إلى السؤال رقم 32**

31. إذا نعم , ماهذه الجهود التي أنت على علم بها؟

32. في اعتقادك ماهي الطرق الملائمة لمكافحة رسائل البريد الإلكتروني الإختراقية في السعودية؟

O تقنية مثل برامج أو مكونات مادية , رجاءً وضّح

O قانونية مثل قوانين جديدة , رجاءً وضّح

O آخر

33. رجاءً لا تتردد في إضافة أي شيء تعتقد أنه قد يكون ذا قيمة لهذا البحث :

- هل تريد نسخة من نتائج هذا الاستبيان؟

○ نعم
○ لا

- إذا نعم , رجاءً اكتب بريدك الإلكتروني :

- **اختياري :** نرغب في جمع مجموعة كبيرة من رسائل البريد الإلكتروني الإختراقية **(السبام)** سواء كانت عربية أو انجليزية وذلك لتحليلها لتحقيق بعض أهداف البحث. هذه المجموعة ممكن أن تستخدم في تحليل و فهم طرق وخدع من يقومون بإرسال هذه الرسائل (السبامرز) والذي بدوره سيساعد في تطوير الفلاتر الحالية أو إيجاد أخرى جديدة. هذه المجموعة ممكن أيضاً أن تساعد في اختبار هذه الرسائل على التقنيات الحالية مما سيؤدي إلى اكتشاف فعالية هذه التقنيات في حظر الرسائل الإختراقية العربية والانجليزية. إذا كنت قادراً على المساعدة الرجاء إرسال هذه الرسائل على الإيميل التالي: hasan.sh.ka@gmail.com
إذا كنت بحاجة إلى مزيد من التفاصيل حول هذا البحث , الرجاء الاتصال بنا على عنوان البريد الإلكتروني التالي : alka0022@flinders.edu.au

شكراً لكم لإكمال الاستبيان

Appendix C: A Letter of Introduction + A Public User Questionnaire (English Version)

Flinders
UNIVERSITY

School of Computer Science,
Engineering and Mathematics

Room (358), Information Science &
Technology Building

GPO Box 2100
Adelaide SA 5001

Tel: (+61 8) 8201 3113
Fax: (+61 8) 8201 2904
Email: Robert.goodwin@flinders.edu.au

http://csem.flinders.edu.au/

CRICOS Provider No. 00114A

LETTER OF INTRODUCTION

Dear Sir/Madam,

This letter is to introduce Mr Hasan Shojaa Alkahtani who is a PhD student in the School of Computer Science, Engineering and Mathematics at Flinders University.

He is undertaking research leading to the production of a thesis or other publications on the subject of "Exploration of Email SPAM, with a focus on its effects and mitigation in Saudi Arabia".

He would be most grateful if you would volunteer to assist in this project, by completing this questionnaire which covers certain aspects of this topic. This questionnaire will investigate email SPAM and its effects on email users in Saudi Arabia. It will also investigate the understanding of email SPAM by email users, how they deal with it, and the efforts to combat it in Saudi Arabia. No more than 30 minutes is required to complete the questionnaire.

A summary of the results will be sent by email to interested respondents.

Be assured that any information provided will be treated in the strictest confidence and none of the participants will be individually identifiable in the resulting thesis, report or other publications. You are, of course, entirely free to discontinue your participation at any time or to decline to answer particular questions.

Any enquiries you may have concerning this project should be directed to me at the address given above or by telephone on (+61 8) 8201 3113, by fax on (+61 8) 8201 2904 or by email to (Robert.goodwin@flinders.edu.au).

Thank you for your attention and assistance.

Yours sincerely

Dr. Robert Goodwin
Senior Lecturer
School of Computer Science, Engineering and Mathematics

This research project has been approved by the Flinders University Social and Behavioural Research Ethics Committee (Project Number: 5074). For more information regarding ethical approval of the project the Executive Officer of the Committee can be contacted by telephone on (+61 8) 8201 3116, by fax on (+61 8) 8201 2035 or by email human.researchethics@flinders.edu.au.

inspiring achievement

• Part 1: Demographic Information

1. Gender:

O Male
O Female

2. What is your age?

3. Nationality:

O Saudi
O Other

4. What language(s) do you speak? **You can choose more than one option**

□ Arabic
□ English
□ Other

5. Highest level of education:

O Primary school **Go to question 7**
O Intermediate school **Go to question 7**
O High school **Go to question 7**
O Diploma
O Bachelor
O Master
O PhD

6. If your level of education was in the last four categories of **question 5**, what was your major area of study: **Select one only**

O Education and Teaching
O Computer Science and Information Technology
O Social Sciences
O Physical and Biological Sciences
O Health Sciences and Medicine
O Other

7. What is your current work status:

O Student **Go to question 9**
O Employed

8. If you are employed, what is the nature of your work? **Select one only**

O Educational
O Medical
O Technical
O Management
O Other

• Part 2: The nature of email SPAM in Saudi Arabia, its effects on the performance of email users, and dealing with it

9. Everyone defines Email SPAM differently, **in your own words**, how would you define email SPAM?

- **Email SPAM definition:**

Email SPAM can be defined as ***"an unsolicited, unwanted, commercial or non-commercial email that is sent indiscriminately, directly or indirectly, to a large number of recipients without their permission and there is no relationship between the recipients and a sender"***.

Email SPAM has many forms such as promotional advertisements from businesses, religious groups, political parties, pornographic websites and forums, and advertisements for a wide variety of products and services including medical, sports and online gaming.

SPAM may also be used for phishing to obtain credit card numbers, usernames, passwords and other personal information.

- **Examples of words and phrases used in SPAM include:**

1) ***"CLICK and WIN", "YOU HAVE WON"*** and ***"YOU WON 1 MILLION DOLLARS"*** are examples for advertisements of businesses and services.
2) ***"VIAGRA"*** and ***"DIET"*** are examples for advertisements of medical and health products.
3) ***"YOUR ACCOUNT NEEDS TO BE UPDATED"*** and ***"INCOMPLETE PERSONAL INFORMATION"*** are examples for phishing.

10. Did you know about SPAM emails prior to reading this survey?
    O Yes
    O No **Go to question 12**

11. If yes, how do you know about SPAM emails? **You can choose more than one option**
    □ Internet Service Providers (ISPs)
    □ The internet and forums
    □ Broadcast media such as radio, TV, newspapers and magazines
    □ Government ministries and commissions
    □ Through school or university education
    □ Other

12. Have you received SPAM emails?
    O Yes
    O No

13. If yes, when was the last time you have received SPAM email?
    - O Last three days
    - O Last week
    - O Last month
    - O Last 3 months
    - O Last 6 months
    - O Last 9 months
    - O Last 12 months

14. If yes, how many SPAM emails do you receive on average weekly?
    Please, estimate □

Note: the following question will ask you to estimate the relative percentage for each language of email SPAM you have received. The percentages should add up to **100 %**.
For example, if the languages of email SPAM that I have received were English, Arabic and Turkish, the relative percentages for each language might be estimated as follows:

Language of email SPAM		Percentage
✓	English	20 %
✓	Arabic	50 %
✓	Other language, please state : **Turkish**	30 %
	Other language, please state :	0 %
Total		**100 %**

So, English SPAM (20%) + Arabic SPAM (50%) + Turkish SPAM (30%) = **100%**

15. What is the language of the SPAM email you receive on average weekly? **You can choose more than one option**

Language of email SPAM		Percentage
	English	%
	Arabic	%
	Other language, please state :	%
	Other language, please state :	%
	Other language, please state :	%
	Languages I do not recognise	%
Total		**100 %**

16. If the language of SPAM was **Arabic**, what types of SPAM have you received on average weekly? **You can choose more than one option**

Note before you answer this question: please estimate the relative percentage for each option you select, the percentages should total **100 %**. See the example in **question 15** for more explanation about estimating the relative percentage.

Type of Arabic email SPAM		Percentage
	Businesses advertisements	%
	Emails from religious groups and political parties	%
	Emails from pornographic websites	%
	Emails from forums	%
	Products and services advertisements	%
	Phishing and fraud	%
	Other:	%
Total		**100 %**

17. If the language of SPAM was **English**, what types of SPAM have you received on average weekly? **You can choose more than one option**

Note before you answer this question: please estimate the relative percentage for each option you select, the percentages should total **100 %**. See the example in **question 15** for more explanation about estimating the relative percentage.

Type of English email SPAM		Percentage
	Businesses advertisements	%
	Emails from religious groups and political parties	%
	Emails from pornographic websites	%
	Emails from forums	%
	Products and services Advertisements	%
	Phishing and fraud	%
	Other:	%
Total		**100 %**

18. Who is your principal email account provider? **Select one only**

O Hotmail
O Yahoo
O Gmail
O Other

19. Approximately, how long have you been using your email account?
    O ________ Days
    O ________ Weeks
    O ________ Months
    O ________ Years

Note: the following question will ask you to choose the appropriate option for your dealing with email SPAM.
For example, when I am not reading the SPAM email at all, I will circle the option "**Never**" in the scale in the following table. If I sometimes read SPAM, I will circle the option "**Sometimes**".

Read the entire email SPAM	[✓ **Never**]	**Sometimes**	**Always**

20. What do you do when you receive SPAM email?
    Please circle the appropriate option for your dealing with SPAM email

Read the entire email SPAM	Never	Sometimes	Always
Delete email SPAM without reading it	Never	Sometimes	Always
Contact with ISP and notify it about email SPAM	Never	Sometimes	Always

21. Have you ever purposely responded to an offer made by a SPAM email?
    O Yes
    O No **Go to question 23**

22. If yes, what benefits did you derive from SPAM emails? **You can choose more than one option**
    □ Purchasing and selling
    □ Learning
    □ Fun
    □ Other ________

23. Have you been affected negatively by email SPAM?
    O Yes
    O No **Go to question 25**

24. If yes, what was the impact of email SPAM? **You can choose more than one option**
    - □ Stealing personal information such as user name, password and credit card numbers
    - □ Losing time and reducing productivity
    - □ Less confidence in using the email
    - □ Filling email inbox
    - □ Computer was infected by a Virus, Worm or other malicious program
    - □ Other impacts: please list them,

25. Are you aware of Anti-SPAM filters?
    - O Yes
    - O No **Go to question 28**

26. If yes, how did you know about these filters? **You can choose more than one option**
    - □ Internet Service Providers (ISPs)
    - □ The internet and forums
    - □ Broadcast media such as radio, TV, newspapers and magazines
    - □ Government ministries and commissions
    - □ Through school or university education
    - □ Other

27. If you have used Anti-SPAM filters, please rate their effectiveness in detecting English and Arabic email SPAM?

Note: please choose the appropriate percentage for the effectiveness of current filters in detecting Arabic and English email SPAM from the following options or estimate other relative percentages based on your opinion

Current Filters\ Percentage	**0 %**	**25 %**	**50 %**	**75 %**	**100 %**
The effectiveness of current filters in detecting Arabic email SPAM	○	○	○	○	○
The effectiveness of current filters in detecting English email SPAM	○	○	○	○	○

Other percentages, please estimate

- **Part 3: The efforts to combat SPAM in Saudi Arabia and the awareness of public users about them**

28. Are you aware of efforts by the government in Saudi Arabia to combat SPAM?
    O Yes
    O No **Go to question 30**

29. If yes, what are the efforts you are aware of?

30. Are you aware of efforts by ISPs in Saudi Arabia to combat SPAM?
    O Yes
    O No **Go to question 32**

31. If yes, what are the efforts you are aware of?

32. In your opinion, what are the appropriate ways to combat SPAM in Saudi Arabia?
    O Technical such as software , hardware , please explain

O Legal such as new laws , please explain

O Other

33. Please feel free to add anything that you think may be of value to this research:

- Do you want a summary of the results of this survey?
  - O Yes
  - O No

- If yes, please provide your email address :

- **Optional:** We wish to collect large corpora of Arabic and English email SPAM to analyse them to achieve some research aims. These corpora could be used to analyse and understand methods and tricks used by spammers, which could help developers to improve the existing Anti-SPAM filters or produce new ones. These corpora could also help in testing SPAM emails on the current email SPAM detection methods which may lead to exploring the effectiveness of these methods in detecting Arabic and English email SPAM. If you are able to help, please send these messages to the following email: hasan.sh.ka@gmail.com

  If you need more explanation about this research, please contact us on the following email address: alka0022@flinders.edu.au

Thank you for completing the survey

Appendix D: A Letter of Introduction + A Business Questionnaire (Arabic Version)

Flinders
UNIVERSITY

School of Computer Science,
Engineering and Mathematics

Room (358), Information Science &
Technology Building

GPO Box 2100
Adelaide SA 5001

Tel: (+61 8) 8201 3113
Fax: (+61 8) 8201 2904
Email: Robert.goodwin@flinders.edu.au

http://csem.flinders.edu.au/

CRICOS Provider No. 00114A

خطاب تعريف

عزيزي/ عزيزتي المحترم/ة

السلام عليكم ورحمه الله وبركاته ... وبعد

هذا الخطاب لتعريف الطالب : حسن بن شجاع القحطاني والذي تم ترشيحه لدرجة الدكتوراه في قسم علوم الحاسب الآلي , الهندسة والرياضيات في جامعة فلندرز في استراليا .
يقوم المرشح حالياً بإجراء الدراسات المتعلقة ببحثه بغرض الحصول على رسالة درجة الدكتوراه أو بغرض نشر أبحاث أخرى حول موضوع "استكشاف رسائل البريد الإلكتروني الإختراقية , مع التركيز على آثارها والتخفيف منها في المملكة العربية السعودية".
هو سيكون ممتن جداً لو تفضلتم بالتطوع بالمساعدة في هذا المشروع البحثي من خلال إكمال هذا الاستبيان الذي يغطي جوانب معينة من هذا البحث. هذا الاستبيان سيحقق في مشكلة رسائل البريد الإلكتروني الإختراقية وآثارها على قطاعات الأعمال التجارية المختلفة في السعودية. هذا الاستبيان أيضاً سيحقق في فهم القطاعات التجارية للرسائل الإختراقية وكيفية تعاملهم معها وأيضاً سيحقق في الجهود المبذولة لمكافحته في السعودية. إجابة هذا الاستبيان لن تأخذ من وقتكم الثمين أكثر من 30 دقيقة .
سيتم إرسال ملخص نتائج البحث عن طريق البريد الإلكتروني إلى المشاركين المهتمين بهذا البحث .
كن متأكداً من أن أي معلومات مقدمة منكم ستُعامل في سرية تامة وأنه لن يتم تحديد أو تمييز أياً من المشاركين في البحث بشكل فردي في نتائج الرسالة أو التقارير أو غيرها من الأبحاث المنشورة . يحق لك بالطبع التوقف عن المشاركة في هذا البحث في أي وقت تشاء كما يحق لك أيضاً رفض الإجابة عن أسئلة معينة في الاستبيان .
أيه استفسارات لديكم بشأن هذا البحث يجب أن توجه إلى العنوان المذكور أعلاه أو عن طريق الهاتف على (8 61+) 8201 3113 أو عن طريق الفاكس على 2904 8201 (8 61+) أو عن طريق إرسال إيميل إلى العنوان (Robert.goodwin@flinders.edu.au) .

أشكركم على اهتمامكم ومساعدتكم.

تقبلوا فائق تحياتي واحترامي

د. روبرت قودوين
قسم علوم الحاسب , الهندسة والرياضيات

لقد تمت الموافقة على هذا المشروع البحثي من لجنة أخلاقيات البحث بجامعة فلندرز (رقم المشروع: 5074). لمزيد من التفاصيل بشأن الموافقة الأخلاقية للمشروع , يمكن الاتصال بالمسؤول التنفيذي للجنة على الهاتف 3116 8201 (8 61+) أو عن طريق الفاكس على 2035 8201 (8 61+) أو عن طريق إرسال إيميل إلى العنوان *human.researchethics@flinders.edu.au*

inspiring achievement

- الجزء الأول : معلومات عامة

1. في أي عام تأسست الشركة ؟ **(ملاحظة : يفضل أن يكون التاريخ بالميلادي)**

2. هل ترى حجم الشركة بأنه:
   - O صغير
   - O متوسط
   - O كبير

3. كم العدد التقريبي للموظفين في الشركة ؟

4. تقريباً , كم عدد العملاء الذين تتعامل معهم الشركة ؟

5. ماهي طبيعة نشاط الشركة ؟

6. هل شركتكم لديها بشكل واضح وحدة عمل أو فريق لإدارة أمن الشبكة؟
   - O نعم
   - O لا **انتقل إلى السؤال رقم 9**

7. إذا نعم , ماهي مسؤوليات هذه الوحدة أو هذا الفريق ؟

8. إذا نعم , تقريباً كم عدد الموظفين الذين يعملون في هذه الوحدة أو هذا الفريق ؟

9. هل هناك موظفين مسؤوليتهم المحددة هي مكافحة رسائل البريد الإلكتروني الإختراقية؟

O نعم

O لا **انتقل إلى السؤال رقم 11**

10. إذا نعم , ماهي مهامهم لمكافحة رسائل البريد الإلكتروني الإختراقية ؟

- الجزء الثاني : طبيعة رسائل البريد الإلكتروني الإختراقية في السعودية , أضرارها على أداء قطاعات الأعمال التجارية , وكيفية تعاملهم معها

11. كل شخص يعرف رسائل البريد الإلكتروني بشكل مختلف, **بأسلوبك الخاص** , كيف تعرف رسائل البريد الإلكتروني الإختراقية؟

- **تعريف رسائل البريد الإلكتروني الإختراقية:**

يمكن تعريف الرسالة الإختراقية على أنها ***" البريد الإلكتروني الغير مرغوب به و الذي يشتمل على محتوى تجاري أو غير تجاري والذي يتم إرساله بشكل عشوائي أو بشكل مباشر أو غير مباشر إلى مجموعة كبيرة من المستلمين بدون أخذ موافقتهم على إرسال هذا الإيميل وكذلك لا تربطهم أي علاقة مع المرسل".***

تأخذ الرسائل الإختراقية العديد من الأشكال , فمن الأمثلة لهذه الرسائل : الإعلانات الترويجية من القطاعات التجارية المختلفة , الرسائل البريدية من المجموعات الإخبارية والسياسية والأحزاب الدينية والمنتديات , الإعلانات لمجموعة واسعة من المنتجات والخدمات والتي تتضمن المنتجات الصحية مثل الترويج لدواء معين أو الاستشارات الطبية مثل إتباع برنامج حمية معين , الرسائل الرياضية , الألعاب الإلكترونية على الإنترنت.

كما يمكن استخدام الرسائل الإختراقية للخداع والتصيد وسرقة المعلومات الشخصية مثل اسم المستخدم وكلمة المرور وكذلك معلومات الحساب البنكية مثل أرقام البطاقات الإئتمانية.

- **أمثلة للكلمات والعبارات المستخدمة في الرسائل الإختراقية:**

1) ***" اضغط واربح" , " لقد ربحت" , " لقد ربحت مليون ريال سعودي"*** : تعتبر هذه العبارات أمثلة للإعلانات التجارية والخدمات.
2) ***"فياقرا" ,"حمية"*** : تعتبر أمثلة للإعلانات والتي تروج لمنتجات صحية وطبية.
3) ***"بيانات حسابك البنكي تحتاج إلى تحديث" , "معلوماتك الشخصية غير مكتملة"*** : تعتبر أمثلة للعبارات التي تستخدم في الخداع وسرقة المعلومات الشخصية.

12. هل كنت تعرف عن رسائل البريد الإلكتروني الإختراقية, وبرامج مكافحتها قبل قراءة هذا الاستبيان ؟

O نعم

O لا **انتقل إلى السؤال رقم 14**

13. إذا نعم , كيف عرفت عن رسائل البريد الإلكتروني الإختراقية وبرامج مكافحتها ؟ **يمكن أن تختار أكثر من خيار**

□ مزودي خدمة الإنترنت (ISPs)

□ الإنترنت والمنتديات

□ وسائل البث الإذاعي و الإعلامي مثل الراديو, التلفزيون , الصحف , المجلات

□ الوزارات والهيئات الحكومية

□ شركات ومنظمات أخرى

□ آخر

14. هل تلقيت رسائل بريد إلكتروني إختراقية (سبام) ؟

O نعم

O لا

15. إذا نعم, متى كانت آخر مرة استلمت فيها رسائل بريد إلكترونية إختراقية؟

O الثلاث الأيام الماضية

O الأسبوع الماضي

O الشهر الماضي

O الثلاث أشهر الماضية

O الست أشهر الماضية

O التسع أشهر الماضية

O الاثني عشر شهر الماضية

16. إذا نعم , كم عدد رسائل البريد الإلكتروني الإختراقية المستلمة أسبوعياً ؟

رجاءً , قدّر العدد :

ملاحظة : السؤال التالي سوف يسألك عن تقدير النسبة المئوية التقديرية لكل لغة رسالة بريد إلكترونية إختراقية تم استلامها. مجموع النسب المئوية لجميع لغات رسائل السبام يجب أن يكون **%100**.

على سبيل المثال, لو كانت لغات رسائل السبام التي قمت باستلامها هي اللغة الانجليزية و العربية و التركية , إذن النسب المئوية التقريبية لكل لغة رسالة سبام مستلمة يمكن أن تقدّر كما في الجدول التالي :

لغة رسائل البريد الإلكتروني الإختراقية		**النسبة المئوية**
✓	الانجليزية	% 20
✓	العربية	% 50
✓	لغة أخرى , رجاءً اذكر : **التركية**	% 30
	لغة أخرى , رجاءً اذكر:	%0
المجموع		**% 100**

نلاحظ من الجدول السابق:
السبام الانجليزي (20%) + السبام العربي (50%) + السبام التركي (30%) = المجموع هو (**%100**)

17. ماهي لغة رسائل البريد الإلكتروني الإختراقية المستلمة أسبوعياً؟ **يمكن أن تختار أكثر من خيار**

لغة رسائل البريد الإلكتروني الإختراقية		**النسبة المئوية**
	الانجليزية	**%**
	العربية	**%**
	لغة أخرى , رجاءً اذكر:	**%**
	لغة أخرى , رجاءً اذكر:	**%**
	لغة أخرى , رجاءً اذكر:	**%**
	لغات لا أستطيع تمييزها	**%**
المجموع		**% 100**

18. إذا كانت لغة الرسائل الإختراقية هي **اللغة العربية** , ماهي أنواع الرسائل الإختراقية التي تم استلامها أسبوعياً؟ **يمكن أن تختار أكثر من خيار**

ملاحظة قبل أن تجيب على هذا السؤال: الرجاء تقدير النسبة المئوية التقريبية لكل خيار تختاره , مجموع النسب المئوية يجب أن يكون **100%**. انظر المثال في **سؤال 17** لمزيد من التفاصيل عن تقدير النسبة المئوية التقديرية.

أنواع الرسائل الإختراقية العربية		**النسبة المئوية**
	الإعلانات التجارية	%
	رسائل من المجموعات الدينية والأطراف السياسية	%
	رسائل من المواقع الإباحية	%
	رسائل من المنتديات	%
	الإعلانات للمنتجات والخدمات	%
	الخداع والاحتيال أوسرقة المعلومات	%
	آخر:	%
المجموع		**% 100**

19. إذا كانت لغة الرسائل الإختراقية هي **اللغة الإنجليزية** , ماهي أنواع الرسائل الإختراقية التي تم استلامها أسبوعياً؟ **يمكن أن تختار أكثر من خيار**

ملاحظة قبل أن تجيب على هذا السؤال: الرجاء تقدير النسبة المئوية التقريبية لكل خيار تختاره , مجموع النسب المئوية يجب أن يكون **100%**. انظر المثال في **سؤال 17** لمزيد من التفاصيل عن تقدير النسبة المئوية التقديرية.

	أنواع الرسائل الإختراقية الانجليزية	**النسبة المئوية %**
	الإعلانات التجارية	%
	رسائل من المجموعات الدينية والأطراف السياسية	%
	رسائل من المواقع الإباحية	%
	رسائل من المنتديات	%
	الإعلانات للمنتجات والخدمات	%
	الخداع والاحتيال أوسرقة المعلومات	%
	آخر:	%
	المجموع	**% 100**

20. ماهي تأثيرات رسائل البريد الإلكتروني الإختراقية على الشركة ؟ **يمكن أن تختار أكثر من خيار**

- □ ضياع الوقت و تقليل الإنتاجية
- □ إنفاق الكثير من المال في شراء و تطبيق وتجديد الفلاتر المستخدمة لمكافحة الرسائل الإختراقية في الشركة
- □ كفاءة خادم البريد الإلكتروني في الشركة قلت بسبب استلام كمية الكبيرة من الرسائل الإختراقية
- □ إصابة أجهزة الكمبيوتر في الشركة بالفيروسات والديدان والبرامج الخبيثة
- □ تأثيرات أخر ى , اذكر

21. هل تستخدمون برامج مكافحة رسائل البريد الإلكتروني الإختراقية أو ما تسمى (Anti-SPAM Filters) لحظر أو منع رسائل البريد الإلكتروني الإختراقية ؟

O نعم

O لا **انتقل إلى السؤال رقم 23**

22. إذا كنتم تستخدمون برامج مكافحة رسائل البريد الإلكتروني الإختراقية ,الرجاء تقدير فعاليتها في اكتشاف رسائل البريد الإلكتروني الإختراقية الانجليزية والعربية ؟

ملاحظة : الرجاء اختيار النسب المئوية المناسبة لفعالية التقنيات في اكتشاف الرسائل الإختراقية العربية والإنجليزية من الخيارات الموجودة أسفل أو تقدير نسب مئوية تقريبية أخرى بناءاً على رأيك

البرامج الحالية / النسبة المئوية	0 %	25 %	50 %	75 %	100 %
فعالية البرامج الحالية في اكتشاف الرسائل الإختراقية العربية	○	○	○	○	○
فعالية البرامج الحالية في اكتشاف الرسائل الإختراقية الإنجليزية	○	○	○	○	○

نسب مئوية أخرى , رجاءً قدر

- الجزء الثالث : وعي قطاعات الأعمال التجارية عن الجهود المبذولة لمكافحة رسائل البريد الإلكتروني الإختراقية في السعودية

23. هل أنت مدرك وعلى علم بجهود حكومة المملكة العربية السعودية لمكافحة رسائل البريد الإلكتروني الإختراقية؟

O نعم

O لا **انتقل إلى السؤال رقم 25**

24. إذا نعم , ما هذه الجهود التي أنت على علم بها؟

25. هل أنت مدرك و على علم بجهود مزودي خدمة الإنترنت في المملكة العربية السعودية لمكافحة رسائل البريد الإلكتروني الإختراقية ؟

O نعم

O لا **انتقل إلى السؤال رقم 27**

26. إذا نعم , ما هذه الجهود التي أنت على علم بها؟

27. هل يوجد هناك جهود توعية مُقدمة من قبل الشركة للموظفين العاملين فيها وكذلك العملاء حول رسائل البريد الإلكتروني الإختراقية والطرق الملائمة لمكافحتها ؟

O نعم , رجاءً وضح

O لا

28. في اعتقادك ماهي الطرق الملائمة لمكافحة رسائل البريد الإلكتروني الإختراقية في السعودية؟

O تقنية مثل برامج أو مكونات مادية , رجاءً وضّح

O قانونية مثل قوانين جديدة , رجاءً وضّح

O آخر

29. رجاءً لا تتردد في إضافة أي شيء تعتقد أنه قد يكون ذا قيمة لهذا البحث :

- هل تريد نسخة من نتائج هذا الاستبيان؟

O نعم
O لا

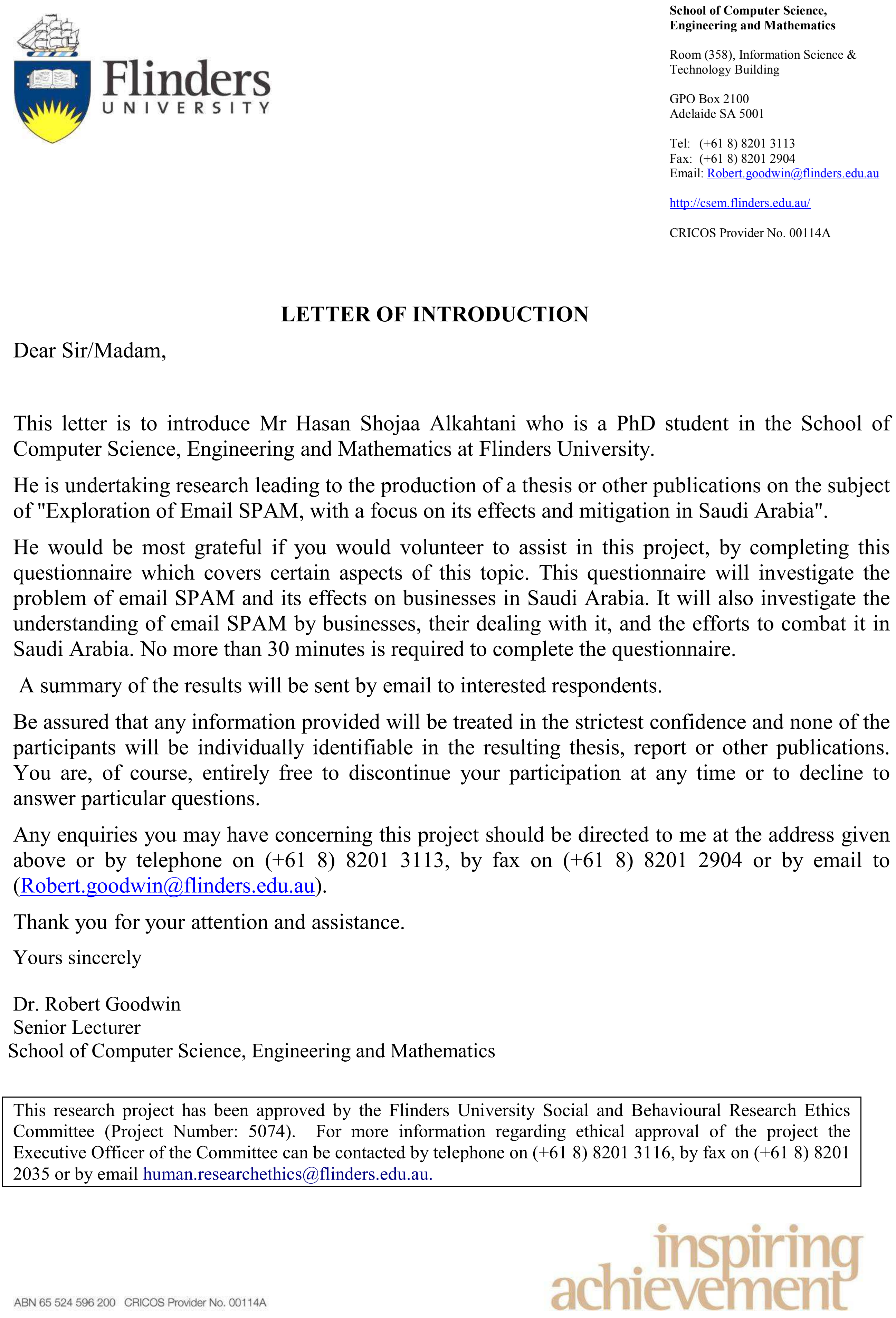
- إذا نعم , رجاءً اكتب بريدك الإلكتروني :

- **اختياري :** نرغب في جمع مجموعة كبيرة من رسائل البريد الإلكتروني الإختراقية **(السبام)** سواء كانت عربية أو انجليزية وذلك لتحليلها لتحقيق بعض أهداف البحث. هذه المجموعة ممكن أن تستخدم في تحليل و فهم طرق وخدع من يقومون بإرسال هذه الرسائل (السبامرز) والذي بدوره سيساعد في تطوير الفلاتر الحالية أو إيجاد أخرى جديدة. هذه المجموعة ممكن أيضاً أن تساعد في اختبار هذه الرسائل على التقنيات الحالية مما سيؤدي إلى اكتشاف فعالية هذه التقنيات في حظر الرسائل الإختراقية العربية والانجليزية. إذا كنت قادراً على المساعدة الرجاء إرسال هذه الرسائل على الإيميل التالي: hasan.sh.ka@gmail.com

إذا كنت بحاجة إلى مزيد من التفاصيل حول هذا البحث , الرجاء الاتصال بنا على عنوان البريد الإلكتروني التالي : alka0022@flinders.edu.au

شكراً لكم لإكمال الاستبيان

Appendix E: A Letter of Introduction + A Business Questionnaire (English Version)



Flinders
UNIVERSITY

School of Computer Science,
Engineering and Mathematics

Room (358), Information Science &
Technology Building

GPO Box 2100
Adelaide SA 5001

Tel: (+61 8) 8201 3113
Fax: (+61 8) 8201 2904
Email: Robert.goodwin@flinders.edu.au

http://csem.flinders.edu.au/

CRICOS Provider No. 00114A

LETTER OF INTRODUCTION

Dear Sir/Madam,

This letter is to introduce Mr Hasan Shojaa Alkahtani who is a PhD student in the School of Computer Science, Engineering and Mathematics at Flinders University.

He is undertaking research leading to the production of a thesis or other publications on the subject of "Exploration of Email SPAM, with a focus on its effects and mitigation in Saudi Arabia".

He would be most grateful if you would volunteer to assist in this project, by completing this questionnaire which covers certain aspects of this topic. This questionnaire will investigate the problem of email SPAM and its effects on businesses in Saudi Arabia. It will also investigate the understanding of email SPAM by businesses, their dealing with it, and the efforts to combat it in Saudi Arabia. No more than 30 minutes is required to complete the questionnaire.

A summary of the results will be sent by email to interested respondents.

Be assured that any information provided will be treated in the strictest confidence and none of the participants will be individually identifiable in the resulting thesis, report or other publications. You are, of course, entirely free to discontinue your participation at any time or to decline to answer particular questions.

Any enquiries you may have concerning this project should be directed to me at the address given above or by telephone on (+61 8) 8201 3113, by fax on (+61 8) 8201 2904 or by email to (Robert.goodwin@flinders.edu.au).

Thank you for your attention and assistance.

Yours sincerely

Dr. Robert Goodwin
Senior Lecturer
School of Computer Science, Engineering and Mathematics

This research project has been approved by the Flinders University Social and Behavioural Research Ethics Committee (Project Number: 5074). For more information regarding ethical approval of the project the Executive Officer of the Committee can be contacted by telephone on (+61 8) 8201 3116, by fax on (+61 8) 8201 2035 or by email human.researchethics@flinders.edu.au.

ABN 65 524 596 200 CRICOS Provider No. 00114A

inspiring
achievement

• Part 1: Demographic Information

1. What year was the company established?

2. Do you see the size of the company as being:
   O Small
   O Medium
   O Large

3. What is the approximate number of employees in the company?

4. Approximately how many customers does the company deal with?

5. What is the nature of your company activity?

6. Does your company have explicitly a business unit or team for managing network security?
   O Yes
   O No **Go to question 9**

7. If yes, what are the responsibilities of this unit or this team?

8. If yes, approximately how many employees are involved in this unit or this team?

9. Are there employees with specific responsibility for combating email SPAM?
   O Yes
   O No **Go to question 11**

10. If yes, what are their tasks to combat email SPAM?

- **Part 2: The nature of Email SPAM in Saudi Arabia, its effects on the performance of businesses, and dealing with it**

11. Everyone defines SPAM differently**, in your own words**, how would you define email SPAM?

▪ **Email SPAM definition:**

Email SPAM can be defined as ***“an unsolicited, unwanted, commercial or non-commercial email that is sent indiscriminately, directly or indirectly, to a large number of recipients without their permission and there is no relationship between the recipients and a sender"***.

Email SPAM has many forms such as promotional advertisements from businesses, religious groups, political parties, pornographic websites and forums, and advertisements for a wide variety of products and services including medical, sports and online gaming.

SPAM may also be used for phishing to obtain credit card numbers, usernames, passwords and other personal information.

▪ **Examples of words and phrases used in SPAM include:**

1) ***“CLICK and WIN ", “YOU HAVE WON”*** and ***“YOU WON 1 MILLION DOLLARS”*** are examples for advertisements of businesses and services.
2) ***“VIAGRA"*** and ***“DIET”*** are examples for advertisements of medical and health products.
3) ***“YOUR ACCOUNT NEEDS TO BE UPDATED"*** and ***“INCOMPLETE PERSONAL INFORMATION”*** are examples for phishing.

12. Did you know about SPAM emails and Anti-SPAM filters prior to reading this survey?
O Yes
O No **Go to question 14**

13. If yes, how do you know about SPAM emails and Anti-SPAM filters? **You can choose more than one option**
□ Internet Service Providers (ISPs)
□ The internet and forums
□ Broadcast media such as radio, TV, newspapers and magazines
□ Government ministries and commissions
□ Other companies and organisations
□ Other

14. Have you received SPAM emails?
O Yes
O No

15. If yes, when was the last time you have received SPAM email?
    - O Last three days
    - O Last week
    - O Last month
    - O Last 3 months
    - O Last 6 months
    - O Last 9 months
    - O Last 12 months

16. If yes, how many SPAM emails do you receive on average weekly?
    Please, estimate

Note: the following question will ask you to estimate the relative percentage for each language of email SPAM you have received. The percentages should add up to **100 %**.
For example, if the languages of email SPAM that I have received were English, Arabic and Turkish, the relative percentages for each language might be estimated as follows:

Language of email SPAM			Percentage
✓	English		20 %
✓	Arabic		50 %
✓	Other language, please state :	**Turkish**	30 %
	Other language, please state :		0 %
Total			**100 %**

So, English SPAM (**20%**) + Arabic SPAM (**50%**) + Turkish SPAM (**30%**) = **100%**

17. What is the language of the SPAM email you receive on average weekly? **You can choose more than one option**

Language of email SPAM		Percentage
	English	%
	Arabic	%
	Other language, please state :	%
	Other language, please state :	%
	Other language, please state :	%
	Languages I do not recognise	%
Total		**100 %**

18. If the language of SPAM was **Arabic**, what types of SPAM have you received on average weekly? **You can choose more than one option**

Note before you answer this question: please estimate the relative percentage for each option you select, the percentages should total **100 %**. See the example in **question 17** for more explanation about estimating the relative percentage.

Type of Arabic email SPAM		**Percentage**
	Businesses advertisements	%
	Emails from religious groups and political parties	%
	Emails from pornographic websites	%
	Emails from forums	%
	Products and services advertisements	%
	Phishing and fraud	%
	Other:	%
Total		**100 %**

19. If the language of SPAM was **English**, what types of SPAM have you received on average weekly? **You can choose more than one option**

Note before you answer this question: please estimate the relative percentage for each option you select, the percentages should total **100 %**. See the example in **question 17** for more explanation about estimating the relative percentage.

Type of English email SPAM		**Percentage**
	Businesses advertisements	%
	Emails from religious groups and political parties	%
	Emails from pornographic websites	%
	Emails from forums	%
	Products and services advertisements	%
	Phishing and fraud	%
	Other:	%
Total		**100 %**

20. What are the effects of email SPAM on the company? **You can choose more than one option**

□ Losing time and reducing productivity

□ Spending a lot of money to buy, implement and update filters or Anti-SPAM programs used in the company

□ The efficiency of organisation's email server was reduced due to the excessive email SPAM

□ Computers of the company were infected by a Virus, Worm or other malicious program

□ Other impacts: please list them,

21. Do you use Anti-SPAM filters to block email SPAM?
    O Yes
    O No **Go to question 23**

22. If you have used Anti-SPAM filters, please rate their effectiveness in detecting English and Arabic email SPAM?

Note: please choose the appropriate percentage for the effectiveness of current filters in detecting Arabic and English email SPAM from the following options or estimate other relative percentages based on your opinion.

Current Filters\ Percentage	**0 %**	**25 %**	**50 %**	**75 %**	**100 %**
The effectiveness of current filters in detecting Arabic email SPAM	○	○	○	○	○
The effectiveness of current filters in detecting English email SPAM	○	○	○	○	○

Other percentages, please estimate

- **Part 3: The efforts to combat SPAM in Saudi Arabia and the awareness of businesses about them**

23. Are you aware of efforts by the government in Saudi Arabia to combat SPAM?
    O Yes
    O No **Go to question 25**

24. If yes, what are the efforts you are aware of?

25. Are you aware of efforts by ISPs in Saudi Arabia to combat SPAM?
    O Yes
    O No **Go to question 27**

26. If yes, what are the efforts you are aware of?

27. Is there awareness provided by the company for employees and customers about SPAM and appropriate methods to combat it?
    - ○ Yes, please explain

    - ○ No

28. In your opinion, what are the appropriate ways to combat SPAM in Saudi Arabia?
    - O Technical such as software , hardware , please explain

O Legal such as new laws , please explain

Other

29. Please feel free to add anything that you think may be of value to this research:

- Do you want a summary of the results of this survey?
  - O Yes
  - O No
- If yes, please provide your email address :

- **Optional:** We wish to collect large corpora of Arabic and English email SPAM to analyse them to achieve some research aims. These corpora could be used to analyse and understand methods and tricks used by spammers, which could help developers to improve the existing Anti-SPAM filters or produce new ones. These corpora could also help in testing SPAM emails on the current email SPAM detection methods which may lead to exploring the effectiveness of these methods in detecting Arabic and English email SPAM. If you are able to help, please send these messages to the following email: hasan.sh.ka@gmail.com

  If you need more explanation about this research, please contact us on the following email address: alka0022@flinders.edu.au

Thank you for completing the survey

Appendix F: A Letter of Introduction + An ISP Questionnaire (Arabic Version)

Flinders
UNIVERSITY

School of Computer Science,
Engineering and Mathematics

Room (358), Information Science &
Technology Building

GPO Box 2100
Adelaide SA 5001

Tel: (+61 8) 8201 3113
Fax: (+61 8) 8201 2904
Email: Robert.goodwin@flinders.edu.au

http://csem.flinders.edu.au/

CRICOS Provider No. 00114A

خطاب تعريف

عزيزي/ عزيزتي المحترم/ة

السلام عليكم ورحمه الله وبركاته ... وبعد

هذا الخطاب لتعريف الطالب : حسن بن شجاع القحطاني والذي تم ترشيحه لدرجة الدكتوراه في قسم علوم الحاسب الآلي , الهندسة والرياضيات في جامعة فلندرز في استراليا .

يقوم المرشح حالياً بإجراء الدراسات المتعلقة ببحثه بغرض الحصول على رسالة درجة الدكتوراه أو بغرض نشر أبحاث أخرى حول موضوع "استكشاف رسائل البريد الإلكتروني الإختراقية , مع التركيز على آثارها والتخفيف منها في المملكة العربية السعودية".

هو سيكون ممتن جداً لو تفضلتم بالتطوع بالمساعدة في هذا المشروع البحثي من خلال إكمال هذا الاستبيان الذي يغطي جوانب معينة من هذا البحث . هذا الاستبيان سيحقق في مشكلة رسائل البريد الإلكتروني الإختراقية وآثارها على مزودي خدمة الإنترنت في السعودية. هذا الاستبيان أيضاً سيحقق في الجهود المبذولة لمكافحة الرسائل الإختراقية إضافةً إلى التحقيق في فعالية التقنيات الحالية في اكتشاف رسائل البريد الإلكتروني الإختراقية العربية والانجليزية . إجابة هذا الاستبيان لن تأخذ من وقتكم الثمين أكثر من 30 دقيقة.

سيتم إرسال ملخص نتائج البحث عن طريق البريد الإلكتروني إلى المشاركين المهتمين بهذا البحث .

كن متأكداً من أن أي معلومات مقدمة منكم ستُعامل في سرية تامة وأنه لن يتم تحديد أو تمييز أياً من المشاركين في البحث بشكل فردي في نتائج الرسالة أو التقارير أو غيرها من الأبحاث المنشورة . يحق لك بالطبع التوقف عن المشاركة في هذا البحث في أي وقت تشاء كما يحق لك أيضاً رفض الإجابة عن أسئلة معينة في الاستبيان .

أيه استفسارات لديكم بشأن هذا البحث يجب أن توجه إلى العنوان المذكور أعلاه أو عن طريق الهاتف على (+61 8) 8201 3113 أو عن طريق الفاكس على (+61 8) 8201 2904 أو عن طريق إرسال إيميل إلى العنوان (Robert.goodwin@flinders.edu.au) .

أشكركم على اهتمامكم ومساعدتكم .

تقبلوا فائق تحياتي واحترامي

د. روبرت قودوين
قسم علوم الحاسب , الهندسة والرياضيات

لقد تمت الموافقة على هذا المشروع البحثي من لجنة أخلاقيات البحث بجامعة فلندرز (رقم المشروع: 5074). لمزيد من التفاصيل بشأن الموافقة الأخلاقية للمشروع , يمكن الاتصال بالمسؤول التنفيذي للجنة على الهاتف (+61 8) 8201 3116 أو عن طريق الفاكس على (+61 8) 8201 2035 أو عن طريق إرسال إيميل إلى العنوان *human.researchethics@flinders.edu.au*

inspiring
achievement

- الجزء الأول : معلومات عامة

1. في أي عام تأسست الشركة ؟ **(ملاحظة : يفضل أن يكون التاريخ بالميلادي)**

2. هل ترى حجم الشركة بأنه:

O صغير

O متوسط

O كبير

3. كم العدد التقريبي للموظفين في الشركة ؟

4. تقريباً , كم عدد العملاء الذين تتعامل معهم الشركة ؟

5. هل شركتكم لديها بشكل واضح وحدة عمل أو فريق عمل لإدارة أمن الشبكة؟

O نعم

O لا **انتقل إلى السؤال رقم 8**

6. إذا نعم , ماهي مسؤوليات هذه الوحدة أو هذا الفريق ؟

7. إذا نعم , تقريباً كم عدد الموظفين الذين يعملون في هذه الوحدة أو هذا الفريق ؟

8. هل هناك موظفين مسؤوليتهم المحددة هي مكافحة رسائل البريد الإلكتروني الإختراقية؟

O نعم

O لا **انتقل إلى السؤال رقم 10**

9. إذا نعم , ماهي مهامهم لمكافحة رسائل البريد الإلكتروني الإختراقية ؟

- الجزء الثاني : طبيعة رسائل البريد الإلكتروني الإختراقية و أضرارها على أداء مزودي خدمة الإنترنت في السعودية

10. كل شخص يعرف رسائل البريد الإلكتروني بشكل مختلف, **بأسلوبك الخاص** , كيف تعرف رسائل البريد الإلكتروني الإختراقية؟

▪ تعريف رسائل البريد الإلكتروني الإختراقية:

يمكن تعريف الرسالة الإختراقية على أنها" ***البريد الإلكتروني الغير مرغوب به و الذي يشتمل على محتوى تجاري أو غير تجاري والذي يتم إرساله بشكل عشوائي أو بشكل مباشر أو غير مباشر إلى مجموعة كبيرة من المستلمين بدون أخذ موافقتهم على إرسال هذا الإيميل وكذلك لا تربطهم أي علاقة مع المرسل".***

تأخذ الرسائل الإختراقية العديد من الأشكال , فمن الأمثلة لهذه الرسائل : الإعلانات الترويجية من القطاعات التجارية المختلفة , الرسائل البريدية من المجموعات الإخبارية والسياسية والأحزاب الدينية والمنتديات , الإعلانات لمجموعة واسعة من المنتجات والخدمات والتي تتضمن المنتجات الصحية مثل الترويج لدواء معين أو الاستشارات الطبية مثل إتباع برنامج حمية معين , الرسائل الرياضية , الألعاب الإلكترونية على الإنترنت.

كما يمكن استخدام الرسائل الإختراقية للخداع والتصيد وسرقة المعلومات الشخصية مثل اسم المستخدم وكلمة المرور وكذلك معلومات الحساب البنكية مثل أرقام البطاقات الإئتمانية.

▪ أمثلة للكلمات والعبارات المستخدمة في الرسائل الإختراقية:

1) ***"اضغط واربح" , "لقد ربحت" , "لقد ربحت مليون ريال سعودي"*** : تعتبر هذه العبارات أمثلة للإعلانات التجارية والخدمات.

2) ***"فياقرا" ,"حمية"*** : تعتبر أمثلة للإعلانات والتي تروج لمنتجات صحية وطبية.

3) ***"بيانات حسابك البنكي تحتاج إلى تحديث" , "معلوماتك الشخصية غير مكتملة"*** : تعتبر أمثلة للعبارات التي تستخدم في الخداع وسرقة المعلومات الشخصية.

11. هل شركتكم حظرت أي رسائل بريد إلكتروني إختراقية في الآونة الأخيرة؟

○ نعم

○ لا

12. إذا نعم , كم عدد رسائل البريد الإلكتروني الإختراقية التي تم حظرها أسبوعياً؟

رجاءً , قدّر العدد :

ملاحظة : السؤال التالي سوف يسألك عن تقدير النسبة المئوية التقديرية لكل لغة رسالة بريد إلكترونية إختراقية تم حظرها. مجموع النسب المئوية لجميع لغات رسائل السبام يجب أن يكون **100%**.

على سبيل المثال, لو كانت لغات رسائل السبام التي قمتم بحظرها هي اللغة الانجليزية و العربية و التركية , إذن النسب المئوية التقريبية لكل لغة رسالة سبام تم حظرها يمكن أن تقدّر كما في الجدول التالي :

	لغة رسائل البريد الإلكتروني الإختراقية	**النسبة المئوية**
✓	الانجليزية	20 %
✓	العربية	50 %
✓	لغة أخرى , رجاءً اذكر : **التركية**	30 %
	لغة أخرى , رجاءً اذكر:	0 %
	المجموع	**100 %**

نلاحظ من الجدول السابق:

السبام الانجليزي (20%) + السبام العربي (50%) + السبام التركي (30%) = المجموع هو (**100%**)

13. ماهي لغة رسائل البريد الإلكتروني الإختراقية التي تم حظرها أسبوعياً؟ **يمكن أن تختار أكثر من خيار**

لغة رسائل البريد الإلكتروني الإختراقية		**النسبة المئوية**
	الانجليزية	**%**
	العربية	**%**
	لغة أخرى , رجاءً اذكر:	**%**
	لغة أخرى , رجاءً اذكر:	**%**
	لغة أخرى , رجاءً اذكر:	**%**
	لغات لا أستطيع تمييزها	**%**
المجموع		**% 100**

14. إذا كانت لغة الرسائل الإختراقية هي **اللغة العربية** , ماهي أنواع الرسائل الإختراقية التي تم حظرها أسبوعياً؟ **يمكن أن تختار أكثر من خيار**

ملاحظة قبل أن تجيب على هذا السؤال: الرجاء تقدير النسبة المئوية التقريبية لكل خيار تختاره , مجموع النسب المئوية يجب أن يكون **100%**. انظر المثال في **سؤال 15** لمزيد من التفاصيل عن تقدير النسبة المئوية التقديرية.

أنواع الرسائل الإختراقية العربية		**النسبة المئوية**
	الإعلانات التجارية	%
	رسائل من المجموعات الدينية والأطراف السياسية	%
	رسائل من المواقع الإباحية	%
	رسائل من المنتديات	%
	الإعلانات للمنتجات والخدمات	%
	الخداع والاحتيال أوسرقة المعلومات	%
	آخر:	%
المجموع		**% 100**

15. إذا كانت لغة الرسائل الإختراقية هي **اللغة العربية** , ما هو مصدر رسائل البريد الإلكتروني الإختراقيه التي تم حظرها أسبوعياً؟ **يمكن أن تختار أكثر من خيار**

ملاحظة قبل أن تجيب على هذا السؤال: الرجاء تقدير النسبة المئوية التقريبية لكل خيار تختاره , مجموع النسب المئوية يجب أن يكون **100%**. انظر المثال في **سؤال 15** لمزيد من التفاصيل عن تقدير النسبة المئوية التقديرية.

مصدر رسائل البريد الإلكتروني الإختراقيه العربية		**النسبة المئوية**
	المملكة العربية السعودية	%
	الدول العربية الأخرى	%
	الدول الغير عربية	%
	غير معروف	%
المجموع		**% 100**

16. الرجاء إدراج أي كلمات دليلية (مفتاحية) أو عبارات قمت بملاحظتها في رسائل السبام العربية.

17. إذا كانت لغة الرسائل الإختراقية هي **اللغة الإنجليزية** , ماهي أنواع الرسائل الإختراقية التي تم حظرها أسبوعياً؟ **يمكن أن تختار أكثر من خيار**

ملاحظة قبل أن تجيب على هذا السؤال: الرجاء تقدير النسبة المئوية التقريبية لكل خيار تختاره , مجموع النسب المئوية يجب أن يكون **100%**. انظر المثال في **سؤال 15** لمزيد من التفاصيل عن تقدير النسبة المئوية التقديرية.

أنواع الرسائل الإختراقية الانجليزية		**النسبة المئوية**
	الإعلانات التجارية	%
	رسائل من المجموعات الدينية والأطراف السياسية	%
	رسائل من المواقع الإباحية	%
	رسائل من المنتديات	%
	الإعلانات للمنتجات والخدمات	%
	الخداع والاحتيال أوسرقة المعلومات	%
	آخر:	%
المجموع		**% 100**

18. إذا كانت لغة الرسائل الإختراقية هي **اللغة الانجليزية** , ما هو مصدر رسائل البريد الإلكتروني الإختراقيه التي تم حظرها أسبوعياً؟ **يمكن أن تختار أكثر من خيار**

ملاحظة قبل أن تجيب على هذا السؤال: الرجاء تقدير النسبة المئوية التقريبية لكل خيار تختاره , مجموع النسب المئوية يجب أن يكون **100%**. انظر المثال في **سؤال 15** لمزيد من التفاصيل عن تقدير النسبة المئوية التقديرية.

مصدر رسائل البريد الإلكتروني الإختراقيه الانجليزية		**النسبة المئوية**
	المملكة العربية السعودية	%
	الدول العربية الأخرى	%
	الدول الغير عربية	%
	غير معروف	%
المجموع		**% 100**

19. الرجاء إدراج أي كلمات دليلية (مفتاحية) أو عبارات قمت بملاحظتها في رسائل السبام الانجليزية.

20. ما هي تأثيرات رسائل البريد الإلكتروني الإختراقيه على مزودي خدمة الإنترنت ؟ **يمكن أن تختار أكثر من خيار**

□ ضياع الوقت وتقليل الإنتاجية

□ إنفاق الكثير من المال في تطبيق وتجديد الفلاتر أو البرامج المستخدمة لمكافحة الرسائل الإختراقية, وكذلك في شراء عرض نطاق ترددي أو سعة إضافية لنظام البريد الإلكتروني

□ فقدان العملاء بسبب تلقيهم كمية كبيرة من الرسائل الإختراقية

□ استهلاك عرض النطاق الترددي بسبب الكمية الكبيرة من الرسائل الإختراقية

□ تأثيرات أخرى , اذكر

21. كم من الوقت يُقضى في إصلاح المشاكل المتعلقة برسائل البريد الإلكتروني الإختراقية أسبوعياً؟ **(ملاحظة : الوقت يجب أن يقدّر بالساعات)**

- الجزء الثالث : الفلاتر المستخدمة من قبل مزودي خدمة الإنترنت لحظر رسائل البريد الإلكتروني الإختراقية و فعاليتها في اكتشاف رسائل السبام العربية والإنجليزية

- **تقنيات مكافحة رسائل البريد الإلكتروني الإختراقية:**

هناك نوعان من التقنيات الرئيسية المستخدمة لتصنيف رسائل البريد الإلكتروني إلى رسائل إختراقية (سبام) أو رسائل غير إختراقية (هامة). هذه التقنيات هي التقنيات المعتمدة على المحتوى والتقنيات المعتمدة على الأصل أو المصدر.

1) **الفلاتر (التقنيات) المعتمدة على المحتوى :** هذه الفلاتر تقوم باكتشاف الرسائل الإختراقية من خلال فحص أو اختبار محتوى رسائل البريد الإلكتروني , بغض النظر عن أصل أو مصدر الرسالة. يوجد هناك العديد من هذه التقنيات مثل تقنيات الكلمات الدليلية أو المفتاحية , تقنيات تعلم الآلة وتقنيات بصمة الإصبع .

2) **الفلاتر (التقنيات) المعتمدة على الأصل أو المصدر** : يتم تصنيف الرسائل الإختراقية عن طريق معلومات الشبكة مثل مصدر عناوين الآي بي (IP) وعناوين البريد الإلكتروني. من الأمثلة لهذه التقنيات القوائم السوداء و القوائم البيضاء وأنظمة استجابة التحدي.

22. ما هي التقنيات المستخدمة في فلاتركم لاكتشاف رسائل البريد الإلكتروني الإختراقية؟ **اختر كل ما ينطبق**

□ التقنيات المعتمدة على المحتوى

□ التقنيات المعتمدة على الأصل أو المصدر

□ لا نقوم بترشيح الرسائل الإختراقية **انتقل إلى السؤال رقم 30**

23. إذا كانت الفلاتر الخاصة بكم لمكافحة الرسائل الإختراقية تعتمد على **المحتوى** , الرجاء وضع **دائرة** على الفلاتر الموجودة في الجدول التالي والتي تقومون باستخدامها في مكافحة الرسائل الإختراقية؟ **اختر كل ما ينطبق**

فلاتر مكافحة رسائل البريد الإلكتروني الإختراقية		
MailWasher	eMailTrackerPro	SpamBayes
SpamFighter	SpamButcher	POPFile
Cactus Spam Filter	SpamSource	Spam Monitor
CleanMail	SpamBully	Spam Buster
AntiSpam Sniper	SpamAssassin	Antispam Scanner
SpamBlocker	Spam Eliminator	Spam Nullifier
iHateSpam	SpamEater	Spam Eraser
Anti-SPAM Guard	SpamWasher	Spam Sleuth
KillSpam	Brightmail Anti-Spam	KasperSky Anti-Spam
يرجى إدراج أي فلاتر أخرى قمتم باستخدامها لحظر رسائل البريد الإلكتروني الإختراقية :		

24. إذا كنتم تستخدمون **الفلاتر المعتمدة على المحتوى** , الرجاء تقدير فعاليتها في اكتشاف رسائل البريد الإلكتروني الإختراقية الانجليزية والعربية ؟

ملاحظة: الرجاء اختيار النسب المئوية المناسبة لفعالية الفلاتر المعتمدة على المحتوى في اكتشاف الرسائل الإختراقية العربية والإنجليزية من الخيارات الموجودة أسفل أو تقدير نسب مئوية تقريبية أخرى بناءاً على رأيك

الفلاتر المعتمدة على المحتوى / النسبة المئوية	**% 0**	**25 %**	**50 %**	**75 %**	**100 %**
فعالية الفلاتر المعتمدة على المحتوى في اكتشاف الرسائل الإختراقية العربية	○	○	○	○	○
فعالية الفلاتر المعتمدة على المحتوى في اكتشاف الرسائل الإختراقية الإنجليزية	○	○	○	○	○

نسب مئوية أخرى , رجاءً قدر

25. إذا كانت الفلاتر الخاصة بكم لمكافحة الرسائل الإختراقية تعتمد على **الأصل أو المصدر** , ماهي أنواع الفلاتر التي قمتم باستخدامها لمنع أو حظر الرسائل الإختراقية؟ **اختر كل ما ينطبق**

□ القوائم السوداء
□ القوائم البيضاء
□ أنظمة استجابة التحدي

26. إذا كنتم تستخدمون **الفلاتر المعتمدة على الأصل أو المصدر** , الرجاء تقدير فعاليتها في اكتشاف رسائل البريد الإلكتروني الإختراقية الانجليزية والعربية ؟

ملاحظة: الرجاء اختيار النسب المئوية المناسبة لفعالية الفلاتر المعتمدة على المصدر في اكتشاف الرسائل الإختراقية العربية والإنجليزية من الخيارات الموجودة أسفل أو تقدير نسب مئوية تقريبية أخرى بناءاً على رأيك

الفلاتر المعتمدة على المصدر / النسبة المئوية	0 %	25 %	50 %	75 %	100 %
فعالية الفلاتر المعتمدة على المصدر في اكتشاف الرسائل الإختراقية العربية	○	○	○	○	○
فعالية الفلاتر المعتمدة على المصدر في اكتشاف الرسائل الإختراقية الإنجليزية	○	○	○	○	○

نسب مئوية أخرى , رجاءً قدر

27. هل تقومون بتحديث فلاتر مكافحة الرسائل الإختراقية التي تستخدموها بشكل منتظم؟

O نعم
O لا

- الجزء الرابع : الجهود المبذولة لمكافحة رسائل البريد الإلكتروني الإختراقية في السعودية, وكذلك جهود مزودي خدمة الإنترنت لتوعية العملاء والموظفين عن الرسائل الإختراقية (السبام)

28. ماهي الجهود التي تبذلها الحكومة لمكافحة رسائل البريد الإلكتروني الإختراقية في المملكة العربية السعودية والتي أنتم على علم بها ؟

29. هل يوجد هناك جهود توعية مُقدمة من قبل مزودي خدمة الإنترنت للعملاء حول رسائل البريد الإلكتروني الإختراقية والطرق الملائمة لمكافحتها ؟

O نعم , رجاءً وضح

O لا

30. هـل هنـاك أي ورش عمـل أو تـدريب مسـتمر تعقـد لمـوظفي الشـركة حـول رسـائل البريـد الإلكترونـي الإختراقية وكيفية السيطرة عليها؟

O نعم

O لا **انتقل إلى السؤال رقم 12**

31. إذا نعم , متى تُعقد هذه الورش , الدورات , المؤتمرات ؟

O كل 1 -3 أشهر

O كل 4 -6 أشهر

O كل 7 -9 أشهر

O كل 10 -12 أشهر

O آخر

32. في اعتقادكم ماهي الطرق الملائمة لمكافحة رسائل البريد الإلكتروني الإختراقية في السعودية؟

O تقنية مثل برامج أو مكونات مادية , رجاءً وضّح

O قانونية مثل قوانين جديدة , رجاءً وضّح

O آخر

33. رجاءً لا تتردد في إضافة أي شيء تعتقد أنه قد يكون ذا قيمة لهذا البحث :

- هل تريدون نسخة من نتائج هذا الاستبيان؟

O نعم
O لا

- إذا نعم , الرجاء كتابة البريد الإلكتروني :

- **اختياري :** نرغب في جمع مجموعة كبيرة من رسائل البريد الإلكتروني الإختراقية **(السبام)** سواء كانت عربية أو انجليزية وذلك لتحليلها لتحقيق بعض أهداف البحث. هذه المجموعة ممكن أن تستخدم في تحليل و فهم طرق وخدع من يقومون بإرسال هذه الرسائل (السبامرز) والذي بدوره سيساعد في تطوير الفلاتر الحالية أو إيجاد أخرى جديدة. هذه المجموعة ممكن أيضاً أن تساعد في اختبار هذه الرسائل على التقنيات الحالية مما سيؤدي إلى اكتشاف فعالية هذه التقنيات في حظر الرسائل الإختراقية العربية والانجليزية. إذا كنت قادراً على المساعدة الرجاء إرسال هذه الرسائل على الإيميل التالي: hasan.sh.ka@gmail.com

إذا كنت بحاجة إلى مزيد من التفاصيل حول هذا البحث , الرجاء الاتصال بنا على عنوان البريد الإلكتروني التالي : alka0022@flinders.edu.au

شكراً لكم لإكمال الاستبيان

Appendix G: A Letter of Introduction + An ISP Questionnaire (English Version)

Flinders
UNIVERSITY

**School of Computer Science,
Engineering and Mathematics**

Room (358), Information Science &
Technology Building

GPO Box 2100
Adelaide SA 5001

Tel: (+61 8) 8201 3113
Fax: (+61 8) 8201 2904
Email: Robert.goodwin@flinders.edu.au

http://csem.flinders.edu.au/

CRICOS Provider No. 00114A

LETTER OF INTRODUCTION

Dear Sir/Madam,

This letter is to introduce Mr Hasan Shojaa Alkahtani who is a PhD student in the School of Computer Science, Engineering and Mathematics at Flinders University.

He is undertaking research leading to the production of a thesis or other publications on the subject of "Exploration of Email SPAM, with a focus on its effects and mitigation in Saudi Arabia".

He would be most grateful if you would volunteer to assist in this project, by completing this questionnaire which covers certain aspects of this topic. This questionnaire will investigate email SPAM and its effects on the Internet Service Providers (ISPs) in Saudi Arabia. It will also investigate the efforts to combat email SPAM in Saudi Arabia as well as the effectiveness of current Anti-SPAM filters in detecting Arabic and English email SPAM. No more than 30 minutes is required to complete the questionnaire.

A summary of the results will be sent by email to interested respondents.

Be assured that any information provided will be treated in the strictest confidence and none of the participants will be individually identifiable in the resulting thesis, report or other publications. You are, of course, entirely free to discontinue your participation at any time or to decline to answer particular questions.

Any enquiries you may have concerning this project should be directed to me at the address given above or by telephone on (+61 8) 8201 3113, by fax on (+61 8) 8201 2904 or by email to (Robert.goodwin@flinders.edu.au).

Thank you for your attention and assistance.

Yours sincerely

Dr. Robert Goodwin
Senior Lecturer
School of Computer Science, Engineering and Mathematics

This research project has been approved by the Flinders University Social and Behavioural Research Ethics Committee (Project Number: 5074). For more information regarding ethical approval of the project the Executive Officer of the Committee can be contacted by telephone on (+61 8) 8201 3116, by fax on (+61 8) 8201 2035 or by email human.researchethics@flinders.edu.au.

ABN 65 524 596 200 CRICOS Provider No. 00114A

inspiring achievement

• Part 1: Demographic Information

1. What year was the company established?

2. Do you see the size of the company as being:
   O Small
   O Medium
   O Large

3. What is the approximate number of employees in the company?

4. Approximately how many customers does the company deal with?

5. Does your company have explicitly a business unit or team for managing network security?
   O Yes
   O No **Go to question 8**

6. If yes, what are the responsibilities of this unit or this team?

7. If yes, approximately how many employees are involved in this unit or this team?

8. Are there employees with specific responsibility for combating email SPAM?
   O Yes
   O No **Go to question 10**

9. If yes, what are their tasks to combat email SPAM?

• Part 2: The nature of Email SPAM, and its effects on the performance of Internet Service Providers (ISPs) in Saudi Arabia

10. Everyone defines SPAM differently**, in your own words**, how would you define email SPAM?

- **Email SPAM definition:**

Email SPAM can be defined as ***"an unsolicited, unwanted, commercial or non-commercial email that is sent indiscriminately, directly or indirectly, to a large number of recipients without their permission and there is no relationship between the recipients and a sender"***.

Email SPAM has many forms such as promotional advertisements from businesses, religious groups, political parties, pornographic websites and forums, and advertisements for a wide variety of products and services including medical, sports and online gaming.

SPAM may also be used for phishing to obtain credit card numbers, usernames, passwords and other personal information.

- **Examples of words and phrases used in SPAM include:**

1) ***"CLICK and WIN", "YOU HAVE WON"*** and ***"YOU WON 1 MILLION DOLLARS"*** are examples for advertisements of businesses and services.
2) ***"VIAGRA"*** and ***"DIET"*** are examples for advertisements of medical and health products.
3) ***"YOUR ACCOUNT NEEDS TO BE UPDATED"*** and ***"INCOMPLETE PERSONAL INFORMATION"*** are examples for phishing.

11. Has your company blocked any email SPAM recently?
    ○ Yes
    ○ No

12. If yes, how many SPAM emails are blocked on average weekly?
    Please, estimate □

Note: the following question will ask you to estimate the relative percentage for each language of email SPAM you have blocked. The percentages should add up to **100 %**.

For example, if the languages of email SPAM that you have blocked were English, Arabic and Turkish, the relative percentages for each language might be estimated as follows:

Language of email SPAM		Percentage
✓	English	20 %
✓	Arabic	50 %
✓	Other language, please state : **Turkish**	30 %
	Other language, please state :	0 %
Total		**100 %**

So, English SPAM (**20%**) + Arabic SPAM (**50%**) + Turkish SPAM (**30%**) = **100%**

13. What is the language of SPAM email you block on average weekly? **You can choose more than one option**

Language of email SPAM		Percentage
	English	%
	Arabic	%
	Other language, please state :	%
	Other language, please state :	%
	Other language, please state :	%
	Languages I do not recognise	%
Total		**100 %**

14. If the language of SPAM was Arabic, what types of SPAM have you blocked on average weekly? **You can choose more than one option**

Note before you answer this question: please estimate the relative percentage for each option you select, the percentages should total **100 %**. See the example in **question 15** for more explanation about estimating the relative percentage.

Type of Arabic email SPAM		Percentage
	Businesses advertisements	%
	Emails from religious groups and political parties	%
	Emails from pornographic websites	%
	Emails from forums	%
	Products and services advertisements	%
	Phishing and fraud	%
	Other:	%
Total		**100 %**

15. If the language of SPAM was Arabic, what was the source of email SPAM have you blocked on average weekly? **You can choose more than one option**

Note before you answer this question: please estimate the relative percentage for each option you select, the percentages should total **100 %**. See the example in **question 15** for more explanation about estimating the relative percentage.

Source of Arabic SPAM		Percentage
	Saudi Arabia	%
	Other Arabic countries	%
	Non-Arabic countries	%
	Unknown	%
Total		**100 %**

16. Please list any keywords or phrases of Arabic SPAM that you have observed.

17. If the language of SPAM was English, what types of SPAM you have blocked on average weekly? **You can choose more than one option**

Note before you answer this question: please estimate the relative percentage for each option you select, the percentages should total **100 %**. See the example in **question 15** for more explanation about estimating the relative percentage.

Type of English email SPAM		Percentage
	Businesses advertisements	%
	Emails from religious groups and political parties	%
	Emails from pornographic websites	%
	Emails from forums	%
	Products and services advertisements	%
	Phishing and fraud	%
	Other:	%
Total		**100 %**

18. If the language of SPAM was English, what was the source of email SPAM have you blocked on average weekly? **You can choose more than one option**

Note before you answer this question: please estimate the relative percentage for each option you select, the percentages should total **100 %**. See the example in **question 15** for more explanation about estimating the relative percentage.

Source of English SPAM		Percentage
	Saudi Arabia	%
	Other Arabic countries	%
	Non-Arabic countries	%
	Unknown	%
Total		**100 %**

19. Please list any keywords or phrases of English SPAM that you have observed.

20. What are the effects of email SPAM on ISPs? **You can choose more than one option**

□ Losing time ad reducing productivity

□ Spending a lot of money to implement and update filters used to combat SPAM, and to buy extra bandwidth and capacity for email system

□ Losing customers due to receiving a large volume of email SPAM

□ Consumption of the bandwidth by excessive email SPAM

□ Other impacts: please list them,

21. How much time do you spend in fixing related SPAM problems on average weekly?
(Note: the time should be estimated in hours)

• Part 3: Anti-SPAM filters used by the ISPs to block email SPAM, and their effectiveness in detecting Arabic and English SPAM:

▪ Techniques of combating email SPAM

There are two main techniques used to classify email as SPAM (junk email) or non-SPAM (Legitimate). These techniques are content based filtering and origin based filtering.

1. **Content based filters:** detect SPAM by examining the content of email messages, irrespective of the origin. There exist several families of content based filtering techniques, including: (a) keywords; (b) machine learning, and; (c) finger printing.
2. **Origin based filters:** SPAM classified by network information such as the source IP and email addresses. Blacklists, Whitelists, and Challenge Response Systems are examples of these techniques.

22. What are techniques used in your filters to detect email SPAM? **Tick all that apply**
    □ Content based filters
    □ Origin based filters
    □ We do not filter SPAM **Go to question 30**

23. If your SPAM filters depend on **content**, please circle any filters from the following lists you have used to block email SPAM? **Choose all that apply**

Anti-SPAM filters		
MailWasher	eMailTrackerPro	SpamBayes
SpamFighter	SpamButcher	POPFile
Cactus Spam Filter	SpamSource	Spam Monitor
CleanMail	SpamBully	Spam Buster
AntiSpam Sniper	SpamAssassin	Antispam Scanner
SpamBlocker	Spam Eliminator	Spam Nullifier
iHateSpam	SpamEater	Spam Eraser
Anti-SPAM Guard	SpamWasher	Spam Sleuth
KillSpam	Brightmail Anti-Spam	KasperSky Anti-Spam
Please list any other filters that you have used to block email SPAM:		

24. If you employ **content based filters,** please rate their effectiveness in detecting English and Arabic email SPAM**?**

Note: please choose the appropriate percentage for the effectiveness of content based filters in detecting Arabic and English email SPAM from the following options or estimate other relative percentages based on your opinion.

Content based filters\ Percentage	0 %	25 %	50 %	75 %	100 %
The effectiveness of content based filters in detecting Arabic email SPAM	○	○	○	○	○
The effectiveness of content based filters in detecting English email SPAM	○	○	○	○	○

Other percentages, please estimate

25. If your SPAM filters depend on **origin**, what types of SPAM filters have you used to block SPAM? **Tick all that apply**
□ Blacklists
□ Whitelists
□ Challenge Response System

26. If you employ **origin-based filters**, please rate their effectiveness in detecting English and Arabic email SPAM**?**

Note: please choose the appropriate percentage for the effectiveness of origin-based filters in detecting Arabic and English email SPAM from the following options or estimate other relative percentages based on your opinion.

Origin based filters\ Percentage	0 %	25 %	50 %	75 %	100 %
The effectiveness of origin based filters in detecting Arabic email SPAM	○	○	○	○	○
The effectiveness of origin based filters in detecting English email SPAM	○	○	○	○	○

Other percentages, please estimate

27. Do you update Anti-SPAM filters that you use regularly?
    O Yes
    O No

- **Part 4: The efforts to combat SPAM in Saudi Arabia, and the efforts of ISPs to inform customers and employees about SPAM**

28. What efforts of the government to combat SPAM are you aware of?

29. Is there awareness provided by ISPs for customers about SPAM and appropriate methods to combat it?
    O Yes, please explain

    O No

30. Are there any workshops or ongoing training conducted for employees of company about SPAM emails and their control?
    O Yes
    O No **Go to question 12**

31. If yes, when these workshops, sessions or conferences are conducted?
    O Every 1-3 months
    O Every 4-6 months
    O Every 7-9 months
    O Every 10-12 months
    O Other

32. In your opinion, what are the appropriate ways to combat SPAM in Saudi Arabia?
    O Technical such as software , hardware , please explain

O Legal such as new laws , please explain

O Other

33. Please feel free to add anything that you think may be of value to this research:

- Do you want a summary of the results of this survey?
  - O Yes
  - O No

- If yes, please provide your email address :

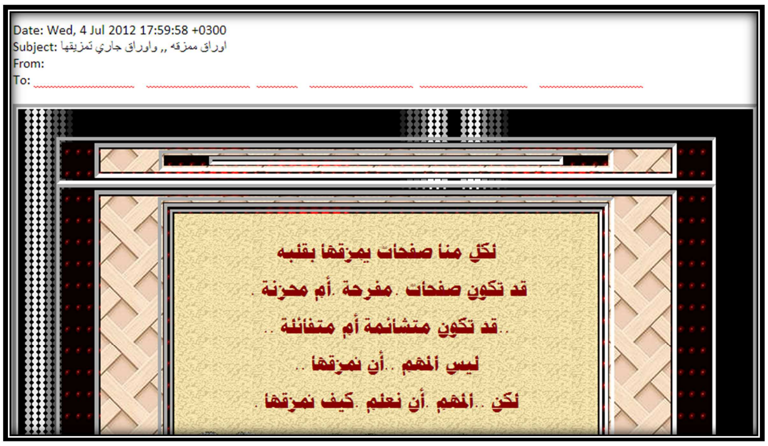
- **Optional:** We wish to collect large corpora of Arabic and English email SPAM to analyse them to achieve some research aims. These corpora could be used to analyse and understand methods and tricks used by spammers, which could help developers to improve the existing Anti-SPAM filters or produce new ones. These corpora could also help in testing SPAM emails on the current email SPAM detection methods which may lead to exploring the effectiveness of these methods in detecting Arabic and English email SPAM. If you are able to help, please send these messages to the following email: hasan.sh.ka@gmail.com

  If you need more explanation about this research, please contact us on the following email address: alka0022@flinders.edu.au

Thank you for completing the survey

Appendix H: Examples of types of Arabic, English and Mixed email spam (contains Arabic and English texts) that received in Saudi Arabia

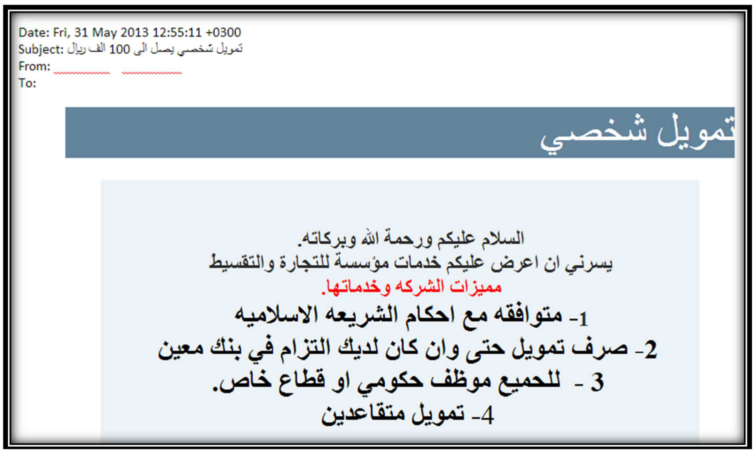
- **Arabic email related to forums**



Date: Wed, 4 Jul 2012 17:59:58 +0300
Subject: اوراق ممزقه ,, واوراق جاري تمزيقها
From:
To:

لكل منا صفحات يمزقها بقلبه
قد تكون صفحات .. مفرحة .. أم محزنة .
.. قد تكون متشائمة أم متفائلة ..
ليس المهم .. أن نمزقها ..
لكن .. المهم .. أن نعلم .. كيف نمزقها .

- **Arabic business advertisement email**

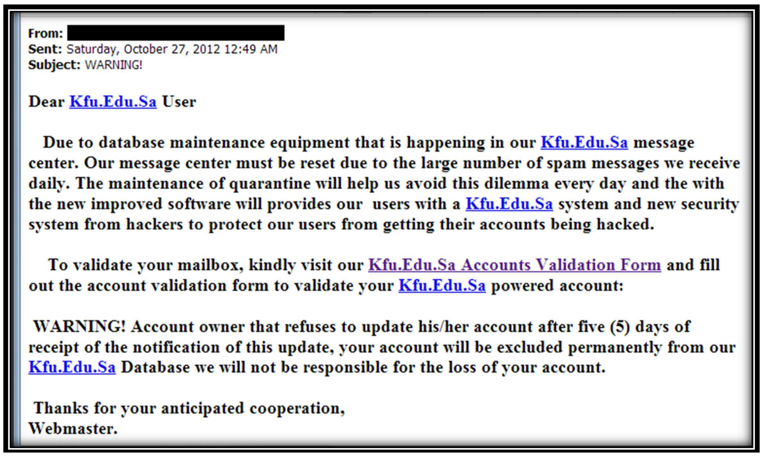


Date: Fri, 31 May 2013 12:55:11 +0300
Subject: تمويل شخصي يصل الى 100 الف ريال
From:
To:

تمويل شخصي

السلام عليكم ورحمة الله وبركاته.
يسرني ان اعرض عليكم خدمات مؤسسة للتجارة والتقسيط
مميزات الشركه وخدماتها.
1- متوافقه مع احكام الشريعه الاسلاميه
2- صرف تمويل حتى وان كان لديك التزام في بنك معين
3 - للحميع موظف حكومي او قطاع خاص.
4- تمويل متقاعدين

- **English phishing email**



From: [redacted]
Sent: Saturday, October 27, 2012 12:49 AM
Subject: WARNING!

Dear Kfu.Edu.Sa User

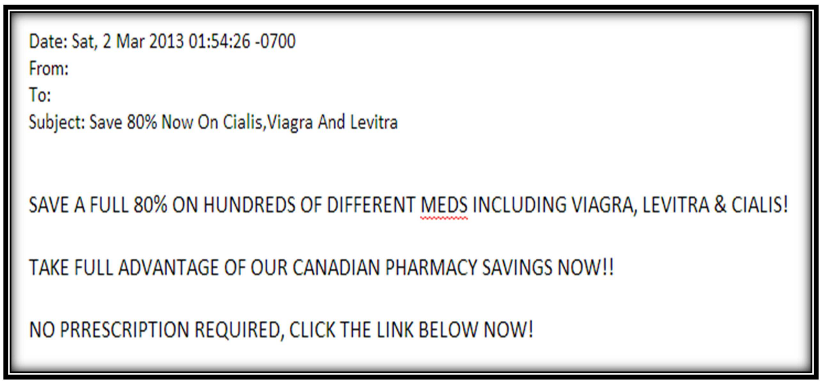
Due to database maintenance equipment that is happening in our Kfu.Edu.Sa message center. Our message center must be reset due to the large number of spam messages we receive daily. The maintenance of quarantine will help us avoid this dilemma every day and the with the new improved software will provides our users with a Kfu.Edu.Sa system and new security system from hackers to protect our users from getting their accounts being hacked.

To validate your mailbox, kindly visit our Kfu.Edu.Sa Accounts Validation Form and fill out the account validation form to validate your Kfu.Edu.Sa powered account:

WARNING! Account owner that refuses to update his/her account after five (5) days of receipt of the notification of this update, your account will be excluded permanently from our Kfu.Edu.Sa Database we will not be responsible for the loss of your account.

Thanks for your anticipated cooperation,
Webmaster.

- **English business advertisement email**



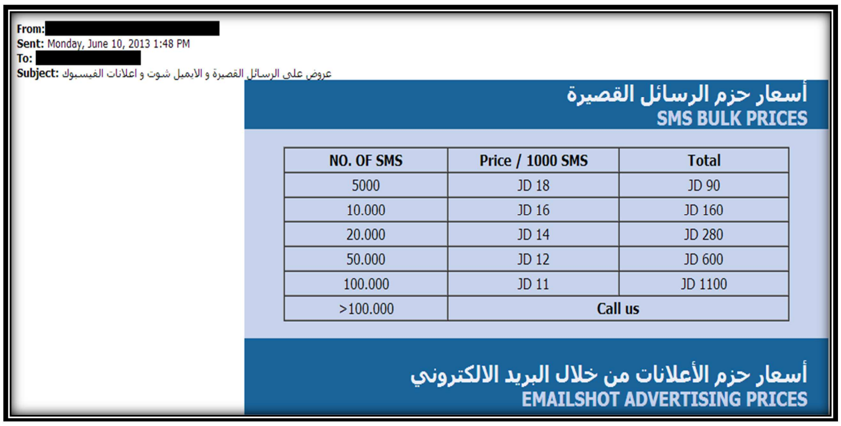
Date: Sat, 2 Mar 2013 01:54:26 -0700
From:
To:
Subject: Save 80% Now On Cialis,Viagra And Levitra

SAVE A FULL 80% ON HUNDREDS OF DIFFERENT MEDS INCLUDING VIAGRA, LEVITRA & CIALIS!

TAKE FULL ADVANTAGE OF OUR CANADIAN PHARMACY SAVINGS NOW!!

NO PRRESCRIPTION REQUIRED, CLICK THE LINK BELOW NOW!

- **Business advertisement email SPAM written in Mixed text (including Arabic and English)**



From: [redacted]
Sent: Monday, June 10, 2013 1:48 PM
To: [redacted]
Subject: عروض على الرسائل القصيرة و الايميل شوت و اعلانات الفيسبوك

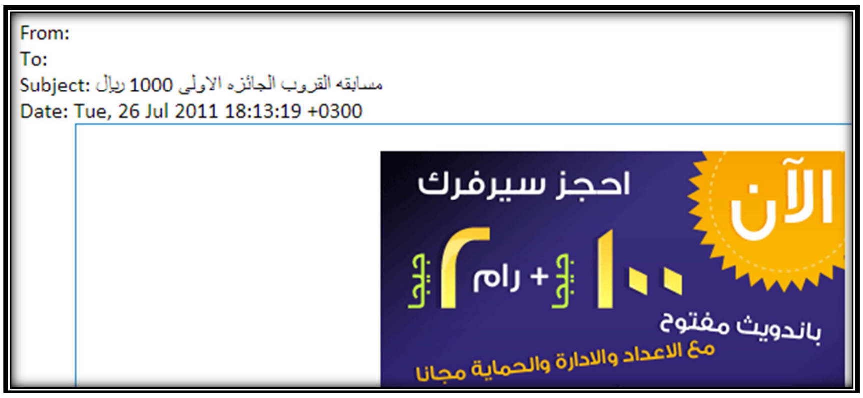
أسعار حزم الرسائل القصيرة
SMS BULK PRICES

NO. OF SMS	Price / 1000 SMS	Total
5000	JD 18	JD 90
10.000	JD 16	JD 160
20.000	JD 14	JD 280
50.000	JD 12	JD 600
100.000	JD 11	JD 1100
>100.000	Call us	

أسعار حزم الأعلانات من خلال البريد الالكتروني
EMAILSHOT ADVERTISING PRICES

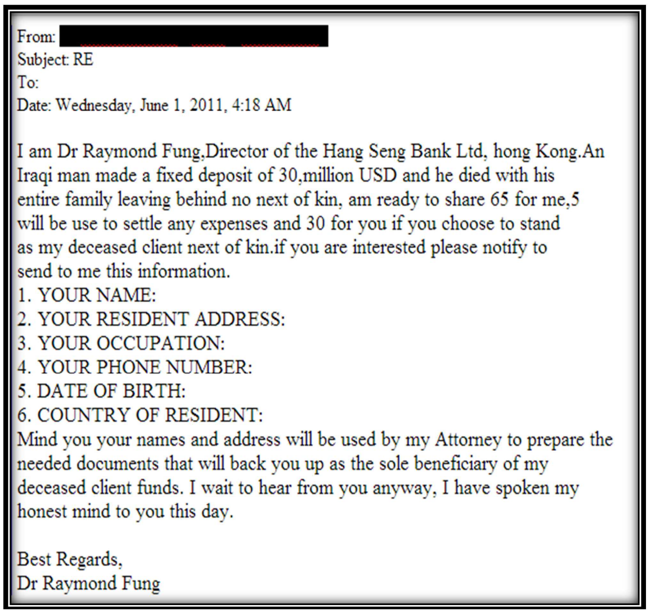
Appendix I: Examples of tricks used by spammers in the headers and bodies of Arabic, English and Mixed email spam

- **Arabic SPAM with a false statement in the subject line of email (the subject of email does not indicate its content)**



From:
To:
Subject: مسابقه القروب الجائزه الاولى 1000 ريال
Date: Tue, 26 Jul 2011 18:13:19 +0300

- **English SPAM with an attractive word “Re” in the subject line of email to obfuscate the recipients about its content**

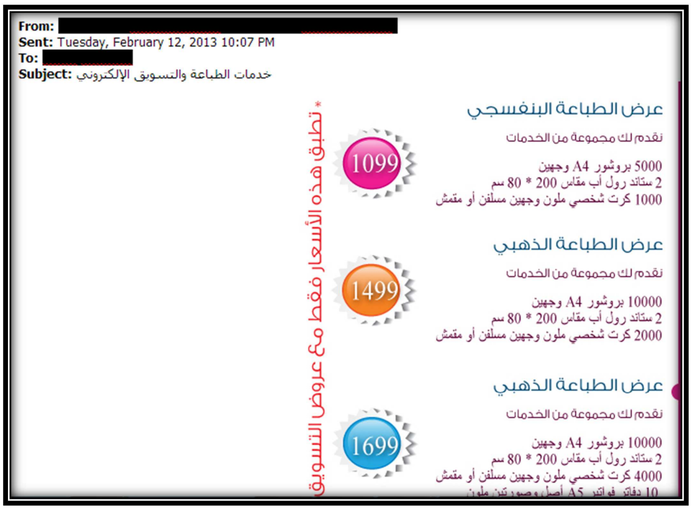


From: [redacted]
Subject: RE
To:
Date: Wednesday, June 1, 2011, 4:18 AM

I am Dr Raymond Fung,Director of the Hang Seng Bank Ltd, hong Kong.An Iraqi man made a fixed deposit of 30,million USD and he died with his entire family leaving behind no next of kin, am ready to share 65 for me,5 will be use to settle any expenses and 30 for you if you choose to stand as my deceased client next of kin.if you are interested please notify to send to me this information.
1. YOUR NAME:
2. YOUR RESIDENT ADDRESS:
3. YOUR OCCUPATION:
4. YOUR PHONE NUMBER:
5. DATE OF BIRTH:
6. COUNTRY OF RESIDENT:
Mind you your names and address will be used by my Attorney to prepare the needed documents that will back you up as the sole beneficiary of my deceased client funds. I wait to hear from you anyway, I have spoken my honest mind to you this day.

Best Regards,
Dr Raymond Fung

- **Arabic email SPAM appeared as a text embedded in an image (image SPAM) to bypass text based Anti-SPAM filters**



From: [redacted]
Sent: Tuesday, February 12, 2013 10:07 PM
To: [redacted]
Subject: خدمات الطباعة والتسويق الإلكتروني

* تطبق هذه الأسعار فقط مع عروض التسويق

عرض الطباعة البنفسجي

نقدم لك مجموعة من الخدمات

5000 بروشور A4 وجهين
2 ستاند رول أب مقاس 200 * 80 سم
1000 كرت شخصي ملون وجهين مسلفن أو مقمش

1099

عرض الطباعة الذهبي

نقدم لك مجموعة من الخدمات

10000 بروشور A4 وجهين
2 ستاند رول أب مقاس 200 * 80 سم
2000 كرت شخصي ملون وجهين مسلفن أو مقمش

1499

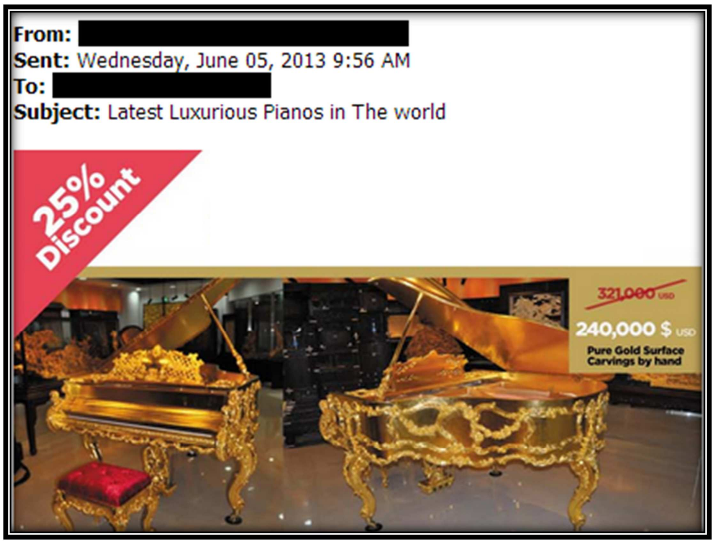
عرض الطباعة الذهبي

نقدم لك مجموعة من الخدمات

10000 بروشور A4 وجهين
2 ستاند رول أب مقاس 200 * 80 سم
4000 كرت شخصي ملون وجهين مسلفن أو مقمش
10 دفاتر فواتير A5 أصل وصورتين ملون

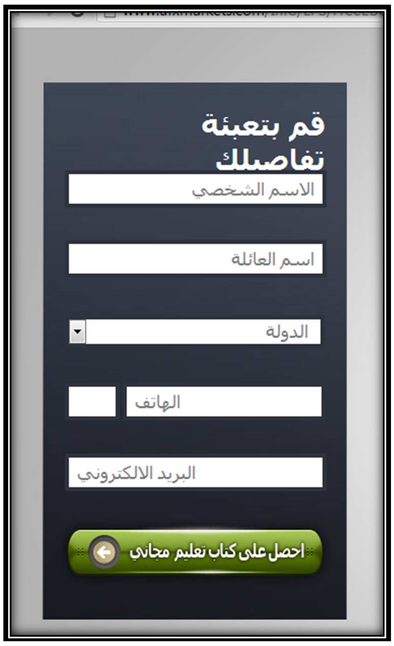
1699

- **English email SPAM appeared as a text embedded in an image (image SPAM)**



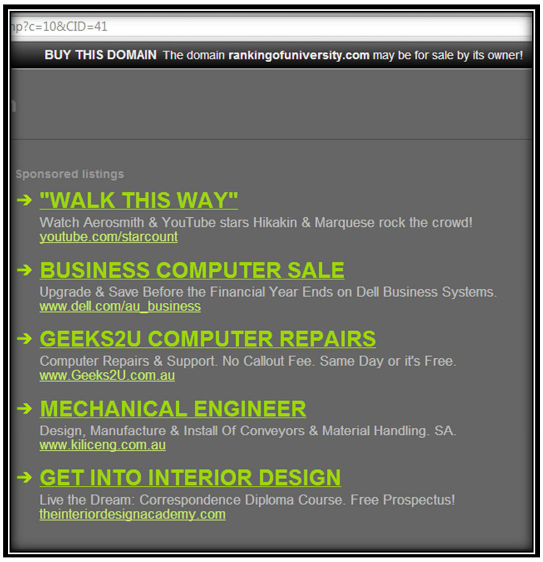
From: [redacted]
Sent: Wednesday, June 05, 2013 9:56 AM
To: [redacted]
Subject: Latest Luxurious Pianos in The world

- **An Arabic business website opened by clicking on a spoofed unsubscribe link involved in Arabic SPAM**



قم بتعبئة
تفاصيلك

- **An English business website opened by clicking onto a spoofed unsubscribe link included in English SPAM**



p?c=10&CID=41

BUY THIS DOMAIN The domain **rankingofuniversity.com** may be for sale by its owner!

Sponsored listings

→ "WALK THIS WAY"
Watch Aerosmith & YouTube stars Hikakin & Marquese rock the crowd!
youtube.com/starcount

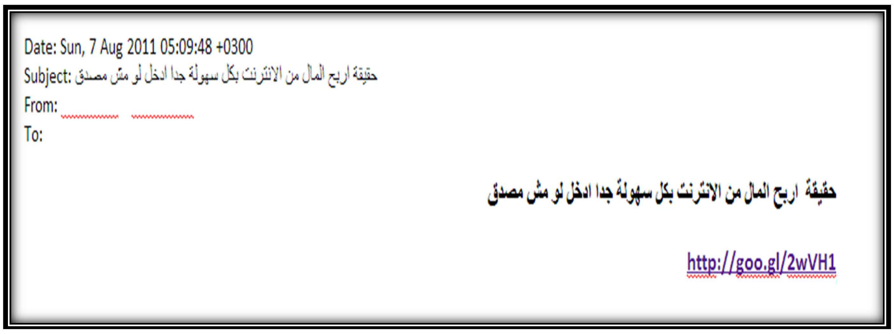
→ BUSINESS COMPUTER SALE
Upgrade & Save Before the Financial Year Ends on Dell Business Systems.
www.dell.com/au_business

→ GEEKS2U COMPUTER REPAIRS
Computer Repairs & Support. No Callout Fee. Same Day or it's Free.
www.Geeks2U.com.au

→ MECHANICAL ENGINEER
Design, Manufacture & Install Of Conveyors & Material Handling. SA.
www.kiliceng.com.au

→ GET INTO INTERIOR DESIGN
Live the Dream: Correspondence Diploma Course. Free Prospectus!
theinteriordesignacademy.com

- **Arabic email SPAM involved a malicious link**

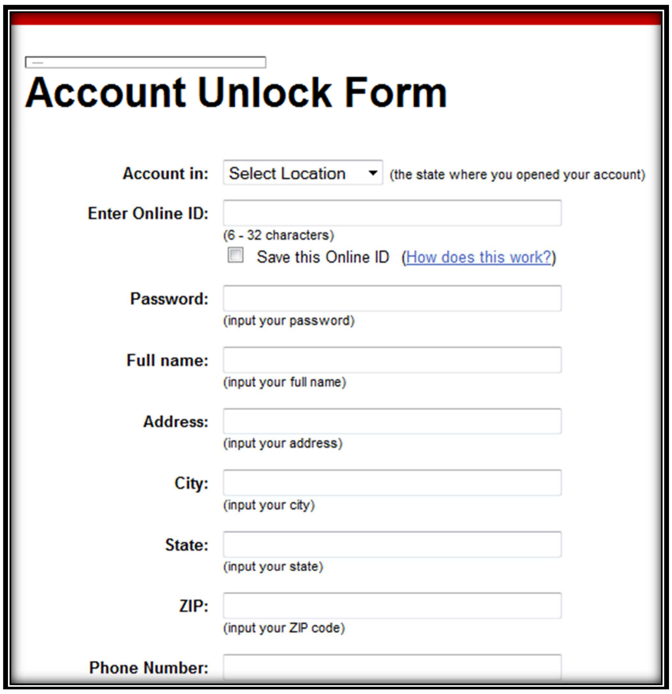


Date: Sun, 7 Aug 2011 05:09:48 +0300
Subject: حقيقة اربح المال من الانترنت بكل سهولة جدا ادخل لو مش مصدق
From:
To:

حقيقة اربح المال من الانترنت بكل سهولة جدا ادخل لو مش مصدق

http://goo.gl/2wVH1

- **A counterfeit webpage was opened by clicking onto a malicious link included in English SPAM to steal confidential information of users**



Account Unlock Form

Account in: Select Location (the state where you opened your account)

Enter Online ID: (6 - 32 characters)
□ Save this Online ID (How does this work?)

Password: (input your password)

Full name: (input your full name)

Address: (input your address)

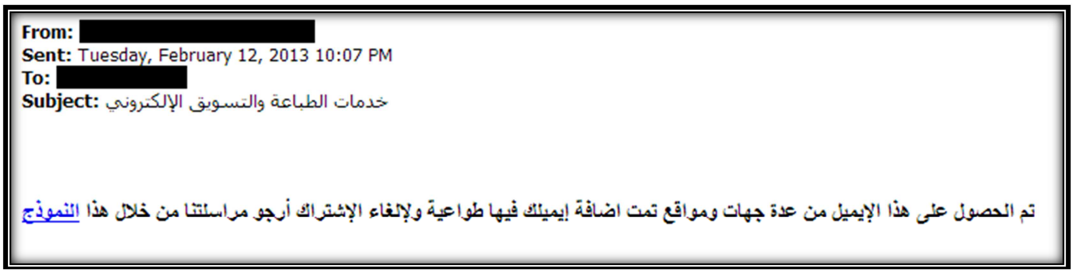
City: (input your city)

State: (input your state)

ZIP: (input your ZIP code)

Phone Number:

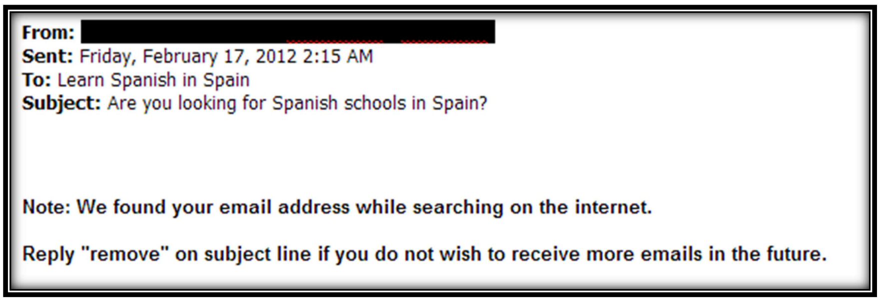
- **The recognition of spammers who send Arabic SPAM about getting the recipients email addresses by searching on the internet**



From: [redacted]
Sent: Tuesday, February 12, 2013 10:07 PM
To: [redacted]
Subject: خدمات الطباعة والتسويق الإلكتروني

تم الحصول على هذا الإيميل من عدة جهات ومواقع تمت اضافة إيميلك فيها طواعية ولإلغاء الإشتراك أرجو مراسلتنا من خلال هذا النموذج

- **The recognition of spammers who send English SPAM about getting the recipients email addresses by searching on the internet**



From: [redacted]
Sent: Friday, February 17, 2012 2:15 AM
To: Learn Spanish in Spain
Subject: Are you looking for Spanish schools in Spain?

Note: We found your email address while searching on the internet.

Reply "remove" on subject line if you do not wish to receive more emails in the future.

Appendix J: Keywords and phrases observed in Arabic and English email spam that received in Saudi Arabia

- **The common keywords and phrases observed in both Arabic and English email SPAM**

Types of email SPAM	Keywords and Phrases	
	Arabic	English
Business advertisements	دولار	Dollar
	استثمار مالي	Financial investment
	تخفيضات	Discounts
	قرض أو تمويل	Loan
	رأسمال	Fund
	اشتر الآن	Buy now
	الكرت الأخضر	Green card
	موضة	Fashion
	قريباً	Soon
	جديد	New
	برامج	Software
	أجهزة	Hardware
	مليون	Million
Religious and Political	حرية	Freedom
	تصويت	Voting
	انتخابات	Elections
Pornographic	جنس	Sex
	فياقرا	Viagra
	رومانسية	Romance
	+18	18+
	للكبار فقط	Adult only
	إباحي	Pornographic
Products and services	فرصة	Opportunity, Chance
	حلول	Solutions
	حمية	Diet
	مجاناً	Free
	ألعاب	Games
	تخفيف الوزن	Weight loss
	مفاجأة	Surprise
Phishing and Fraud	تحديث	Update
	ترقية	Upgrade
	تحذير	Warning
	تحويل	Transfer
	لديك رسالة جديدة	You have a new message
	اضغط هنا	Click here
	كلمة المرور	Password
	الحساب	Account
	فيزا كارد	Visa card
	ماستر كارد	Master card
Other types	تدريب	Training
	تعليم	Education
	ورشة عمل	Workshops
	مؤتمر	Conference
	دعوة	Invitation
	عروض	Offers
	حب	Love

Types of email SPAM	Keywords and Phrases	
	Arabic	English
	وظائف	Careers
	هدية	Gift
	جائزة	Prize
	أخبار	News
	مسابقة	Competition
	دردشة أو شات	Chatting
	التحيات	Greetings
	قصص	Stories
	ترفيه	Entertainment
	انضم إلينا	Join us
	حظ	Luck
	حمّل هنا	Download here
	اضغط واربح	Click and win
	مبروك لقد ربحت	Congratulation you have won

- **The specific keywords and phrases used in Arabic email SPAM**

Types of email SPAM	Keywords and Phrases (Arabic)
Business Advertisements	تجارة العملات
	لايصدّق
	لايفوتك
	فرص استثمارية
	عروض مميزة
	للراغبين الجادين
	جوائز قيّمة
	أزياء
	أناقة
	كشخة
	موديلات
	سارع فالعرض محدود
	اربح المال من الانترنت
	تسويق
	للميزين فقط
	خصم
	قروض ميسّرة
	تمويل مريح
	للبيع بسعر خيالي
	أسهم
	عروض العيد
	أقل الأسعار
Religious and Political	ثورة
	صوتك أمانة
	الطائفية
	محاكمة
	ديكتاتور
Pornographic	ثقّف نفسك جنسياً
	للرجال فقط
Forums	اشترك في المنتدى

Types of email SPAM	Keywords and Phrases (Arabic)
	لقد تلقيت دعوة
	نرحب بانضمامك
	المشتركين
	القروب
	المشاركات
	المواضيع
	عضو
	الردود
Products and services	حصرياً
	حساب مجاني
	طوّر مشروعك
	الحياة أسهل
	خدمات مجانية
	خاص
	أفضل المنتجات
	هل تعلم؟
	بشرى
	علاج
	أدوية
	إعلان هام
	لأول مرة
	حقق أحلامك
	دخل إضافي
Phishing and Fraud	ضاعف أمولك
	أموال
	جنيه استرليني
	شارك واربح
Other types	دورة تدريبية
	وظائف برواتب مجزية
	أعمال خيرية
	إبدأ اللعبة مجاناً
	فضيحة
	مضحك
	نصائح
	إهداء
	روعة
	نكت
	تعارف
	مشهور
	عاجل
	مغامرات مثيرة
	إعجاب
	فن
	طرب
	إبداع
	اعتذار
	وداع
	تهانينا
	زواج

Types of email SPAM	Keywords and Phrases (Arabic)
	تبرعات
	شريك العمر
	شكر
	فرح
	فرصة الفوز
	عزيزي الفائز
	أهنئك
	كيف تصبح مليونيراً؟

- **The specific keywords and phrases used in English email SPAM**

Types of email SPAM	Keywords and Phrases (English)
Business Advertisement	$USD
	% Off
	Low interest
	Partnership
	Deposit
	Refund
	Huge Income
	Bonus
Religious and Political	Violence
	Protesters
	Christmas
Pornographic	Guaranteed results
	Girls
Products and services	Award
	Promotion
	Special day
	Attention
Phishing and Fraud	Validation
	Maintenance
	Reactivate
	Revalidate
	Reconfigure
	Confirm
	Information missing
	Account not updated
	Online banking access denied
	Verify
	Incomplete personal information
	Order now
Other types	Regards
	Lottery
	Music
	Fun
	e-Card
	Cialis
	Warm regards

Types of email SPAM	Keywords and Phrases (English)
	Winner
	Become a rich man

- **The English Translation of Arabic Keywords and Phrases used in Email SPAM**

Keywords and Phrases	
ARABIC	**ENGLISH**
تجارة العملات	Currencies Trading
لايصدّق	Unbelievable
لايفوتك	Do not Miss it
فرص استثمارية	Investment Opportunities
عروض مميزة	Special Offers
للراغبين الجادين	For Serious Wishing
جوائز قيّمة	Great Prizes
أزياء	Fashion
أناقة	Style
كشخة	Dressy
موديلات	Models
سارع , فالعرض محدود	Hurry, Offer is Limited
اربح المال من الانترنت	Gain a Money from the Internet
تسويق	Marketing
للمميزين فقط	Only for Distinctive People
خصم	Discount
قروض ميسّرة	Simplified Loans
تمويل مريح	Comfortable Financing
للبيع بسعر خيالي	For Sale at Fantastic Price
أسهم	Stocks
عروض العيد	Eid Offers
أقل الأسعار	Lowest Price
ثورة	Revolution
صوتك أمانة	Your Voice is Secretariat
الطائفية	Sectarianism
محاكمة	Judgment
ديكتاتور	Dictator
ثقّف نفسك جنسياً	Educate Yourself Sexually
للرجال فقط	For Men Only
اشترك في المنتدى	Subscribe to Forum
لقد تلقيت دعوة	You have Received an Invitation
نرحب بانضمامك	Welcome to Join Us
المشتركين	Subscribers
القروب	Groups
المشاركات	Posts
المواضيع	Subjects
عضو	Member
الردود	Replies
حصرياً	Exclusive
حساب مجاني	Free Account

Keywords and Phrases	
ARABIC	**ENGLISH**
طوّر مشروعك	Develop Your Project
الحياة أسهل	Life Easier
خدمات مجانية	Free Services
خاص	Special
أفضل المنتجات	Best Products
هل تعلم؟	Do you know?
بشرى	Tidings
علاج	Treatment
أدوية	Medications
إعلان هام	Important Announcement
لأول مرة	For First Time
حقق أحلامك	Achieve Your Dreams
دخل إضافي	Additional Income
ضاعف أموالك	Increase Your Money
أموال	Funds
جنيه استرليني	Sterling
شارك واربح	Share and Win
دورة تدريبية	Training Course
وظائف برواتب مجزية	Jobs with Rewarding Salaries
أعمال خيرية	Charities
إبدأ اللعبة مجاناً	Start the Game for Free
فضيحة	Scandal
مضحك	Funny
نصائح	Tips
إهداء	Gifting
روعة	Magnificence
نكت	Jokes
تعارف	Dating
مشهور	Famous
عاجل	Urgent
مغامرات مثيرة	Exciting Adventures
إعجاب	Impressing
فن	Art
طرب	Glee
إبداع	Creativity
اعتذار	Apology
وداع	Farewell
تهانينا	Congratulations
زواج	Marriage
تبرعات	Donations
شريك العمر	Partner
شكر	Thanks
فرح	Joy
فرصة الفوز	Chance to Win
عزيزي الفائز	Dear Winner
كيف تصبح مليونيراً؟	How to Become a Millionaire?
ألعاب	Games
ريجيم	Diet
مسابقة	Competition
إربح مليون ريال سعودي	Win One Million Saudi Riyals
تعليم	Education

Keywords and Phrases	
ARABIC	**ENGLISH**
انضم إلينا	Join Us
بطاقة خضراء للسفر إلى أمريكا	Green card to Travel to USA
"18+ فمافوق"	18+ and Older
الرومانسية	Romance
مفاجآت	Surprises
برامج	Programs
جائزة	Prize
إباحية	Pornographic
اعمل من المنزل	Work From Home
قبل وبعد	Before and After
حلول عاجلة	Urgent Solutions
أفلام عالية الجودة	High Quality Movies
البرامج الأكثر فاعلية	The Most Effective Programs
هل الإنترنت صديقك المفضل؟	Is the Internet Your Best Friend?
أنت ستحصل على الأرباح	You will Get the Profits
هل تريد الحصول على دخل إضافي؟	Do you Want to Earn Extra Income?
دع الذهب يضاعف أعمالك	Let Gold Doubles Your Business
فرصة حقيقية لتحقيق أحلامك	A Real Opportunity to Achieve Your Dreams
لأول مرة في الشرق الأوسط	For the First Time in the Middle East
دعوة مجانية لحضور مؤتمر	A Free Invitation to Attend a Conference
هل أشتري آيباد أم لابتوب؟	Do I buy an iPad or a Laptop?
ممنوع الدخول لمن هم أقل من 18	No Entry to Those Who are Less than 18
تحذير هام جداً	A Very Important Warning
قروض ميسرة وبدون فوائد	Simplified Loans and Interest Free
نصائح لمقابلة عمل ناجحة	Tips for a Successful Job Interview

Appendix K: Questionnaires data management

- **Public Users Questionnaire (Part 1: Demographic Information)**

Question	Code
Region	
Central	1
Eastern	2
Western	3
Southern	4
Northern	5
Gender	
Male	1
Female	2
Age (Al-A'ali 2007; Bujang & Hussin 2010)	
15-25	1
26-35	2
36-45	3
>45	4
Nationality (Abdul-Muhmin & Al-Abdali 2011; Mohamed 2011)	
Saudi	1
Non-Saudi	2
Education Level (Abdul-Muhmin & Al-Abdali 2011; Bujang & Hussin 2010; Sait & Al-Tawil 2007)	1
High School	2
Diploma	3
Bachelor	4
Master	5
PhD	
Study Discipline	
Education and Teaching	1
Computer Science and Information Technology	2
Social Sciences	3
Physical and Biological Sciences	4
Health Sciences and Medicine	5
Other	6
Work Status (Bujang & Hussin 2010)	
Student	1
Employed	2
Work Position	
Educational	1
Medical	2
Technical	3
Management	4
Other	5

- **Business and ISP Questionnaires (Part 1: General Information)**

Question	Code
Business\ISP Size (based on the number of employees used by the EU[9]) (Kraft 2008; Pressey, Winklhofer & Tzokas 2009)	1
Small (1-49 employees)	2
Medium (50-249 employees)	3
Large (250 employees and more)	
Establishment Year (based on the internet entrance in Saudi Arabia) (Al-Ghamdi 2010; Al-Tawil 2001)	1
Before 1994 (Old)	2

[9] European Union

Question	Code
1994 till now (New)	
Business Sector (Business Questionnaire) (Ramady & Sohail 2010)	
Production and Manufacturing	1
Finance and Investment	2
Technology and Telecommunication	3
Consulting Services	4
Other businesses	5
Having business unit or team to manage network security	
Yes	1
No	2
Responsibilities of business units or teams regarding network security	
Setting up and updating Internet security software and hardware	1
Reporting security attacks to CITC	2
Designing security policies for businesses	3
Providing technical support for users regarding security issues	4
Having specific employees to combat email SPAM	
Yes	1
No	2
Tasks of employees regarding email SPAM	
Applying and updating Anti-SPAM filters	1
Reporting emails SPAM to CITC	2
Adding emails SPAM into Blacklists	3

- **Public User, Business and ISP Questionnaires (Part 2: Email SPAM Questions)**

Question	Code
Definition of Email SPAM	
UBE	1
Email was sent from unknown senders and without recipients' permission to receive it	2
Email was sent randomly and contain malicious programs such as Viruses	3
UCE	4
Annoying email that was not related to recipients' work	5
Awareness about email SPAM and Anti-SPAM filters (Public Users and Business Questionnaires)	
Yes	1
No	2
Knowledge source about email SPAM and Anti-SPAM filters	
ISPs	
Yes	1
No	2
Internet and forums	
Yes	1
No	2
Broadcast media such as TV	
Yes	1
No	2
Government	
Yes	1
No	2
School or university education (Public User Questionnaire)	
Yes	1
No	2
Other companies and organisations (Business Questionnaire)	
Yes	1

Question	Code
No	2
Receiving Email SPAM (Public Users and Businesses) \ Blocking Email SPAM (ISPs)	
Yes	1
No	2
Average number of email SPAM received on average weekly (Public User Questionnaire)	
Less than 5 SPAM	1
5-15 SPAM	2
16-25 SPAM	3
>25 SPAM	4
Email account provider	
Hotmail	1
Yahoo	2
Gmail	3
Other	4
Experience in using email	
Less than 8 years	1
8 years and more	2
Dealing with Email SPAM	
Read the entire email SPAM	
Never	1
Sometimes	2
Always	3
Delete the email SPAM	
Never	1
Sometimes	2
Always	3
Contact with ISP and notify it about email SPAM	
Never	1
Sometimes	2
Always	3
Responding to email SPAM	
Yes	1
No	2
Positive impact of email SPAM	
Purchasing and selling	
Yes	1
No	2
Learning	
Yes	1
No	2
Fun	
Yes	1
No	2
Effects of email SPAM on the performance of public users	
Yes	1
No	2
Negative impact of email SPAM	
Stealing personal information such password (Public Users)	
Yes	1
No	2
Losing time and reducing productivity (Public Users, Businesses and ISPs)	
Yes	1
No	2
Less confidence in using email (Public Users)	
Yes	1

Question	Code
No	2
Filling email inbox (Public Users)	
Yes	1
No	2
Computer infection by malicious programs such as Viruses (Public Users and Businesses)	
Yes	1
No	2
Spending money to buy or update Anti-SPAM filters (Business and ISPs)	
Yes	1
No	2
Reducing the efficiency of organisation's email server due to SPAM (Businesses)	
Yes	1
No	2
Losing customers due to receiving a large volume of email SPAM (ISPs)	
Yes	1
No	2
Consumption of the bandwidth by excessive email SPAM (ISPs)	
Yes	1
No	2
Using Anti-SPAM filters to block email SPAM (Businesses and ISPs)	
Yes	1
No	2

- **ISP Questionnaire (Part 3: Anti-SPAM Filters used and their effectiveness in Detecting Arabic and English Email SPAM)**

Question	Code
Types of Anti-SPAM filters used to block email SPAM	
Content-based filters	
Yes	1
No	2
Origin-based filters	
Yes	1
No	2
Types of content-based filters	
Iron Port	
Yes	1
No	2
Brightmail	
Yes	1
No	2
Barracuda	
Yes	1
No	2
McAfee	
Yes	1
No	2
Norman	
Yes	1
No	2
Sophos	
Yes	1
No	2
Forefront	
Yes	1

Question	Code
No	2
Symantec	
Yes	1
No	2
Mfiltro	
Yes	1
No	2
Kaspersky	
Yes	1
No	2
Types of origin-based filters	
Blacklists	
Yes	1
No	2
Whitelists	
Yes	1
No	2
Challenge Response Systems	
Yes	1
No	2
Updating Anti-SPAM filters regularly	
Yes	1
No	2

- **Public User and Business (Part 3), and ISP (Part 4) Questionnaires (Efforts to Combat Email SPAM in Saudi Arabia and the Awareness about them)**

Question	Code
Awareness about government efforts to combat SPAM (Public Users and Businesses)	
Yes	1
No	2
Government Efforts to combat SPAM (Public Users, Businesses and ISPs)	
Technical Efforts by CITC and KACST	1
Awareness Efforts by CITC	2
Receiving ISPs' reports regarding SPAM issues	3
Awareness about ISPs efforts to combat SPAM (Public Users and Businesses)	
Yes	1
No	2
ISPs' Efforts to combat SPAM (As perceived by Public Users and Businesses)	
Using Anti-SPAM filters	1
Providing awareness information about SPAM	2
Reporting SPAM related issues to CITC	3
Awareness of customers about email SPAM and Anti-SPAM filters (Businesses and ISPs)	
Yes	1
No	2
Conducting workshops and training for employees about email SPAM and Anti-SPAM filters (ISP)	
Yes	1
No	2
Period of conducting workshops and training about email SPAM (ISP)	
Every 1-3 months	1
Every 4-6 months	2
Every 7-9 months	3
Every 10-12 months	4

• Spammers' Tricks Questions

Question	Code
Using attractive words and false statements in the subject line of email spam	
Attractive words	1
False statements	2
Using different formats in writing content of email spam	
Text	1
Text embedded in an image	2
Adding links or attachment into the content of email spam	
Links	1
Attachments	2
Types of attachments	
Images (e.g. gif and jpeg)	1
Pdf files	2
Text (txt) files	3
Executable (exe) files	4
The percentages of malicious links and attachments	
Malicious links	
Yes	1
No	2
Malicious attachments	
Yes	1
No	2
Types of malicious links	
Fake bank's website link	1
Forged unsubscribe link	2
Hiding or obfuscating email addresses\Identity	
Yes	1
No	2