

0402594312



# **An approach to preventing spam using Access Codes with a combination of anti-spam mechanisms**

By  
Akhtar Hussain Khalil

A doctoral thesis submitted in partial fulfilment of the  
requirements for the award of Doctor of Philosophy of  
Loughborough University

March 2009

© By Akhtar Hussain Khalil 2009

*Dedicated to*

*My parents and my family*

## Acknowledgment

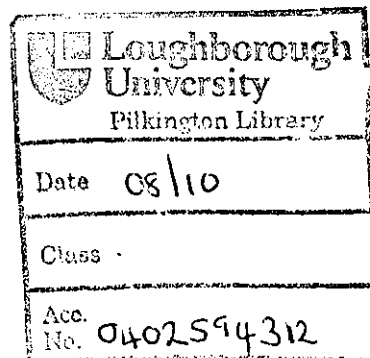
First of all I am highly grateful to Almighty God who provided me not only the opportunity but also the knowledge, skills and abilities for the successful completion of my project.

I have a very long list to offer my gratitude, but if I spell out all, this page of "Acknowledgements" would become a thick "Who's Who". Thanks to all those who made this project a success.

I begin with the name of my Project Supervisor, Prof. David J. Parish for giving me directions, motivation and advice every time I needed. You are not only a very good supervisor but a very good human too who is always kind to others and understands the problems or limitations of the students. It is only with your support and suggestions that I achieved fruitful results in time.

I will fail my duty if I do not thank my research colleagues in the High Speed Networks (HSN) group for their support and for sharing their knowledge. These include Dr. Mark Withall, Dr. Marcelline Shirantha de Silva, John Whitley, Xiaoming Wang, Shah and Jin Fan.

Last but not the least I would like to extend my special thanks to my close friends Yaqoob Juma Yaqoob Al-Raisi, Konstantinos Kyriakopoulos and Ghulam Mujtaba for being with me during some tough periods. I will cherish the sweet memories with you.



## **Abstract**

Spam is becoming a more and more severe problem for individuals, networks, organisations and businesses. The losses caused by spam are billions of dollars every year. Research shows that spam contributes more than 80% of e-mails with an increased in its growth rate every year. Spam is not limited to emails; it has started affecting other technologies like VoIP, cellular and traditional telephony, and instant messaging services. None of the approaches (including legislative, collaborative, social awareness and technological) separately or in combination with other approaches, can prevent sufficient of the spam to be deemed a solution to the spam problem.

The severity of the spam problem and the limitations of the state-of-the-Art solutions create a strong need for an efficient anti-spam mechanism that can prevent significant volumes of spam without showing any false positives. This can be achieved by an efficient anti-spam mechanism such as the proposed anti-spam mechanism known as “Spam Prevention using Access Codes”, SPAC. SPAC targets spam from two angles i.e. to prevent/block spam and to discourage spammers by making the infrastructure environment very unpleasant for them.

In addition to the idea of Access Codes, SPAC combines the ideas behind some of the key current technological anti-spam measures to increase effectiveness. The difference in this work is that SPAC uses those ideas effectively and combines them in a unique way which enables SPAC to acquire the good features of a number of technological anti-spam approaches without showing any of the drawbacks of these approaches. Sybil attacks, Dictionary attacks and address spoofing have no impact on the performance of SPAC. In fact SPAC functions in a similar way (i.e. as for unknown persons) for these sorts of attacks.

An application known as the “SPAC application” has been developed to test the performance of the SPAC mechanism. The results obtained from various tests on the SPAC application show that SPAC has a clear edge over the existing anti-spam technological approaches.

## Abbreviations and Acronyms

<b>SPAC</b>	Spam Prevention using Access Code (The proposed mechanism)
<b>VoIP</b>	Voice over IP
<b>IP</b>	Internet Protocol
<b>SPIT</b>	Spam over Internet Telephony
<b>RFC</b>	Request For Comments
<b>QoS</b>	Quality of Service
<b>TCP</b>	Transport Control Protocol
<b>CIA</b>	Confidentiality, Integrity and Availability
<b>DoS</b>	Denial of Service
<b>DDos</b>	Distributed Denial of Service
<b>URL</b>	Universal Resource Locator
<b>HTTP</b>	Hypertext Transfer Protocol
<b>MTM</b>	Man-in-the-Middle
<b>TPL</b>	Trusted Persons List
<b>BPL</b>	Blocked Persons List
<b>NPL</b>	New Persons List
<b>SPIM</b>	Spam over Instant Messaging
<b>CAPTCHA</b>	Completely Automated Public Turing Test To Tell Computers and Humans Apart
<b>IM</b>	Instant Messaging
<b>SMS</b>	Short Messaging Service
<b>OECD</b>	Organisation for Economic Co-operation and Development
<b>ITU</b>	International Telecommunication Union
<b>UCE</b>	Unsolicited Commercial Emails
<b>CEAS</b>	Conference on Email and Anti-Spam
<b>FTC</b>	Federal Trade Commission
<b>CAN-SPAM</b>	Controlling the Assault of Non-Solicited Pornography and

	Marketing (Act)
<b>ESP</b>	Email Service Provider
<b>SpotSpam</b>	Self-regulatory Plan on Tackling Spam (project)
<b>MoU</b>	Memorandum of Understanding
<b>EU</b>	European Union
<b>EC</b>	European Commission
<b>SMTP</b>	Simple Mail Transfer Protocol
<b>MTA</b>	Mail Transfer Agent
<b>DNS</b>	Domain Name System
<b>URI</b>	Uniform resource Identifier
<b>MX</b>	Mail Exchanger

## **Publications based on part of this thesis**

Akhtar H Khalil and David J. Parish, "*SPAM Prevention using Access Code (AC)*", Proceedings of Post Graduate Network, PG Net 2007, Liverpool John Moores University, Liverpool, June 2007, pp 154-158, ISBN 1-9025-6016-7.

Akhtar H. Khalil, David J. Parish, "SPIT Prevention Using the Concept of Access Number", Proceedings of the Fourth IASTED International Conference on Communication, Network and Information Security, CNIS 2007, Berkeley, California, USA, September 2007, pp 71-76, ISBN 978-0-88986-697-3

Akhtar H Khalil and David J. Parish, "*Spam Prevention using Access Code (AC)*", Multi Service Network, MSN 2008 Coseners, Abingdon, July 2008



## Table of Contents

1	Introduction .....	1
1.1	Introduction .....	1
1.2	Importance of Network Communication.....	1
1.3	Security .....	2
1.3.1	Network Infrastructure security.....	2
1.3.2	Content security .....	3
1.4	Techniques used by attackers .....	3
1.5	The CIA triad.....	3
1.5.1	Confidentiality .....	4
1.5.2	Integrity .....	4
1.5.3	Availability .....	4
1.5.4	Non-repudiation.....	5
1.6	Network Threats and Attacks .....	5
1.6.1	Denial of Service (DoS) .....	5
1.6.2	Distributive Denial of Service (DDoS) Attacks .....	6
1.6.3	Man-in-the-Middle .....	7
1.6.4	Theft of service .....	7
1.6.5	Eavesdropping .....	8
1.6.6	Impersonation .....	8
1.6.7	SPAM .....	8
1.7	Research Problem .....	9
1.8	Research Questions.....	10
1.9	Contributions .....	10
1.10	Thesis Overview .....	13
1.11	Summary.....	14
2	Spam .....	15
2.1	Introduction .....	15
2.2	History of spam .....	17
2.3	Causes of spamming.....	18
2.4	Statistics of spam .....	19
2.5	Problems Caused by Spam .....	20
2.5.1	Problems/Losses to business communication.....	20
2.5.2	Exposure to Malicious contents.....	21
2.5.3	Phishing .....	22
2.5.4	Opportunity Cost [23].....	26
2.5.5	Loss of Corporate Assets .....	27
2.5.6	Effects on the Network Resources.....	27
2.5.7	Spam contents - Annoying and offensive.....	27
2.5.8	The Legal Risk of Spam .....	28
2.5.9	Impact on Internet Usability .....	28
2.6	SPIT (Spam over Internet Telephony).....	29
2.7	Differences between email and voice spam .....	29

2.8	Summary.....	31
3	State-of-the-Art Anti-spam Mechanisms.....	32
3.1	Introduction .....	32
3.2	Anti-Spam Legislations: .....	32
3.3	Social Awareness.....	36
3.4	Collaborative Approaches .....	37
3.5	Anti-spam Technological approaches .....	38
3.5.1	Payments.....	38
3.5.2	White and Black Lists.....	39
3.5.3	Grey Listing.....	40
3.5.4	Sender Verification (Challenge/Response) .....	41
3.5.5	Cryptographic puzzles .....	43
3.5.6	Reputation Systems .....	43
3.5.7	Modifying Transmission Protocols .....	44
3.5.8	Content Filtering.....	44
3.6	Summary.....	53
4	Spam Prevention using Access Code, SPAC .....	54
4.1	Introduction to the SPAC mechanism .....	54
4.1.1	Identity (ID).....	55
4.1.2	Access Code (AC) .....	55
4.2	Spammer Vs Legitimate Client .....	56
4.3	Data Base of a User on the Server.....	57
4.3.1	Trusted Persons List (TPL).....	58
4.3.2	Blocked Persons List (BPL) .....	58
4.3.3	New Persons List (NPL).....	59
4.4	Working of the System .....	59
4.5	Accessing the Access Code (AC) from the server .....	63
4.6	Different Case Studies .....	64
4.7	Charging Mechanism.....	75
4.8	Summary.....	76
5	The SPAC Application .....	77
5.1	Introduction .....	77
5.2	Flow Chart.....	78
5.3	The SPAC Application .....	79
5.3.1	Accessing the SPAC home page: .....	79
5.3.2	Register .....	80
5.3.3	Login.....	81
5.3.4	Extension for Third Party Application / View Request.....	91
5.3.5	Off SPAC Messages .....	93
5.3.6	Auto Spam Test .....	95
5.4	Summary.....	96
6	Experiments and Results .....	98
6.1	Introduction .....	98
6.2	Number of free tokens .....	99
6.3	Results of different tests .....	100
6.4	Third party application (E-commerce websites) .....	109

6.5	Summary.....	110
7	Performance Comparison .....	112
7.1	Introduction .....	112
7.2	Comparison with the state-of-the-art anti-spam technological approaches.....	112
7.2.1	Payment .....	112
7.2.2	White and black lists.....	114
7.2.3	Greylisting .....	116
7.2.4	Challenge/response .....	117
7.2.5	Cryptographic puzzles .....	118
7.2.6	Reputation Mechanisms .....	118
7.2.7	Content Filtering.....	119
7.3	Summary.....	122
8	Conclusion .....	123
8.1	Conclusion .....	123
8.2	Areas of Application.....	124
8.3	Future Work.....	125
	Appendix A – Dreamweaver .....	133
	Appendix B – PHP (Hypertext Preprocessor [83]).....	134
	Appendix C - MySQL .....	137

# 1 *Introduction*

---

## **1.1 Introduction**

This thesis is about the prevention of spam in emails and spam in Voice over IP (VoIP) also referred to as SPIT (Spam over Internet Telephony). There are other types of spam as well which are related to short messaging service (SMS), Instant Messaging (IM), blogs etc. This thesis considers two forms of spam in a single research solution (i.e. spam in email and spam in VoIP). The thesis proposes a mechanism called SPAC (Spam Prevention using Access Codes) for prevention of email spam and SPIT. However, SPAC can be applied to prevent other forms of spam as well. To avoid the complexity, however this thesis limits itself to only spam email and SPIT. In the following paragraphs, the importance of network communication is discussed followed by network security issues and threats with particular focus on **spam**. The last sections discuss the research problem, question and contribution of the thesis.

## **1.2 Importance of Network Communication**

The Internet is an essential part of our life. Network communication is affecting the way we live. These days Internet is being used as an essential part in every field of life e.g. surveys, research, innovations, businesses, education, government, updating individuals and tackling emergencies to name a few. However, the open access of the Internet opens

users, companies and organisations to malicious threats and attacks which threaten the confidentiality, integrity and accessibility of their valuable data and resources [1].

## **1.3 Security**

There is a range of product services and data resources that the Internet offers. As the world becomes more firmly interconnected, the need for improving network security increases. Not only are the financial effects of loss greater than ever, systems are now more exposed to a wider variety of threats and in particular to risks caused by the greater access to systems offered by the Internet. The fast expansion in communication areas and the failure of current network architectures to deliver the desired security and privacy increase the need to research and embed security into the network architecture [2]. The number of books and research in network security is increasing as are the number of attacks on networks.

The data on which businesses depend comprises vital private and business assets. This includes personal, financial information, business secrets, business strategies, innovative ideas, designs, orders and plans. Loss through malicious attack or an accident could have severe business and financial consequences. A lack of public trust in a business's privacy, confidentiality, and integrity levels may lead to loss of customers and eventually can threaten the survival of an organisation/company.

Network security concerns are of two types [2]:

- Network infrastructure security
- Content security

### **1.3.1 Network Infrastructure security**

Network infrastructure security comprises securing of the physical devices that provide network connectivity and prevent unauthorized access to the management software and applications that reside on them.

### 1.3.2 Content security

Content security involves protecting information. This information may be in the form of packets being transmitted over the network or it may be stored on network attached devices. To secure the contents during transmission, different tools must be implemented on top of the underlying protocols which manage how packets are formatted, addressed and delivered. In addition security tools and protocols should be employed on the end systems as well.

## 1.4 Techniques used by attackers

In order to achieve their objectives, attackers normally create malicious programs and methods. Essential information in a computer network may also be collected before launching an attack. The following two methods are normally used in information gathering:

- **Port Scanning:** A port scanner program is used by an attacker to automatically detect any open port in a remote system. This enables him to take control of the machine. Its analogy is a thief looking for an unlocked door or window of a house. Scanners are normally TCP port scanners that attack TCP/IP ports and services (Telnet or FTP, for example), and record the response from the target.
- **Packet Sniffing:** A packet sniffer is software that captures 'packets' travelling across a network. By looking into the contents of the packets, hackers try to achieve valuable information like usernames, passwords, addresses, card numbers or the contents of e-mails.

## 1.5 The CIA triad

In this section we discuss a few topics related to security and the main threats to network security. The three key aspects of information security are referred to as the CIA triad: confidentiality, integrity and availability. Some additional aspects of security e.g., non-

repudiation or accounting have not been included in the CIA triad. In short we can say that network security provides four services: confidentiality (privacy), integrity, availability and non-repudiation [3].

### **1.5.1 Confidentiality**

Confidentiality means that the transmitted message/information must not be revealed to unauthorized persons or subjects. It must make sense only to the intended and authorized recipients. These recipients may be individuals, processes or devices. The data must be unintelligible to all others. Physical protection and encryption is normally used to ensure the privacy of a message. Strong user authentication systems and difficult to guess passwords help restrict access to unauthorised communications and data.

### **1.5.2 Integrity**

Integrity has been categorised into two categories: (1) data integrity and (2) source integrity. Data integrity is maintained by not allowing any changes in the data (either accidental or malicious) throughout its transmission, from origin to destination. Data integrity certifies that the information has not been corrupted (intentionally or unintentionally) before its reception by the intended recipient. Source integrity means that a pretender (imposter) has not sent the message. The receiver needs to confirm the sender's identity. Digital signatures can give message integrity in a secure communication.

### **1.5.3 Availability**

Availability promises that the services are available and accessible (at a rate fast enough for a system to perform its tasks) when needed by legitimate users. In the absence of service availability confidentiality and integrity are irrelevant. Service availability attacks have direct impacts on businesses and customers. In addition to the opportunity cost related to service unavailability, it results in loss of revenue and profit, system downtime

and loss of productivity. Due to these reasons service availability attacks are considered as the most damaging to networks. A Denial of Service (DoS) attack or the spread of a computer virus can cause service unavailability.

#### **1.5.4 Non-repudiation**

It means that a receiver must be able to prove the true sender of a received message. In other words the sender must not be able to deny sending a message that he/she did send. Digital signature can provide non-repudiation.

## **1.6 Network Threats and Attacks**

Threat and vulnerability are used interchangeably for issues related to network security. However there is a difference between the terms threat, attack, and vulnerability [4]. According to Webster's dictionary threat is defined as, "The expression of an intention to inflict evil or injury on another; the declaration of an evil, loss, or pain to come." The actual attempt to impact the network and its assets is referred to as attack. Vulnerability is defined by (Request for Comments) RFC 3067 [5] as, "A flaw or weakness in a system's design, implementation, or operation and management that could be exploited to violate the system's security policy." In this section we discuss some common network threats which are encountered on the Internet for almost any application including VoIP.

### **1.6.1 Denial of Service (DoS)**

The purpose of a denial of service attack (DOS) is the unavailability of the services of the targeted machine (host, switch, router, server etc) or application. In such an attack, the under attacked machine ceases or slows down its operation due to lack of allocated resources (memory, processing or communication) or by a system crash. These attacks are the most harmful to networks due to its direct impacts to business resulting in loss of customers, revenue and profit, system downtime and loss of productivity. Services like E-911 (Emergency Response Service 911 on VoIP), are susceptible to such attacks and can result in catastrophic damage. Encryption and firewall can protect VoIP networks, but



they also introduce significant delay. Above all due to the VoIP's real-time nature, data is never stored in a VoIP scenario and any packet loss is not retransmitted like ordinary data networks. Due to the small size of voice packets in voice networks loss of some packets doesn't affect the voice transmission. "Packet losses as low as one percent can make a call unintelligible, depending on the compression scheme used. A five percent loss is catastrophic, no matter how good the codec" [6]. It means that VoIP networks can easily be targeted by computer worms because the loss of bandwidth could knock out the network. There are two main classes of DoS attacks: flooding and exploitation [7]

### ***DoS Flooding Attacks***

In this attack large volumes of traffic are targeted at a system to exhaust its resources like bandwidth, CPU processing, memory or even storage. The flow of traffic exceeds the capacity of the host, server, application or circuit and exceeds the resource's capacity to operate. "A DoS flooding attack can be directed at different levels of the TCP/IP protocol stack" [7].

### ***Exploitation DoS***

Here the attacker discovers some implementation flaw which can cause a target to fail. As a result of a successful attack the targeted resource is incapable to offer services. Attackers commonly abuse URL encoding in HTTP or HTTPS requests. By doing so the attackers force a web application into doing something the designers didn't predict. Exploitation attacks are also known as implementation DoS attacks [7]. These attacks are not limited to web applications. "An attacker can send malformed SIP, H.323 and RTP packets to VoIP servers to exploit protocol implementation vulnerability, force a failure condition, and ideally, "get root" privileges (e.g. system administrator level access)" [7].

## **1.6.2 Distributive Denial of Service (DDoS) Attacks**

In a DDoS attack, a huge number of compromised systems attack a single target. An attacker delivers virus or worm to a large number of hosts and installs a Trojan program

on these hosts. This gives him control of these hosts. The attacker then uses the resources of all these hosts and targets a system. The flood of incoming messages from the compromised systems causes denial of service for users of the targeted system. The attacker exploits vulnerability in one computer system and makes it the DDoS “master”. The master program then directs Trojan programs (referred to as zombies) or multiple compromised systems. In this way, the attacker uses the resources of all these compromised systems and instructs all the machines to concurrently launch an attack against a target. Apart from the focus on the targeted host there are many victims (the compromised systems) in a DDoS attack.

### **1.6.3 Man-in-the-Middle**

In this attack, the attacker can examine and capture messages between the two hosts by inserting himself in between them. In a successful attack, the attacker is able to read, introduce and change messages between the two hosts. In addition to that the attacker can also prevent the packets destined for a recipient. MITM attack may include eavesdropping, replay attacks, phishing etc. “MITM is typically used to refer to active manipulation of the messages, rather than passively eavesdropping” [8].

In such an attack the attacker either impersonates (pretending to be the authorised user) or spoofs the address of a legitimate client [8]. The attack can be executed at many levels of the TCP/IP architecture. It can also be executed at the application layer [7].

### **1.6.4 Theft of service**

In theft of service attacks, the attacker obtains valuable services without paying for them. This is achieved by illegal means such as by stealing shared secret keys, private keys or the credentials of some legitimate client. A redirection attack is one of the worst cases of theft of service. In such an attack the service or contents (e.g. a voice message) requested by an authorized host or party can be redirected to an unauthorized host.

### **1.6.5 Eavesdropping**

As we discussed, privacy and confidentiality are the key aspects of security. Eavesdropping is actually the attempt to collect information or in other words to leak out privacy and confidentiality. In such an attack the attacker copies or listens to communication between the two hosts. It can be passive (i.e. collect, store and analyse the data) or active (i.e. decoding/translation of packets). As a result of a successful attack, an eavesdropper can discover credentials, calling patterns or other sensitive information. The captured data (audio, video, or text messaging) can be replayed or rebroadcast. [7]

### **1.6.6 Impersonation**

Here an attacker pretends to be another user or host, especially one that the intended victim trusts. For example, in phishing attacks [9] an attacker sends an email to a user and pretends to be from a trust worthy organisation (like bank, university etc). In such attacks, the attacker also directs the user to his own web site instead of the user's online banking site. The misleading website is very similar to the user's actual online banking site. If the victim fall a prey then the attacker gets his banking and/or his other sensitive information. "Address spoofing is a common form of impersonation attack" [7].

### **1.6.7 SPAM**

Unsolicited bulk messages are informally known as spam. Spam is one of the biggest threats to the internet and it costs billions of dollars every year. The impacts of spam range from loss of productivity to theft of personal information, legal and social issues to exploitation of the precious Internet and computer resources. One of the main reasons of sending spam is that it is extremely cheap for a spammer. Network maintainers, recipients (legitimate customers, companies/organisations etc) have to pay the costs of spamming. There is no state-of-art mechanism which can prevent sufficient spam and proves to be a magic bullet for the spam problem. This research focuses on spam prevention. Spam is discussed in detail in the next chapter with particular focus on the problems/losses caused by it. The inefficiency of the existing anti-spam techniques, the extent of the losses

caused by spam and high demand for an anti-spam mechanism provides motivation for research in this area.

## 1.7 Research Problem

Most of the Internet users receive unwanted messages in the form advertisements, winning notifications, fraudulent emails and malicious executable codes. The Internet resources are badly exploited by spammers for the distribution of information and for achieving financial gains. The problem is that we cannot differentiate (with full confidence) between a spammer and a legitimate client based on the contents of a header or a message. This is due to the anonymity that is inherent to the email and VoIP infrastructure. The spammer always pretends to be a legitimate client. Above all the spammers spoof addresses and PCs of innocent remote users can be easily controlled to fulfil the job of a spammer by sending auto-generated messages (bots). Various mechanisms have been proposed for spam prevention but all of these methods have a number of limitations with the likelihood of false positives and false negatives. Spam causes many problems including employee productivity loss, expenses on anti-spam measures and IT staff, exposure to viruses, spyware, adware, loss of corporate assets, misuse of network resources etc. The total world wide financial losses caused by spam are estimated at several billion USD per year [10] and [11]. This scenario forced the related authorities to combat spam with a number of approaches including legislative and technological solutions to name a few. As far as the legislative approaches are concerned so far they have shown very limited effectiveness due to the problem that the spammers spoof addresses of innocent users and senders of spam can not be traced back with sufficient reliability. In addition, state-of-the-art anti-spam technological measures have a number of limitations and negative impacts on the distribution of legitimate messages. As discussed in [12], the current filtering mechanisms are heuristic approaches, even with the best filters that are the most deployed type of technological anti-spam measures, these anti-spam measures sometimes misclassify legitimate email as spam. Another problem is that with the state-of-the-art anti-spam measures, the legitimate users of the Internet or Network have to pay the cost of fighting spam as these measures allocate the resources of

the recipients of emails. In cases like VoIP where we deal with real-time audio, the problem becomes even more severe because filtering and/or other security processes can affect the voice quality. Because of this reason most of the techniques that work (up to some extent) for spam prevention, fail completely in the scenario of SPIT (so called spam over internet telephony, SPIT).

## **1.8 Research Questions**

In order to solve the above research problem, the research needs to answer the following research questions:

- How to develop a mechanism which can filter messages of spammers from those of legitimate clients?
- How to design a mechanism which is equally efficient for preventing text spam messages (email) and for preventing SPIT (Spam over Internet Telephony)?
- How to develop a mechanism which has no negative impacts of the state-of-the-art techniques (especially of false positive, false negative and inconvenience)?
- How to develop an infrastructure so unpleasant for the spammer that he gives up and goes away?

## **1.9 Contributions**

1. The major contribution of the thesis is that it provides an anti-spam mechanism which prevents uses Access Codes to prevent the different spam attacks including spoofing attacks, sybil attacks (attacks in which the spammers change their identity or make new accounts), dictionary attacks, auto-spam (bots) etc.
2. The second contribution of the thesis is that it provides an integrated framework by combining a number of existing anti-spam techniques to achieve better results. This is done with the use of Access Codes.

3. The third contribution of the thesis is related to the unique feature of SPAC which does not show any false positive. As discussed in section 3.5.8, the existing approaches for preventing spam come with a likelihood of false positive and false negative. The results shown in chapter 6 shows that there are no false positive related to SPAC
4. The fourth contribution of the thesis is in developing a mechanism which is effective for preventing SPIT (spam over internet telephony) without affecting the voice quality. The reason is that the spammers are filtered out in the connection establishment phase. Current state-of-the-art anti-spam techniques fail in the scenario of VoIP or they provide inconvenience specially related to the Quality of Service (QoS) which is not acceptable in VoIP.
5. This work proposes a mechanism which considers not only spam email but also spam in VoIP referred to as SPIT. Spam email and spit have the same goals. However, the technical differences in these different forms of spam make the problem of spam too complex if it is considered in general in a single overview [13]. The mechanism proposed (SPAC) in this thesis is applicable for preventing spam in its different forms (such as spam email, SPIT).
6. SPAC takes into account the effectiveness from the user perspective but also takes into account the prevention of the consumption of precious Internet resources which is not possible with the existing content filtering techniques. As mentioned by Enrico Blanzieri and Anton Bryl [13], filtering on the destination point solves the problems caused by spam only partially. This prevents end-users from wasting their time on junk messages but since all the messages are delivered nevertheless, this mechanism does not prevent resource misuse. In addition existing content filtering mechanisms do a lot of processing on the contents of the message. This uses the network resources and adds to the carbon foot print of IT. Above all the network gets congested due to the huge amount of spam data. As opposed to this mechanism, SPAC does not allow spam traffic (data) to use the network resources. In fact a spammer, (who is not and

who can not be in the Trusted Persons List, TPL list discussed later) can not get even a connection which is required before sending the data. So, the spam traffic (data) does not access the network which saves internet resources in addition to preventing end users from the nuisance of spam.

7. The SPAC mechanism provides an alternative solution for the problems caused by the lack of technical definition of spam as discussed in section 3.5.8. Messages related to jokes, legitimate marketing or political messages might be spam for most of us but many would still like to receive them. This problem can not be solved by existing spam filters. SPAC solves this problem by providing a SPAC enable and SPAC disable feature as discussed in chapter 6 (section 5.3.2). Also as discussed by Andy Walker [14] in the existing mechanisms, during registration, if a user agrees to receive email from an organisation (opt-in) then the email that the organisation sends is not considered as spam. This happens with a lot of users. In order to stop such emails the user needs to go back to the company's website and find out how to opt-out (unsubscribe). Most reputable companies have a mechanism for unsubscribing but not every company does have. It means that the control is with the company. A similar incident was mentioned by Andy Walker [14] in his book where he mentions, "Out of curiosity, I bought a list of 10 million Canadian email addresses for \$49. The company claimed they were all opt-in email addresses, meaning that the owners of the addresses had agreed to be put on the list. I found one of my addresses that was used for inbound mail only, however. It was never used to opt into anything". The author didn't opt-in for the emails but the company claimed that all these users including the author did. SPAC gives this control to the user. At the same time to know about the mechanism of unsubscribing is difficult because it is different for each company. This also adds to the inconvenience of the existing mechanism. As discussed in section 5.3.2, instead of going through the opt-out mechanism of each company individually, a SPAC user can enable or disable such messages (from all the companies) with a single setting in his own account.

8. Since SPAC prevents spam on the connection establishment phase so it can prevent spam in various forms (such as voice, audio, text and image) without affecting their quality. This means that SPAC can be applied to a number of other technologies such as cellular telephony, traditional telephony and instant messaging service.
9. SPAC targets spam from two angles i.e. to prevent/block spam and to discourage spammers by making the infrastructure environment very unpleasant for them.

## **1.10 Thesis Overview**

The thesis is organized in the following way:

Chapter 1 (this chapter) gives an insight into the research work. It discusses the importance of network communication and the network security threats and issues. It also discusses the research problem, the questions to be addressed by the thesis and the contribution of the thesis.

Chapter 2 is dedicated to spam. It reviews the definition, history, aims and problems caused by spam. This chapter also introduces spam in VoIP (referred to as SPIT). The differences between spam email and SPIT is also presented in this chapter. The chapter reveals that spam is a major threat to the most attractive applications of Internet like email and VoIP and expresses the extreme need for an anti-spam solution.

Chapter 3 explores the state-of-the-art approaches for preventing spam. The chapter discusses the legislative, social, collaborative and technological anti-spam approaches and shows the inefficiency of these state-of-the-art approaches and legislations. The outcomes of the detailed discussion in this chapter generate a number of research questions and areas which need to be addressed in the war against spam.

Chapter 4 of the thesis proposes a novel anti-spam mechanism which is called “Spam Prevention using Access Code”, SPAC. The chapter discusses the detailed



working of the mechanism and analyses the system in different case studies. The last section of the chapter is a brief glance on the novelty verification of the mechanism.

Chapter 5 overviews the SPAC application which has been used to perform different tests and experiments. It shows how to use the SPAC application and gives results and screen shots of some practical experiments.

Chapter 6 provides the different results obtained from different tests and experiments.

Chapter 7 compares the performance of SPAC with state-of-the-art mechanisms and mentions the other areas of its applications.

Chapter 8 summarises the conclusion and gives a glance of the future research work.

## **1.11 Summary**

Network communications are an integral part of our daily routines and affect the way we live. The open access of the Internet opens individual users and companies to malicious threats and attacks which threaten the confidentiality, integrity and accessibility of their valuable data, resources and businesses. The three key aspects of information security are referred to as the CIA triad: confidentiality, integrity and availability. Unsolicited bulk messages are informally known as spam. Spam is one of the biggest threats to the internet and it costs billions of dollars every year. The objective of this research work is to develop an efficient anti-spam mechanism which can stop enough amount of spam. The thesis makes significant contributions to this particular research area.

# 2 Spam

## 2.1 Introduction

As discussed in [15], the word Spam originally referred to “*a canned meat product*” (shown in the following figure) sold by the Hormel Foods Corporation.

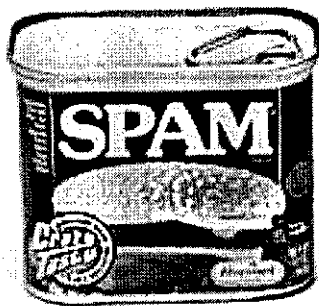


Fig 1, SPAM family of products [16]

The word spam was later adopted by the Monty Python comedy team in their SPAM song which was a comedy sketch set in a restaurant where every dish contained spam, the canned meat product [15]. Since then, many other uses of the term have emerged. Most later uses of "spam" refer to undesirable repetition. In technology's literature the word

spam appears for unsolicited or unwanted bulk messages (e.g. emails, mobile phone calls, sms, blogs, VoIP, instant messaging etc). The use of the word spam for unwanted messages in electronics is named after the Monty Python sketch [15]. Besides the literal meaning of spam, there is no single agreed technical definition of spam. [11], [22] However, “unsolicited bulk messages” are informally referred to as spam. Sometimes spam is also called junk mail [17], [18]. One of the widely accepted definition of spam is, “Internet spam is one or more unsolicited messages, sent or posted as part of a larger collection of messages, all having substantially identical content” [19]. According to TREC Spam Track, spam is “unsolicited, unwanted email that was sent indiscriminately, directly or indirectly, by a sender having no current relationship with the user” [20]. ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT [11] has classified the characteristics of spam. According to ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT the primary characteristics of spam include: unsolicited, electronic message, sent in bulk and commercial. There are some arguments about the commercial characteristics of spam [13]. The remaining characteristics as recognized in many definitions are the secondary characteristics. These characteristics have been given in the following table:

<b>Primary characteristics</b>	<b>Secondary characteristics</b>
Electronic message	Uses addresses collected without prior consent or knowledge
Sent in bulk	Unwanted
Unsolicited	Repetitive
Commercial	Untargeted and indiscriminate
	Unstoppable
	Anonymous and/or disguised
	Illegal or offensive content
	Deceptive or fraudulent content

**Table 1: Primary and secondary characteristics of spam [11]**

Despite the lack of a single definition, the widely accepted and agreed upon general characteristics of spam are [21]: Electronic message<sup>1</sup>, unsolicited message and message in large volumes.

<sup>1</sup> Includes email, SMS, VoIP, IM services, Mobile phone Multimedia Messaging Services (MMS) etc

All these definitions convey the message that spam is unsolicited and sent in bulk. The problem is that solicitation can not be confirmed because opinions about a message can be divided. As discussed in section 3.5.8, a message can be unwanted (unsolicited) for most of us but some of us would like to receive it. So far there is no precise definition that can be used by servers to filter spam from legitimate messages (also known as non-spam, genuine message or ham) [22]. It should be noticed that the term spam is normally used to refer to spam in email. However, the word spam can be used for spam in other applications such as in VoIP, sms, blogs, mobile phone calls and IM etc. In VoIP, spam is also called SPIT (Spam over Internet Telephony) and spam in instant messaging is also called SPIM (spam in instant messaging). The work of this thesis focuses on spam prevention in email and spam in VoIP (SPIT). So, in this thesis we will use the word spam for both types of spam messages (spam email and spit). However, where necessary we would be using the particular terminology. The following section reviews a brief history of spam. The chapter presents a summary of the problems caused by spam. We will also discuss SPIT and the differences between email and voice spam in this chapter.

## **2.2 History of spam**

According to Guido Schryen [23], spam was first considered a problem in an RFC in 1975 [24] and it first appeared in the publications of ACM in 1982 [25]. Michael Specter [26] and Templeton [27] have discussed the story of the first spam message which was sent by a marketing representative named Gary Thuerk from Digital Equipment Corporation (DEC) in the spring of 1978. He wanted to inform people about a powerful new computer system to be introduced by his company (DEC). He decided that the best way to introduce the product was to send a message to every Arpanet (the predecessor of the Internet) address. A very interesting story about this first spam is discussed in detail in [26] and [27]. With the push of the send button Thuerk became the father of spam. The reaction to the first spam sent by Thuerk was instant and completely aggressive. One of the recipients wrote, "This was a flagrant violation of the Arpanet". Thuerk was harshly scolded. However DEC sold more than twenty of the computer systems for a million

dollars each. Selling this number of computers was a great achievement for DEC in terms of profit. However, even at that time several viewed the network as a rising symbol of intellectual freedom and did not see any harm in such actions. The Internet pioneer Richard Stallman was in favour of the action by Thuerk. After a few days of the DEC email, he wrote, "The amount of harm done by any of the cited unfair things the net has been used for is clearly very small". He was among those people who were in the favour of the openness that defines the web and opposed any action that places limits to the exploitation of this powerful new tool. According to him, the network provided a unique opportunity to advertise jobs and an entirely new way to sell products. He added: "Would a dating service on the net be frowned upon ....? I hope not. But even if it is, don't let that stop you from notifying me via net mail if you start one." This historic story also reveals a number of facts about spam including the different reactions and its benefits (to the spammer) to name a few. It also shows that (like today) opinions about a spam message were divided. It was unsolicited for some but not for others like Richard Stallman, the Internet pioneer. Michael Specter [26] says that he has no idea whether anyone on the Arpanet tried to help Stallman find a date, but thousands of people have tried to help him (Michael Specter). He claims that he receives dozens of email from dating and adult sites today. Specter means to say that the problem of spam which was underestimated for many years has become a looming problem these days. Some interesting examples of early spam attacks are given by Templeton [28].

## **2.3 Causes of spamming**

Spammers send spams because they want to make quick money. The price of spamming is only a fraction of the cost of other marketing and delivery methods with which a merchant can reach thousands and millions of potential customers. Very small percentages of the recipients of spam emails click the offers in a spam email and buy the advertised product. [29]. However, this small percentage of a huge number (some million) of recipients makes spamming more cost-effective than any other advertising method. Bekman [30] has discussed the following findings. Spammers can send a very large amount of spam emails at the cost of a few hundred dollars. Assuming a profit of

\$25 from each sale, a spammer can make an immediate profit of \$10K by sending slightly more than 2 million spam emails which is not a difficult job for bots. A study by ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT [11], in 2003, reveals that the accumulated economic impact of phishing attacks (see section 2.5.3) was estimated at USD 222 billion. In addition a phisher gets an average profit of USD 5000 per successful transaction.

## 2.4 Statistics of spam

In their book on “Understanding Voice over IP Security”, Johnston and Piscitello [7] mentioned that spam exceeds legitimate messages by 4 times. The following figure shows a dramatic increase in the percentage of spam from Dec 2001 to Nov 2005.

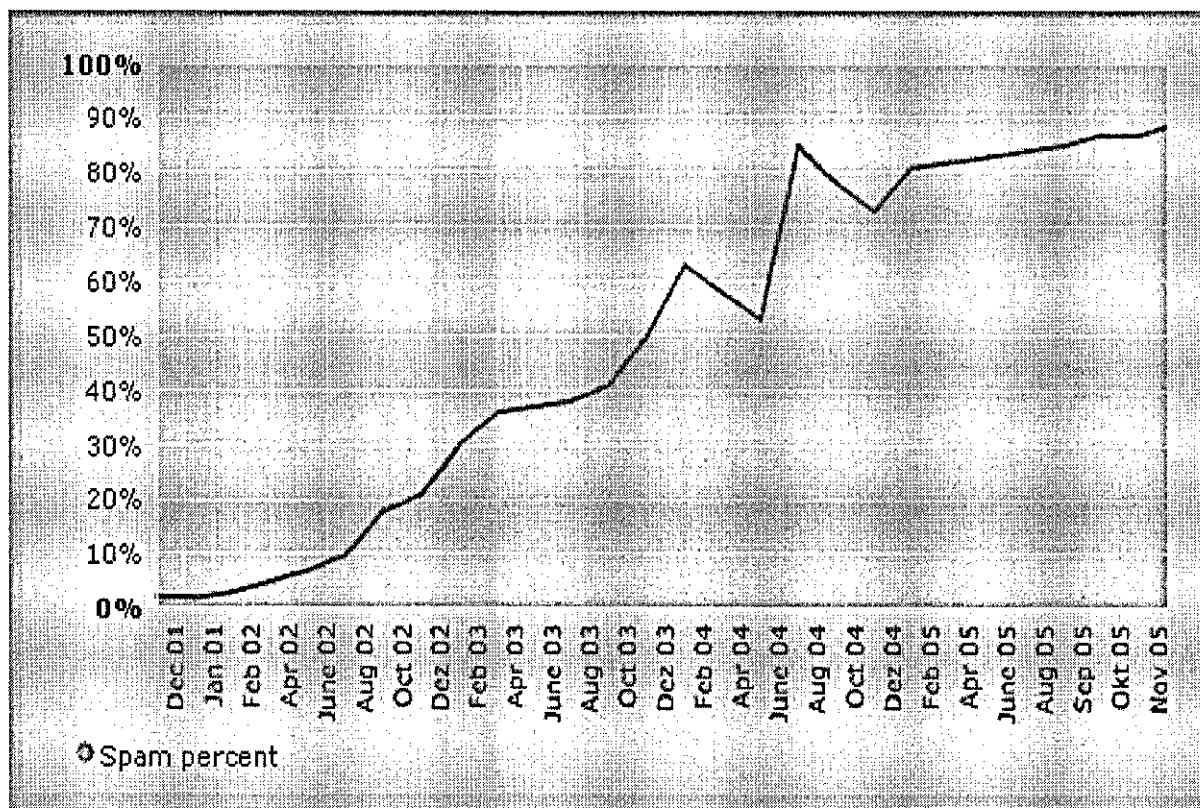


Fig 2, percentage of spam emails in the total email messages [31]

Most of the studies like [21] and [31] show that spam emails constitute 80% to 90% of the emails. The reason for the difference in different studies is that each study considers different definitions of spam and is conducted at different times. Some recent studies show that the percentage of spam emails has increased to more than 90%. Since the state-of-the-art technologies do not cope well with the spam problem, the growth rate is expected to increase in future.

## **2.5 Problems Caused by Spam**

Spam is a nuisance that has its impacts not only on organisations but also on individuals. Looking through spam and sorting out emails waste a significant amount of time. This in turn causes not only loss of work productivity but also irritates the users. The problems caused by spam have been discussed in detail in [32], [10], [33], [34]. In this section we discuss some major problems caused by spam.

### **2.5.1 Problems/Losses to business communication**

The frightening spam problem has been undervalued due to a lack of investigation of the economical impacts of spam. The damages estimated by Ferris Research [33] was 8.9 billion USD for U.S.-American companies and around 2.5 billion USD for European companies (in 2003). In terms of loss of productivity, Marten Nelson of Ferris Research found that damage caused by spam (for US-American companies) is up to 4 billion USD annual. The damage for U.S.-American and European ISPs (Internet Service Providers) is calculated to be around 500 million USD. Key findings of the Nucleus Research in 2003 [10] included the following:

“The average employee receives 13.3 spam messages per day.

Time spent per person managing spam ranges from 1 minute to 90 minutes per day, with an average of 6.5 minutes.

**Average lost productivity per employee per year: 1.4%**

*Calculation:* 6.5 minutes/day divided by 480 total minutes/day

**Average cost of spam per employee per year: \$874**

*Calculation:* 1.4% times 2080 hours at an average fully loaded cost of \$30/hour

**For every 690 employees, a full-time IT staff person will be needed just to manage spam.”**

Nucleus found that spam impacted the productivity of some employees to the extent that they invested in desktop filters and learned to use them to combat their spam problem. Even with desktop filters, these individuals still spent an average of 12.5 minutes per day – nearly twice the average – screening and managing incoming mail, at a cost of \$1,625 per year in lost productivity. In addition the study by Nucleus revealed that apart from productivity and IT impact, many companies worry that even with filters; unsolicited e-mail sent to employees may provoke legal action

All the above figures are for 2003. Recent research shows far more losses than these. *Ferris Research Analyzer Information Service [33] estimated that the total worldwide financial losses caused by spam in 2005 were \$50 billion*

The Radicati Group estimated that 4.9 trillion spam e-mails would be sent in 2003 with an increased growth rate in the future. [11]

The estimated numbers are different for different studies. The reason is that different studies considers different types of spam, harms and assumptions etc. However all studies reveal that the spam causes hundreds billion USD loss per annum with an increase in growth rate.

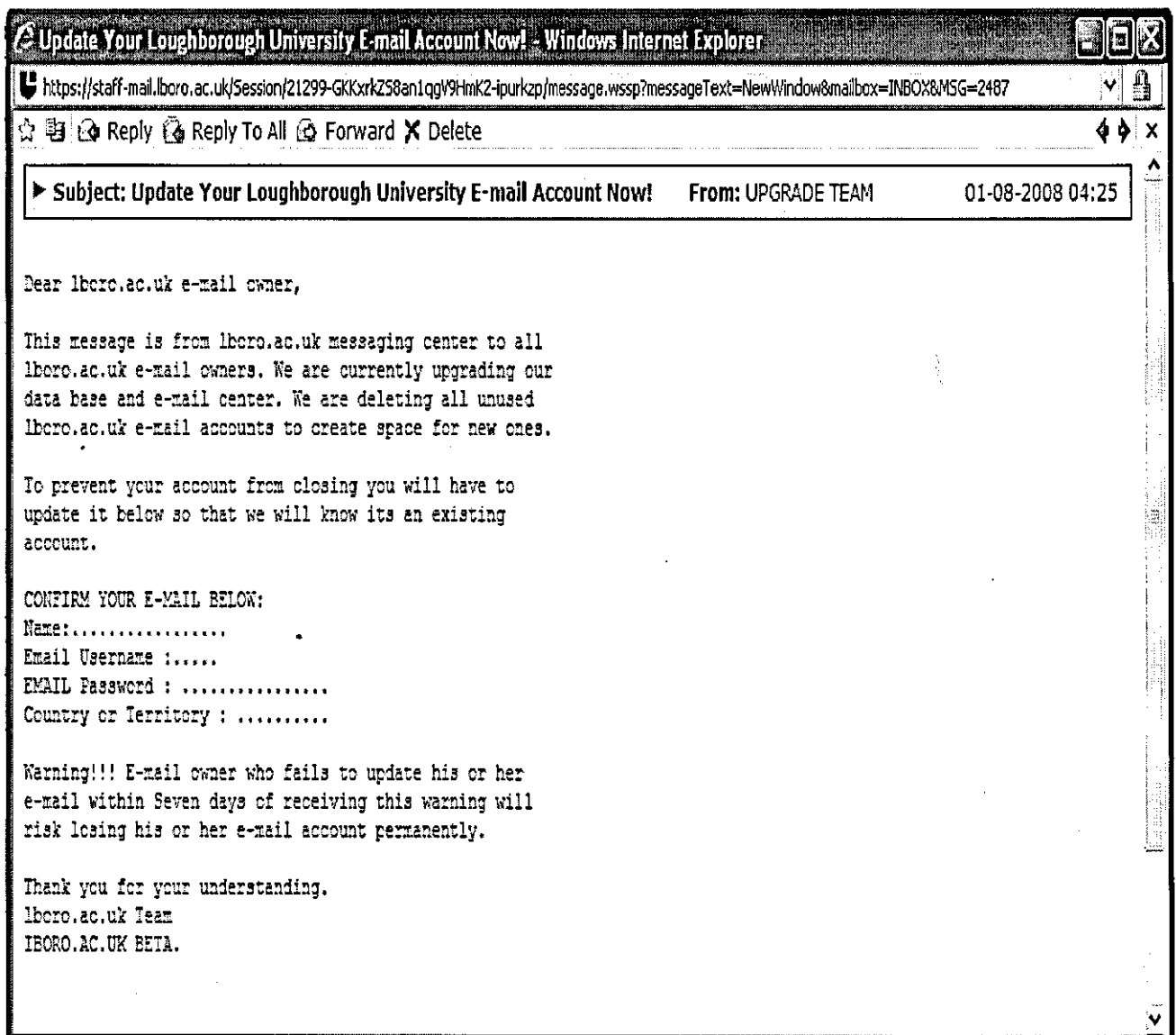
## **2.5.2 Exposure to Malicious contents**



One of the most threatening malicious spam contents are viruses. Spams are sometimes used by spammers to carry hidden viruses that can get into a system and infect it. The infected system then impacts the entire network. “Spyware and adware are other problems created by spam that disrupt the flow of your business”. [34] It is reported by Symantec’s Threat Team in their biannual report on Internet Threat that more than 30,000 pcs per day were being recruited into secret networks to spread spam and viruses in the first 6 months of 2004 [35].

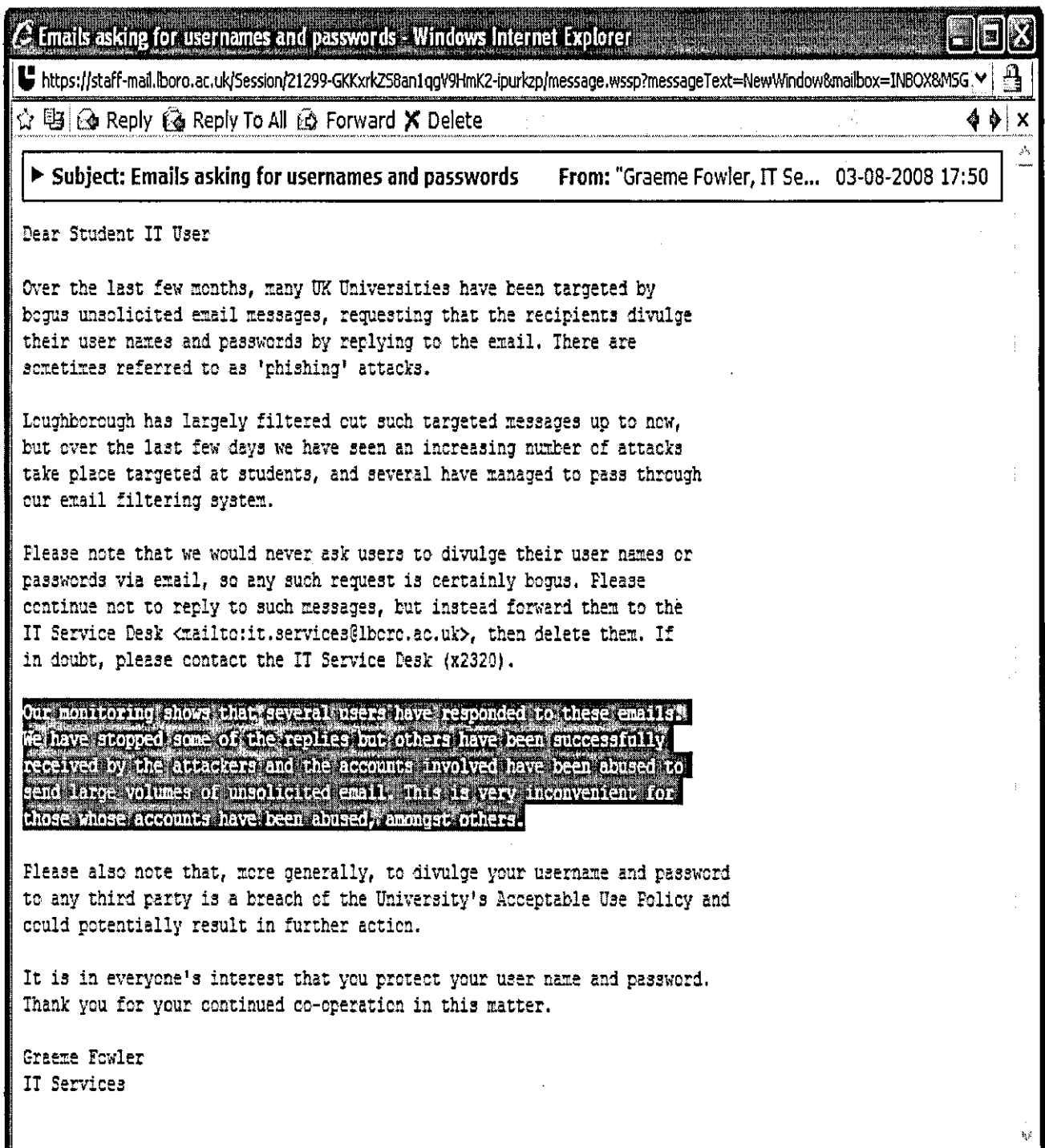
### **2.5.3 Phishing**

Phishing is one of the most irritating classes of spam which is used for online frauds. These fraudulent emails are also called “scam”. Such emails misguide or lead to a deceptive activity on the part of sender. In a Phishing attack, a spammer accesses personal information of the victim (like credit/debit card details or other personal information). In a phishing attack, a spammer typically asks the victim to click on a hyperlink which takes them to a fake web site. These fake web sites are in the control of spammers. They do so by pretending to be from trust worthy institutions (such as a bank, your employer, university or server administration). This is also called “brand spoofing”. While writing this section of the thesis, the author received a phishing email pretending to be from his university and which asks him for his username and password. This phishing email is as shown in the figure below:



**Fig 3, example of phishing email from the author's personal e-mail inbox**

In response to this phishing email, the email from the IT services department of my university (Loughborough) shows that most of the users did not respond to the message but it is amazing that many students of the university (who are mature and are in higher education) responded to these phishing emails. The email from the IT services is shown in the figure below. The highlighted text shows that several students replied to the email. This shows the severity of the phishing problem which is delivered as spam.

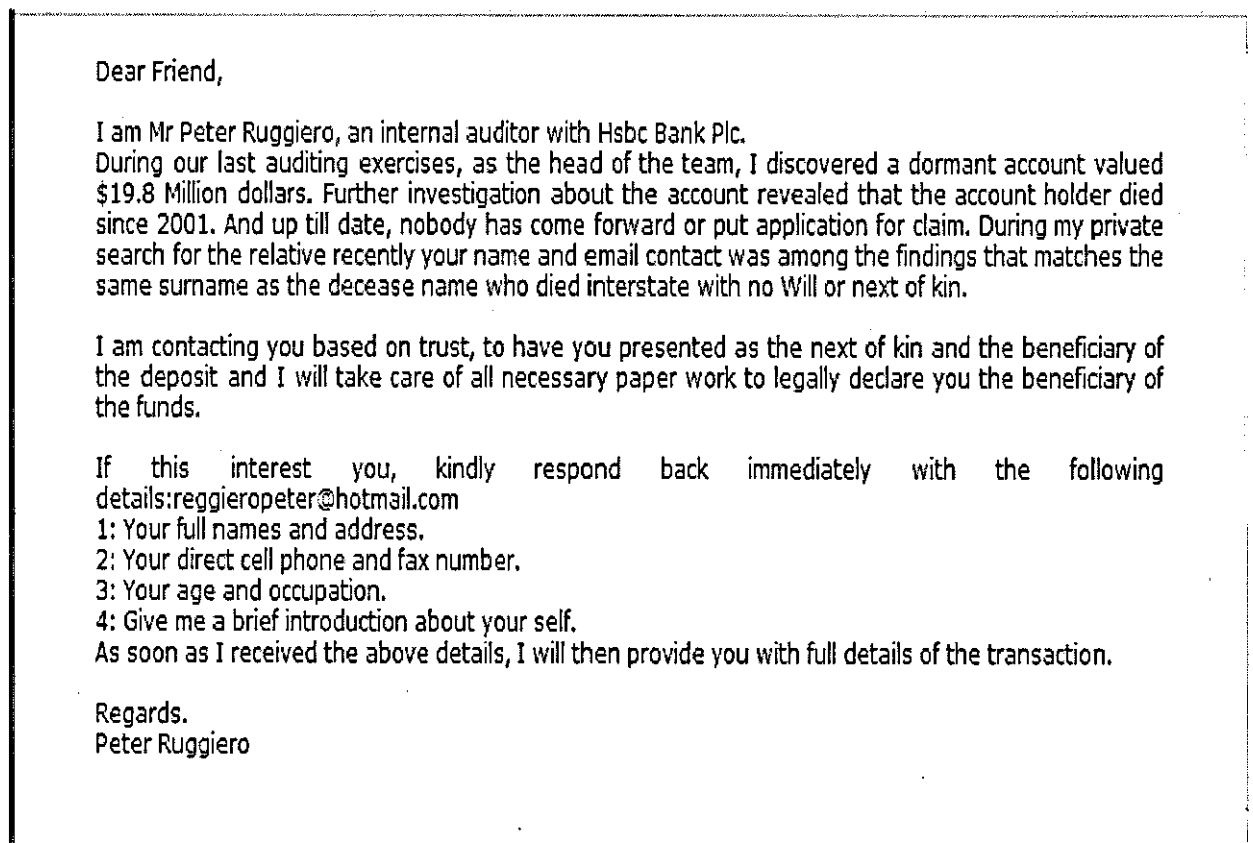


**Fig. 4, email received from IT support in response to the phishing email in fig 3**

The contents of this practical email show that:

- Filtering which is the most popular technique for spam prevention can not stop all spam.
- Even highly educated people fall a prey to phishing attacks
- Spam produces inconvenience not only for the immediate victims but also for other network users
- The second last paragraph discusses some policies which show that spam messages can result in legal issues

This is an example from a university but such attacks develop more severe problems and cause huge losses to big companies and organisations. In an email to the Loughborough email account users, the Director of IT, Loughborough University, revealed that 80% of the incoming emails to Loughborough University are filtered as spam and this figure continues to rise. Another example of such phishing attacks is shown in fig 5 where an attacker is interested in the details of the recipient.



**Fig 5, example 3 of a phishing attack**

In such an attack the attacker would first like to gain the trust of the recipient. If the recipient replies then later on the attacker asks him/her for more and more confidential/personal information. Mostly these types of phishing attacks also contain a hyperlink and the recipient is asked to click on the hyperlink to access his or her online banking account. This leads the user to a web page that is similar to the web page of HSBC in design but is actually a fake web page under the control of the spammer. They promise money but first demand money from the users as a service charge to provide start up costs or some insurance. To convince the recipient, they make a very good story and ask him/her to keep it confidential. In addition spams pretend to be replies to inquiries to get people into opening these messages. Opening a mysterious email containing a virus puts the system, the entire network and customer information at risk. At the same time it is difficult to drop an email just on the basis of suspicions. In some Phishing attacks the attacker asks the victim to call a telephone number that is controlled by a phisher. Phishing is not restricted to emails; it also attacks the victim in instant messaging, sms messages and also on mobile or VoIP phones.

#### **2.5.4 Opportunity Cost [23]**

A massive spam attack may result in the break of an email system which can lead a company or an individual miss an important email. This will cause direct loss of revenue. In addition, the most popular anti-spam mechanisms like content filtering do not give 100% accurate results. Sometimes they result in false positives and false negatives. Classifying a legitimate message as a spam message (false positive) can block the email from a potential customer which can be potentially very costly for the companies and organisations (as discussed in section 2.5 of chapter 2). This may also result in the loss of existing customers and the opportunity to obtain new customers.

### **2.5.5 Loss of Corporate Assets**

Inappropriate content is often leaked by spam into corporate email accounts. Much of the material in junk email is offensive and adult material. This can lead to disputes, costly ramifications as well as legal issues. Human resource departments would surely like to avoid these problems [34].

### **2.5.6 Effects on the Network Resources**

One of the major impacts of spam is the consumption of computer and network resources. Spam exceeds the required amount of memory and computational power and consumes a lot of network bandwidth. This reduces the amount of data that can be transmitted in a fixed amount of time. Organizations may experience storage loss through backing up emails that are spam [34]. The spam statistics in section 2.4 show that spam exceeds 4 times legitimate messages. This means that as compared to legitimate emails, our internet resources are used 4 times more by spam emails. At the same time this huge amount of spam causes additional traffic in the Internet which in turn produces the need for additional storage space and additional computational power [37]. All these loss of network resources increases the cost of operations which is also passed on to customers.

### **2.5.7 Spam contents - Annoying and offensive**

The advertising messages in spam annoy the users. These advertising messages are mostly of no interest for the users. Especially in case of VoIP the repeating ringing tones of spam calls early in the morning (say at 2 AM) can be a major threat and fear not only for the recipient of a VoIP user but also for the people surrounding him. Sometimes the contents of spam are offensive and distasteful. It exposes adults and children to adult contents.

### **2.5.8 The Legal Risk of Spam**

Spammers use adware and spyware to turn the computers of innocent persons into spam slaves which force their network to do the work of the spammer. If an individual or a company's computers become victims of such an attack, they can face a lot of problems due to the outgoing spam from their computers. Although many companies neglect this problem, however, if spam can be traced to their computer systems they can be totally banned by some email servers. This will prevent legitimate email from going through. At the same time these innocent persons can be involved legally for violation of the CAN-SPAM rules (discussed in section 3.2), even though their systems were just zombies, not willing participants" [34]. In case of successful phishing attacks, a recipient gives personal information or some confidential information to an attacker. This can lead to severe legal issues and can put him/her in risk of losing a job or money (in case of credit/debit card details). The second last paragraph of fig 4 in section 2.5.3, states that divulging the username and password to any third party is a breach of the University's Acceptable Use Policy and could potentially result in further action. The same figure also shows that several users have responded to the phishing emails targeted at the university's email accounts. Apart from that there exist other legal problems associated with spam. These include problems related to advertising pornography, pyramid schemes etc [38].

### **2.5.9 Impact on Internet Usability**

Some of the impacts of spam on Internet usability have been mentioned in [39]. By making the cost of using it prohibitively high, spam makes the Internet less capable of supplying information. A lot of time and money is spent on deleting and downloading spam. This makes the use of online information less effective. In places like the United States, phone service and Internet access is usually paid by monthly fee. In such places the issue becomes a matter of time rather than money. Even in such cases the cost increases for service providers and network maintainers. This affects the overall cost/fees. Statistics show that spam is increasing each year. If people are going to receive large amounts of spam, they may be forced to give up much of the Internet experience because

they must take so much time going through spam. By the time they go through the spam, they may not be able to afford to remain online any longer [39].

## **2.6 SPIT (Spam over Internet Telephony)**

The problem of spam becomes even tougher when considering it in the case of VoIP. In VoIP spam calls are referred to as SPIT (spam over internet telephony). SPIT is one of the two major threats to VoIP and is expected to become a more severe and serious problem in the near future. The real challenge is to block a spam call before the telephone rings. Most of the antispam techniques which helped in limiting the amount of spam, fail totally in preventing SPIT because in VoIP we deal with real time voice which doesn't accept the processing time without affecting the QoS [40]. Above all it is not a good idea to drop a call just on the basis of suspicion. This can drop important calls [40] which will be of loss to the VoIP user.

## **2.7 Differences between email and voice spam**

SPIT is similar to email spam in many aspects. For example the causes/aims and impacts of spit and spam are similar. However, spam email deals with text while SPIT works with voice. Therefore, some of the popular techniques used for preventing spam emails fail in the scenario of SPIT due to the real time behaviour of VoIP. In addition, when compared to spam email the impacts of SPIT are more immediate and in most cases more severe.

Currently the rate of SPIT calls is very low when compare to the rate of spam emails. However, with the increase in the popularity of VoIP, it is expected that SPIT will occur with a frequency similar to email spam in the near future. Although email spam is a nuisance, the impacts of SPIT on a user are far more irritating than email spam. SPIT upsets instantly and a SPIT call will disturb not only you (the recipient) but also your colleagues surrounding you. Emails don't interrupt the users and can be checked at intervals. However phone calls are interrupting and they need attention. The recipient of a



call either answers it or at least checks to see who the caller is. This means that the reaction of users to phone calls is as soon as they receive it.

Deleting a spam email is quite simple when compared to dealing with SPIT calls especially when they are received at midnight or early in the morning. At the same time phone or VoIP calls are more urgent, interrupting, and attention-getting than an email due to which we can not turn off our mobile phone. In cases where we want to convey urgent messages, we use telephony services so a VoIP user will receive an immediate ring in the case of a SPIT call. Receiving a number of SPIT calls (say equal in number to the average number of spam emails a user receives) in the voice mail would be very annoying. Relative to email spam it takes more time to delete spit calls because in the later case the user would have to listen to every message (at least a portion of it) before deleting it. This means increased loss of productivity as compare to email spam. In addition the impact of SPIT on network resources is more severe when compared to email spam because it has the potential of clogging up the network [41] and consumes more bandwidth. This may also force the users to leave VoIP.

Another issue with SPIT is that we cannot perform filtering based on the contents of the call. Content filtering is the most popular approach for preventing spam but most of these antispam techniques which helped in limiting the amount of spam, failed totally while dealing with VoIP. As QoS is not an issue with email so an email is first received by a server that can apply many filtering strategies and then makes it available to be downloaded by the user. In contrast, VoIP deals with voices rather than text. *“To recognize voices and to determine whether the message is a spit or not is still a very difficult task for a computer”* [40].

Another reason which makes it hard or nearly impractical to use content filtering mechanisms for preventing SPIT calls is that in case of VoIP a recipient finds out about the message (or contents of the call) when he or she has received the call and is listening to it. This means that content filtering techniques can be applied only after the user has received the call. That is by the time when the user is subjected to all the impacts of

spam. Also VoIP is very sensitive to delay and requires a certain degree of QoS. Using filtering techniques on the contents of voice could significantly degrade the quality of sound. Also it is relatively very easy for the user to delete an email spam. However, receiving a SPIT call means the telephone rings and disturbs the users at any time of day. Also it is not good to drop a call just on the basis of doubt. This can drop important calls [40] which will be of loss to the VoIP user. In cases where we want to convey urgent messages, we use telephony services. So in case of VoIP phones we have to set our phone to ring mode for the unidentified persons and we will receive ring tones for regular calls and also for SPIT calls which will of course be unsatisfactory.

David Endler and Mark Collier [42] have expressed the severity of SPIT relative to email spam in the following phrase which shows the user reaction to SPIT:

*"I don't know what is worse. Digging through my voicemail and deleting all the SPAM or getting 25 calls a day trying to sell me Viagra. I think I am going to just turn this stupid phone off".*

## 2.8 Summary

Spam is a major threat to the efficiency of the Internet. Spam emails are 4 times (80%) more frequent than ham emails. The lack of a single agreed definition of spam contributes to the inefficiency of the filtering mechanisms. Spam creates a lot of problems including impacts on businesses, phishing, loss of corporate assets, legal issues, impacts on network resources and exposure to viruses, spyware and Adware to name a few. The losses caused by spam are billions of dollars every year. The growth rate of spam is constantly increasing due to the inefficiency of the state-of-the-art antispam techniques. It is expected that spammers will badly exploit the network resources if researchers don't come up with an efficient anti-spam mechanism. The huge increase in spam is the cause for the unexpected overload in bandwidth, server storage capacity and loss of end-user productivity for email systems. The problem of SPIT is more severe than spam.

# 3 *State-of-the-Art Anti-spam Mechanisms*

---

## **3.1 Introduction**

A number of measures have been proposed and implemented to prevent spam but the fact is that none of the approaches has been able to prevent sufficient spam to solve the problem. To solve the problem the anti-spam approaches can be divided into four broad categories. These include legislative, technological, collaborative and social awareness. The legislative approaches make spam illegal whereas the social awareness (or educating users) aim at educating the users to learn about spam, its causes, problems and how to minimise or get rid of this nuisance. The collaborative approaches work for possible types of cooperation between national authorities, companies and users. The purpose of technological approaches is to prevent the delivery of spam and/or to make it difficult [43]. In this chapter we discuss these anti-spam approaches in detail. In addition to that we will discuss some potential problems caused by the existing anti-spam mechanisms. We will see that current mechanisms are not efficient enough to overcome the problems caused by spam.

## **3.2 Anti-Spam Legislations:**

The severity and the extent of problems caused by spam (as discussed in section 2.5) both to users and organisations have compelled many countries and federal states to implement legislation for preventing spam. Many countries including the United States, the European Union, Australia, Canada and United Kingdom have recently implemented legislation to combat Unsolicited Commercial Communication. Some of the most prominent legislations regarding spam prevention are European Union (EU) Privacy and Electronic Communications Directive, and US CAN-SPAM Act.

The **Privacy and Electronic Communications Directive 2002/58/EC** [44] was passed by European Parliament in July 2002. Each EU member had to implement this legislation by 31<sup>st</sup> October 2003. This law focused on: Protecting the rights of people by reducing spam and guaranteeing the individual's control over personal relationships and contacts. The directive regulates some sort of opt-in approaches (which requires that the sender has the recipient's permission prior to sending). The directive also requires each marketing email to contain information about the Opt-out mechanism (in which the recipient is provided with information on how to decline and stop the receipt of further emails from the sender). An overview of the directive is given by Nicola Lugaresi in [45]. Article 13 (paragraph 1) of the Directive [44] states: "The use of automated calling systems without human intervention (automatic calling machines), facsimile machines (fax) or electronic mail for the purposes of direct marketing may only be allowed in respect of subscribers who have given their prior consent." Further paragraph 4 of article 13 mentions: "In any event, the practice of sending electronic mail for purposes of direct marketing disguising or concealing the identity of the sender on whose behalf the communication is made, or without a valid address to which the recipient may send a request that such communications cease, shall be prohibited."

The EU Directives apply to all unsolicited commercial communications received on and sent from networks in the EU. However the enforcement and identification of spammers becomes quite complicated when emails are generated in third countries.

**US CAN-SPAM Act of 2003** is another anti-spam legislation. The acronym CAN-SPAM is derived from the bill's full name: ***Controlling the Assault of Non-Solicited***

***Pornography and Marketing Act.*** It is the United States' first national standard for the distribution of commercial e-mail and needs the Federal Trade Commission (FTC) to impose its provision [46]. It allows the senders to send unsolicited commercial emails (UCE) until the recipient refuses receipt. It places several other restrictions. It prohibits: deceptive subject lines, harvest email addresses on the web, use of illegally captured computers to relay the messages and false or misleading header information. It requires that commercial messages be identified as advertisements and the senders must include their valid physical postal address. The message must contain an opt-out link for the recipient.

The anti-spam activists commonly refer to CAN-SPAM Act as the YOU-CAN-SPAM Act (means that with this law you can spam) because the opt-out rule in the bill does not require e-mailers to get permission before they send marketing messages [46].

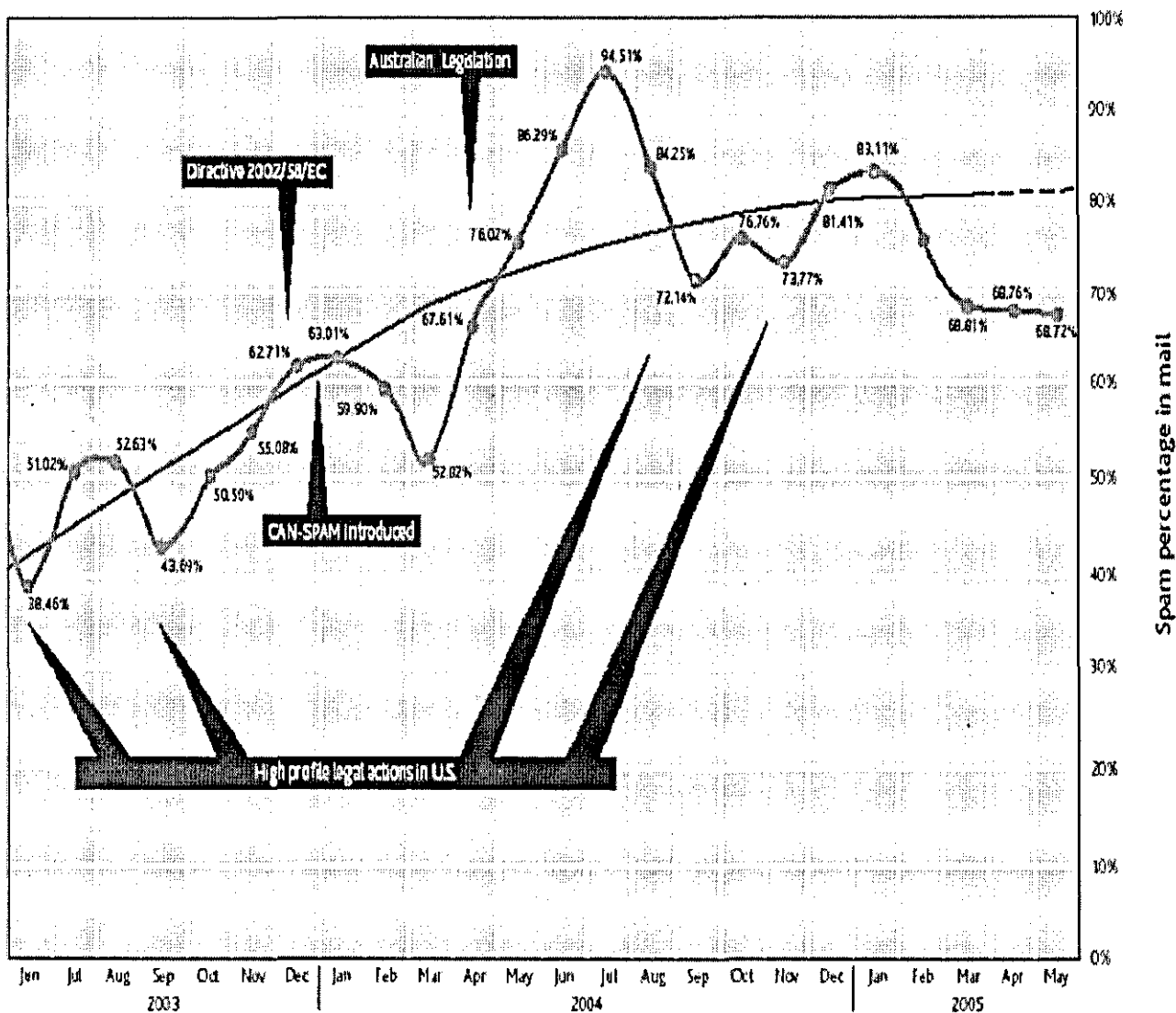
It is worth mentioning that anti-spam laws have not used the term "spam" because its legislative semantics have not yet been defined. For more information on this topic, refer to the paper on, "Combating spam through legislation: A comparative analysis of US and European approaches" [38]. The overview of anti-spam legislation of different countries can be seen in the surveys carried out by the International Telecommunication Union (ITU) [21] and the Organisation for Economic Co-operation and Development (OECD) [47]. The studies of ITU and OECD reveal that many countries have no or non-effective anti-spam legislation. There is no legislation information for large parts of the world like Africa, Latin America, Large parts of Asia and the Middle East.

The legislative approach has achieved very little success in the war against spam and fig 7 shows, that the compliance with the CAN-SPAM act was low from the very beginning. It became even lower in the following years. The figure clearly shows the increase in the growth rate of spam. Apart from Sybil attacks (where the spammers change their identities), the spammers use address spoofing and zombies (controlled PCs of innocent users) which make it very difficult to trace them. In addition, the lack of a widely agreed and well accepted spam definition makes it impossible to have a homogenisation of

worldwide anti-spam legislations. Also many countries still do not have anti-spam legislations which complicates the enforcement and identification. All of these problems contribute to creating difficulties in the regulation of spam and the implementation of the anti-spam laws.

There is a broad consensus that spam must be fought with a combination of anti-spam measures including legislation, collaboration, technological approaches and social awareness. As discussed by [12] the legislative approach can be fully exploited only if we have global harmonisation of anti-spam legislations, international cooperation for sharing information and cross-border investigations and prosecutions related to spam and proper funding and training of the anti-spam organisations and bodies (both technical and legislative).

The following figure shows the percentage of world wide internet email identified as spam from 2003 to 2005.



**Fig 6, percentage of world wide internet email identified as spam [21]**

The resultant line shows that spam increased from 40% in 2003 to 80% in 2005. It means that in 2005 the total number of spam emails was 4 times that of ham emails.

### 3.3 Social Awareness

An approach to the spam problem from a different angle is by social awareness. The efforts which come in this type of approach include educating the users about the causes

and problems of spam and the procedures that can be implemented to reduce it. Warning the users of phishing attacks also comes under this category. The idea is that when people receive emails and messages from banks, employers, organisations and/or universities they do not fall a prey to phishing as they know how to deal it in each case as received. The course on “Spam and Spyware” [48], [49] offered at the university of Calgary, Alberta, Canada is a good example of this category of anti-spam approaches. The course is offered in the computer science department and is available at both the undergraduate and graduate levels. The course looks at examining the problems caused by spam, the different anti-spam approaches and their performance. Students are also given hands-on-experience on developing software for spamming, spyware and how to prevent them. All these practical experiments are carried out in a secure lab.

### **3.4 Collaborative Approaches**

To achieve better spam filtering this approach uses the collaboration of users, companies, organisations and authorities. They share their knowledge and exchange opinions about spam and ham. (See for example [50], [51]). Reports are also gathered from users on a mail server (such as Google’s Gmail<sup>1</sup>). Abuse systems (systems which are meant to help the Internet users to report and control spam and other network abuse) are good examples of collaborative approaches. However abuse systems are subjected to the same drawbacks as shown by reputation systems (discussed in section 3.5.6). The Honey Pot project [52] is one approach which aims to identify email address harvested with the help of specially generated emails. However, this approach can not prevent all and has an issue of privacy that arises due to the exchange of data between users. At the same time this approach needs collaboration from the users who are not always willing to participate. In addition it is also subjected to false praise. The Self-regulatory Plan on Tackling Spam, SpotSpam project [53] recently launched by the EU is an important step towards achieving the goals of Abuse systems that is to help users submit spam complaints to the SpotSpam database at the international level. One other popular example of the international cooperation in the war against spam is the Memorandum of Understanding (MoU) [54] between the UK,

---

<sup>1</sup> <http://gmail.google.com/>



USA and Australia. Another example is the London Action Plan (LAP) [55] (formed in October 2004) in which member organisations from 27 different countries met to discuss the international cooperation against spam.

## 3.5 Anti-spam Technological approaches

Many techniques have been proposed and implemented to prevent spam. However, as mentioned in [56], “None of the schemes is the “magic bullet” that some proponents claim”. In this section we discuss some popular anti-spam technologies.

### 3.5.1 Payments

Research shows that one of the main reasons for spam is the fact that the cost of sending spam is almost zero for a spammer. It has been proposed (e.g. in [57] and [59]) to charge the sender of the email, small payment for sending an email. The payment is kept so small that it remains negligible for a legitimate user but potentially high for the spammers. This would prevent a spammer from broadcasting millions of messages. However this is a difficult trade off and the implementation of such a payment infrastructure can be an ambitious endeavour [40].

The Zmail protocol [58] is one of the versions of the payments’ approach. In this approach the sender pays a small fee to the receiver. The idea is to give neither damage nor loss to a common email user who sends and receives nearly equal amount of messages. At the same time it aims at making spamming costly for the spammers. Another version of the payment mechanism has been discussed in the paper on “Bankable Postage for Network Services” [59]. In this mechanism the sender is charged a small amount of money (in the form of a ticket) but if the receiver decides that the call/email is not spam then the sender is paid back the money (or ticket). It is worth mentioning that the proposers of this mechanism concluded in [58] that, *“There remain numerous questions about the ticket server. As discussed earlier, deployment of the service raises several tricky issues. Many of the same issues arise with other schemes for email*

*payment: in all cases, deployment is difficult, involving fundamental and disruptive changes to the way that Internet email works. It's not at all clear that we can achieve such changes”.*

There is a lot of risk involved in this mechanism especially in the presence of address spoofing and zombies. This is also subjected to false praise. The overview of “The Penny Black Project” [60] from Microsoft Research reveals that the payment can be in the form of several currencies: CPU cycles, memory cycles, turing test etc. The approach of Microsoft Research is fundamentally an economic one. The idea behind this approach is that spammers would have to invest heavily in hardware in order to send high volumes of spam.

The payments approach has many concerns and can not be implemented. The email and VoIP users are unwilling to pay for anything extra for emails and VoIP. Also these payment mechanisms are subjected to false praise and create a significant problem in delivering legitimate/solicited bulk emails. The mechanism introduced by the “Penny Black Project” can cause loss to legitimate users in the presence of bots and zombies. The payment approach is against the open flavour of Internet and is expected to punish legitimate users for the misbehaviours of others. Brad Templeton [61] has discussed the limitations, drawbacks and challenges involved with such payment systems.

### **3.5.2 White and Black Lists**

In white lists a user explicitly states which persons are allowed to contact him/her. Each user has two types of lists. Persons in the white list are trusted persons and are given the connection. However addresses in the black lists are for spammers and they cannot be given the SMTP connection. These lists can be maintained locally or can be provided publicly. There are different variants of white and black lists including DNS-based (Domain Name System-based) white and black lists, Uniform resource Identifier (URI) based white and black lists. Skype also uses a similar technique.

[62] has discussed the DNS-Based Blacklists and its limitations. The main disadvantage of this system is that it is very difficult to up-to-date the white and black lists. The spammers change their identities and/or addresses (using Sybil attacks) which lead to false negatives. Blacklisting the spoofed address of a legitimate user (which is abused by a spammer) can block the ham emails of customers (false positives) which can be of great loss to the customers. On the other hand the problem with the white list is that it is not possible to receive messages by someone who has not yet been put on the white list explicitly (Introduction Problem).

### 3.5.3 Grey Listing

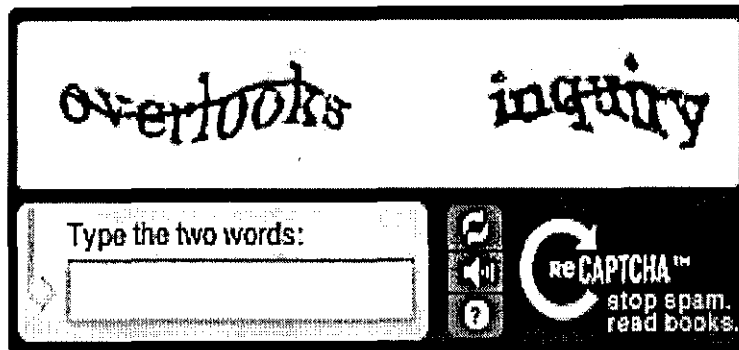
Greylisting as discussed in detail by Evan Harris [63] is based on the assumption that spammers do not always retry sending their messages and if a spammer retries, he would be listed in the blacklist which would prevent him from sending email in the future. For every SMTP transaction, this technique looks for three pieces of information called a “triplet” which is stored in the receiving MTA (Mail Transfer Agent). The three pieces of information are: The IP address of the server attempting the delivery, the email address of the sender (arguments of the MAIL FROM command) and the email address of the recipient (arguments of the RCPT command). According to this rule if the triplet has been never seen before (that is if this message is neither in the whitelist nor in the blacklist) then respond to the server by temporarily refusing delivery. Greylisting is based on the assumption that spammers mostly use applications which ignore error messages and they don’t retry. [63] “Mail server implementations that conform to the SMTP specifications will wait for a while, and then attempt to resend the message” [7]. Based on the receipt of a retransmitted message the grey listing email server will conclude that the email is legitimate.

Spammers have developed new software packages that retry delivery to other MX (Mail Exchanger) hosts for a domain if delivery through one MX fails. Also it’s not effective unless ALL of the MX hosts for a particular domain use mail software that incorporates it. In addition to this grey listing can cause delay in the delivery of a legitimate piece of

mail. "A legitimate piece of mail may get significantly longer than expected if there are enough MX hosts in the mix, even to the point that the sending server may give up and bounce the mail" [63]. Considering the fact as discussed previously that the spam traffic is 4 to 5 times more prevalent than the legitimate email traffic, Greylisting results in an increase of email traffic because most of the emails will be resent. This also leads to the conclusion that this technique does not take into account the network resources' abuse. In addition the users may experience unwanted delays due to congestion on the network. Since the IP addresses of the host changes, Greylisting can result in false positives (where a legitimate email is not received by the recipient) in cases where an email never passes the triplet rule. Also the mechanism fails against address spoofing where a spammer can spoof IP address and the address of the MAIL FROM command (sender's address).

#### **3.5.4 Sender Verification (Challenge/Response)**

The idea behind this mechanism is that spammers send millions of spam emails and are unable to reply to any response sent back to them by the receiving host. The sender is provided with a challenge to prove his/her identity. The challenge is chosen such that it is difficult for computers but easy for humans. These challenges can be simple tests which request the sender's email client to perform a mathematical computation or provide him with an image and ask him to enter the word given in the image. Once the sender responds successfully to the challenge, his message is delivered. Due to the challenge and response characteristics of the verification mechanism, it is also called the challenge-response mechanism. This mechanism distinguishes human senders from auto-generated spams (also called bots) and prevents the spammers from using automated methods to send spams. One such approach is **CAPTCHA** (Completely Automated Public Turing Test to Tell Computers and Humans Apart) algorithms [64]. **CAPTCHA** is a type of challenge-response test, used to ensure that the response is not generated by a computer. The figure below shows an example of a CAPTCHA procedure. The distorted text given in the figure below can be read by a human but it can not be read by current computer programs.



**Fig 7, distorted Text which can be read by humans but not by current computer programs [64]**

In case of VoIP, it has been proposed that the same result can be achieved by introducing noise or music in the audio signal and then asking the user to repeat the voice which will prevent spitters from using speech recognition. These are also called Turing tests.

The problem with this approach is that that the distorted text or the enriched (with noise or music) audio signal should not be too cluttered to be understood by a legitimate user. Especially in the case of VoIP the problem of different accents and different languages can complicate the communication process and it would be very difficult for a caller to understand the noise enriched audio voice. Introduction of noise to the audio signals will add to the inconvenience. At the same time if we make it too easy the intelligent recognition softwares used by spammers could identify the letters and numbers (in the distorted text) or voice (in case of VoIP). Another drawback of this mechanism is that apart from bots it can not prevent other forms of spam (those sent manually by spammers). The mechanism also complicates the sending of solicited bulk messages (like newsletter) which will become impossible or costly in the presence of challenges. This is because each challenge will need a human resource.

In addition spammers bypass this mechanism by social engineering attacks. For example whenever an auto-spam generating computer comes across a challenge (given on some legitimate site), it sends the challenge to a porn viewer on the web pages of a porn

website. On accessing the adult web page the spammers promise the user access to the adult site by responding to the challenge. The moment the user enters the letters and numbers, they are fed into the relevant legitimate sites and the spammers bypass the challenge step. The spammer succeeds in sending spams. In this case the porn site viewer is entering the letters and numbers for the spammers.

### **3.5.5 Cryptographic puzzles**

In cryptographic puzzles [65], before establishing the connection, a caller/sender is asked to solve a small puzzle which consumes computational resources like CPU and bandwidth. [66] (Note: This technique is not a challenge-response mechanism in which the server provides a challenge for each message). This approach is based on the assumption that the computational power of a spammer is limited so the request for the number of parallel connections in a spam attack would weaken the computational power of a spammer in a massive spam attack. However, using zombies (which are virus-infected machines and are used to fulfil the jobs of spammers) spammers can hijack computers and can gain unlimited computational power. As discussed in the “Honeynet Project & Research Alliance” [67], spammers can send (some million) spam e-mails by distributing the total required CPU time required among many hosts. In such a scenario the idea of cryptographic puzzles not only fails but also causes a significant loss of resources to legitimate users. Another drawback of this approach is that a legitimate user with a slow machine can experience unacceptable delays due to the challenges [40].

### **3.5.6 Reputation Systems**

In this system a recipient (organisation) is given a hint about the reputation of the sender/caller before receiving the message (email/call). Based on the reputation of the sender/caller, the recipient accepts (when reputation is good) or rejects (when reputation is poor). This system is susceptible to false praise which can lead to both false positives (classifying legitimate messages as spam) and false negatives (classifying spam as legitimate messages). In addition the system doesn't provide any protection against

address spoofing which can result in false negatives. The situation can become even more severe when the spammers send thousands of spam messages using the spoofed address of a legitimate user. In such a case the address of the legitimate innocent user would be blacklisted and his/her reputation will be badly destroyed. This will block all his messages which would be a great loss not only for the customer but also for the service provider. The system is ineffective against Sybil attacks (in which a spammer changes his identity) or opens a new account. Reputation systems have been discussed in [68]

### **3.5.7 Modifying Transmission Protocols**

One of the drawbacks of existing email transmission protocols (like SMTP) is that they don't provide a mechanism for checking the identity of the message source. It has been proposed to enhance or even substitute the existing standards of email transmission in order to overcome this drawback [13]. However there are a number of obstacles in this kind of approach. Similarly the problem of fake IP addresses in email messages and finding ways to overcome it by changes in the existing standards has been discussed by Goodman in [69].

### **3.5.8 Content Filtering**

As mentioned in [7], [13] and the study by Siponen and Sucke [70], content filtering is the most popular anti-spam technique. It works on heuristic methods and more than 80 percent of the current antispam solutions available on market today make use of the Bayesian filtering methods [71]. The technique first identifies word sets and a database of words as spam. The filtering mechanism then checks the contents (body), header or both parts of the email against these word sets and data base of words. Based on the result of the comparison, the filtering mechanism can classify the email as either spam or ham. Some filters work on collaborative approaches where different servers share information about spam emails and update their filters.

All the filter-based approaches show false negatives and false positives. The false negative is a nuisance for the users but the later one can be of great loss to a company or organisation. Tuning the parameters of the spam filters to intercept spam and ensuring that legitimate emails are not filtered is a difficult balance. A single potential customer's email (say application of a potential student for admission into a university) that is filtered as spam (false positive) is potentially very costly to the company or organisation (in the case of student to a university). Currently there is no filtering mechanism that can give 100% accurate results (i.e. to find an accurate and precise difference between legitimate and spam messages) and in fact we do not think that such a filtering mechanism is possible due to the lack of a single technical definition of spam (which can be used by filters) and for the reasons given below.

Companies can set harsher parameters in SPAM filters, and also enable so-called 'Bayesian-filtering' that allows SPAM to be identified based on trends and probabilities. This would stop a significant amount of SPAM, but at the same time it would block a significant number of legitimate emails which is not acceptable to the users for the reasons given above. The only option left is to try and strike the right balance for users. That is to set some lineant parameters in spam filters which will stop a good percentage of spam emails but still allowing an annoying amount of SPAM in order to get ensure a certain level of probability that legitimate emails are not filtered.

Spams pretend to be replies or follow-ups to previous enquiries. In addition spammers keep on changing the way their message is written (structure). They do so by using misspellings, punctuation, spaces, images (more recently) and other methods to evade the filters. Thus filters need to continue learning and training which is not totally exhaustive and is very costly as well. For example instead of using the word "Hot" they may write "H0t" (note the number 0 instead of the second letter o) to deceive the filters. More recently spammers use images which make it very difficult for the filters to achieve their goals. The existing state-of-the-art algorithms can be used for image-based filtering only if the specific features are extracted from the images. Extracting these features is very difficult for the state-of-the-art filters. Aradhya et al. [72] and Wu et al. [73] have



proposed some approaches that could be used to extract features from images in image-based filtering but they are not exhaustive.

It is important to mention that all of these anti-spam mechanisms can be part of a comprehensive approach to fighting spam. None of these mechanisms can detect and block sufficient spam to be deemed a solution to the spam problem themselves [56].

### ***Problems caused by lack of a single technical definition of spam***

Christopher Lueg [32] has discussed the problems related to the lack of a single technical definition of spam. In many cases, opinions about a spam message may be divided. We take the example of political messages which may be 'unsolicited' for some people but many might be willing to receive them. Similarly jokes may be unsolicited for some but many simply like to receive them. "So far there is no precise technical definition of (email) spam that could be used by mail servers to verify "solicitation" exists. In fact we do not think that such a definition is possible" [32]. We consider the example of the following email given by Lueg [32] to illustrate the point:

"Date: Sat, 29 Mar 2003 09:15:07 -0800  
From: Kate <beatwar@mega781.com>  
To: [author's email address]  
Subject: Should America be at War?  
Message-ID: <155195208-1463747838-1048958107@b.Mega78.com>

*With the current situation in the Middle East, we're conducting a United States Consumer survey about the WAR WITH IRAQ and we'd like to invite you to help us out. Let your opinions be heard amongst Americans and the world! After answering the question above, you qualify to receive a complimentary\* USA T-shirt. Your opinion counts!  
[...]"*

In the context of the war in Iraq it is more than likely that quite a few concerned people would be willing to look at the email as it fits their emotional situation. On the other hand many people would consider it as unsolicited.

“There is a need to define what spam is, as the frequently used term “spam” is not a legal term, which may involve some misunderstandings about the real object of the discipline” [38].

Due to the lack of a precise technical definition of spam, all the content filtering algorithms have the drawback of finding an accurate and precise difference between legitimate and spam messages. They may result in false negatives (classifying a spam message as legitimate). At the same time they may result in false positives (considering a legitimate message as spam and thus blocking or preventing it) which will annoy the user. The later case can be of great loss to companies, organizations or customers. Given below are two examples which illustrate how content-based filtering methods may prevent delivery of legitimate content or email:

- i. A friend tells another friend about a really great deal on a product that they found. Based on the words in the email, the filters may categorise it as spam (advertisement email) and will block it [74].
- ii. A similar incident happened when members of the British parliament did not receive messages relating to the "Sexual Offences Bill" under discussion (Heise Online News, 2003). Assumed to be porn, these messages had been filtered by anti-spam filters [74].

A number of solutions have been proposed (such as [75] and [76]) to overcome the problem of the lack of technical definition and to train the filters. But the fact is revealed by the Taughnack report which states that, **“Although none of the schemes is the “magic bullet” that some proponents claim, some of them, particularly when used in combination with each other, can help limit the amount of spam that users receive”** [56].

It is worth mentioning that content filtering does not take into account the prevention of the precious network and/or Internet resources. As mentioned by Enrico Blanzieri and Anton Bryl [13] that filtering solves the problems caused by spam only *partially* which prevents end-users from wasting their time on junk messages. But it should be noticed that since all the messages are delivered nevertheless, therefore this mechanism does not prevent resource misuse. At the same time these techniques are probabilistic (prevent the delivery of spam with a certain probability). The spammers know that a small percentage of their spam emails would be delivered successfully. This convinces the spammers to send more spam to increase the number of delivered emails. This will lead to large amount of resource consumption. In addition existing content filtering mechanisms do a lot of processing while checking the contents (body) of the message. This uses the network and processor resources and the network gets congested due to the huge amount of spam data.

Many filtering research works claim that the accuracy of their filtering algorithms is above 90%. However it is worth mention that these are the results during the experimental evaluation of these filtering algorithms. These results are based on empirical testing and their data sets are small when compared to the global data set of spam. At the same time their data sets may not be up to date. Due to these reasons we believe that in practice the actual accuracy of these filtering algorithms is less than the results shown by the empirical tests. Also due to these reasons we can not compare the results of different filtering algorithms because different algorithms use different definitions of spams and different data sets for evaluating their performance.

### ***Examples of filtering mechanisms:***

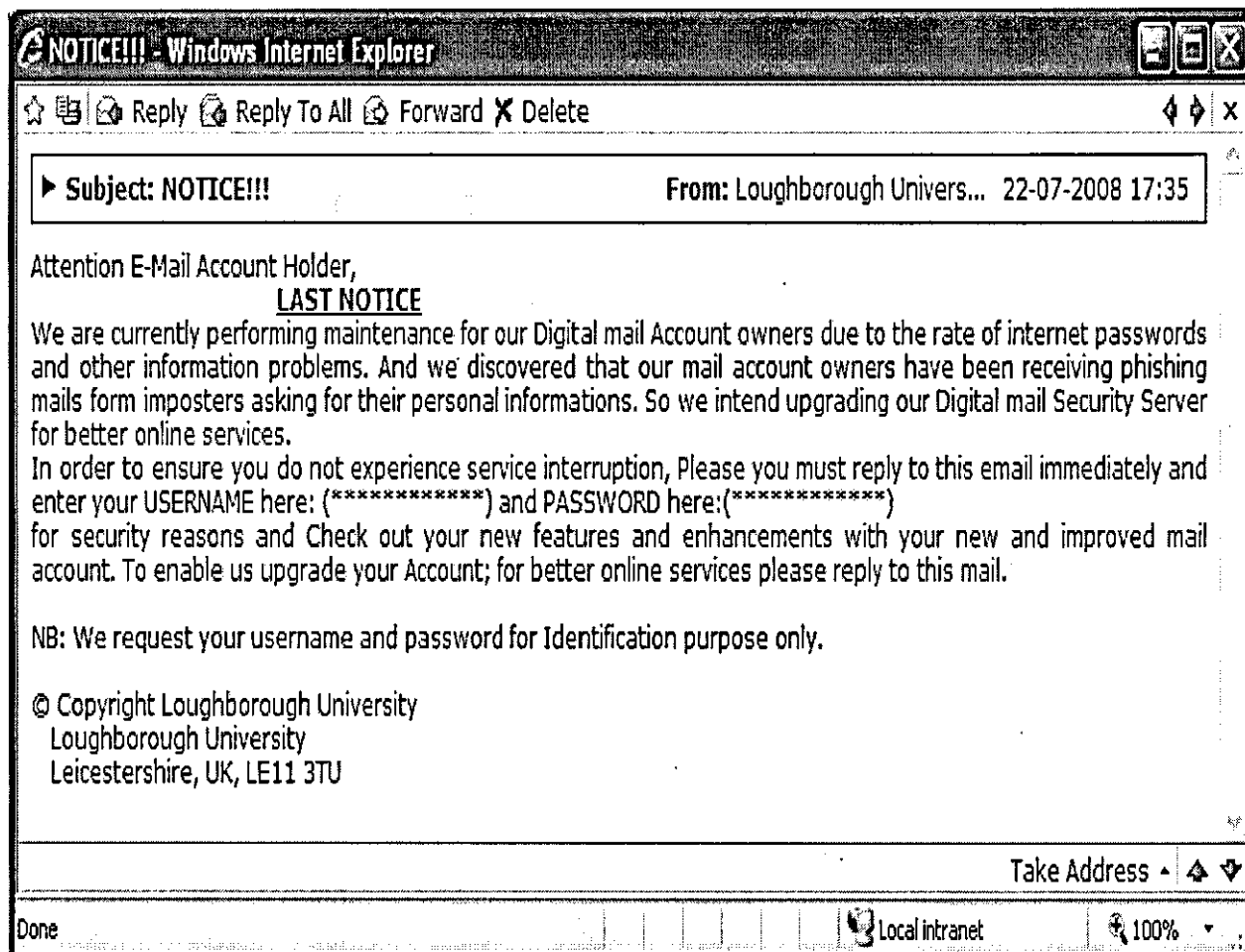
There are many types of filtering mechanisms. Some popular filtering mechanisms have been discussed in [77], [78], [79] and [80]. In this section we discuss two popular spam filtering techniques: Rule-based filtering and Signature-based filtering.

### ***Rule-based filtering:***

In such a filtering mechanism the contents of the email are checked against the database of words. A simple rule can be: If the subject contains “HOT GIRLS” and the body of the email contains “Find a friend” then the rule can decide that it is a spam. However, this filtering mechanism can be easily overcome by spammers with an effort as little as misspelling or using words which can classify it as a ham email. For example, replacing the letter “O” in “HOT” with a number zero “0”. That is “H0T Girls” instead of “HOT GIRLS”.

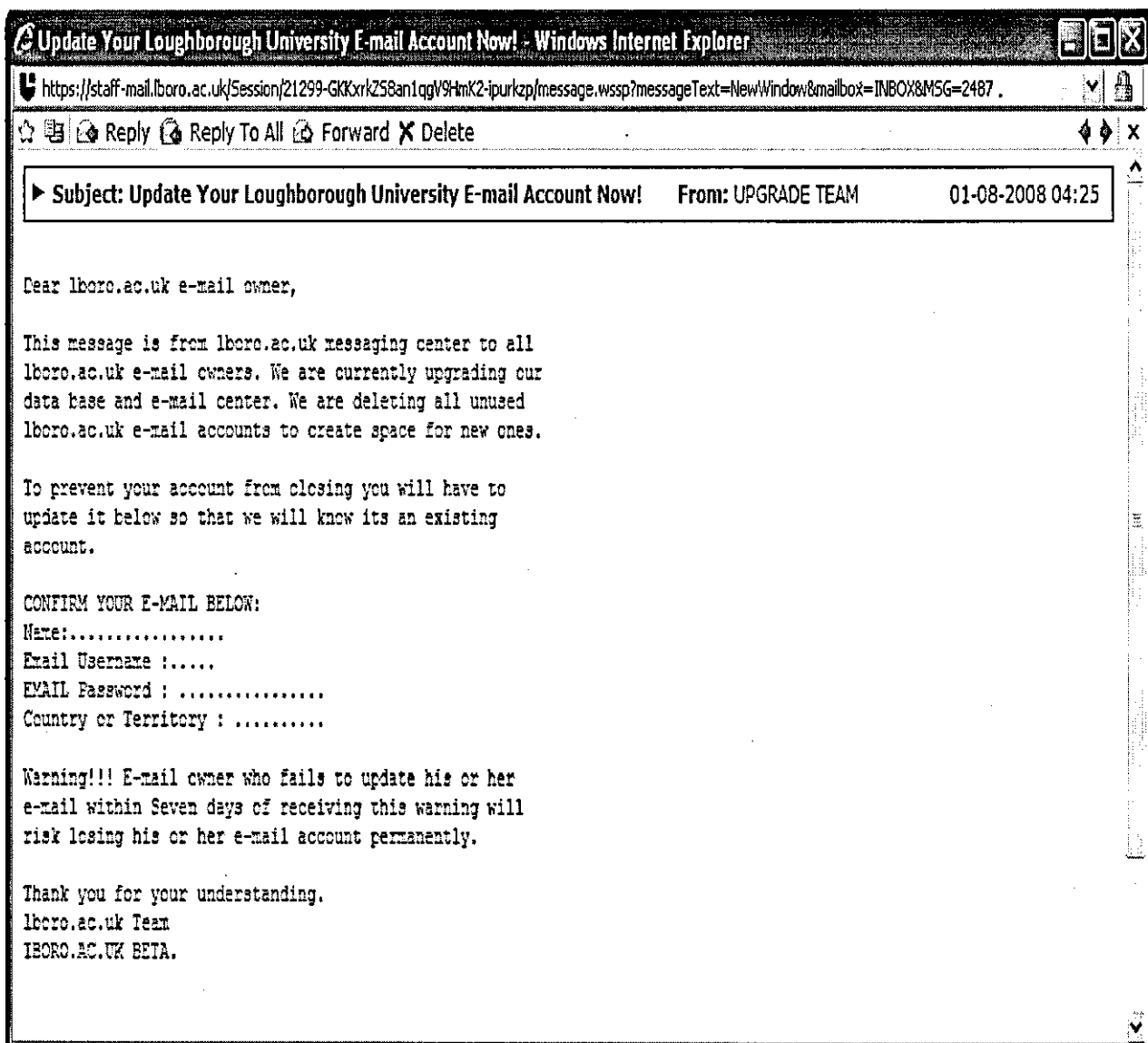
***Signature-based filtering:***

Using an algorithm such as a hash function, this method first calculates the signature of the whole message. Instead of comparing the whole contents, it then compares this signature against a database of known spam signatures. The main disadvantage of such an approach is that with slight variations in the contents of the message the resulting signature is different and it is difficult to update the known spam signatures database. Spammers change the contents of the same messages. This is shown in the following figure that the author received.



**Figure 8, example 1 of a phishing email from the author's email inbox**

As a result of this phishing attack, the spammer obtained usernames and passwords of several accounts. Some students also reported this to the IT service. The author understands that the IT service people would have configured their filters to tackle this spam message in the future. However, the spammers are also aware of this. After a few days, the spammer attacked with a second email (shown below), the message was same (in terms of meaning) but the content and structure was different. So, these two emails are same in meaning but will give different signatures. The spammer was successful even in the second attempt.



**Fig 9, spammer of email in fig 5, changed contents and structure of email to deceive the filters**

Figure 10 shows an overview and classification of the SPIT prevention methods.

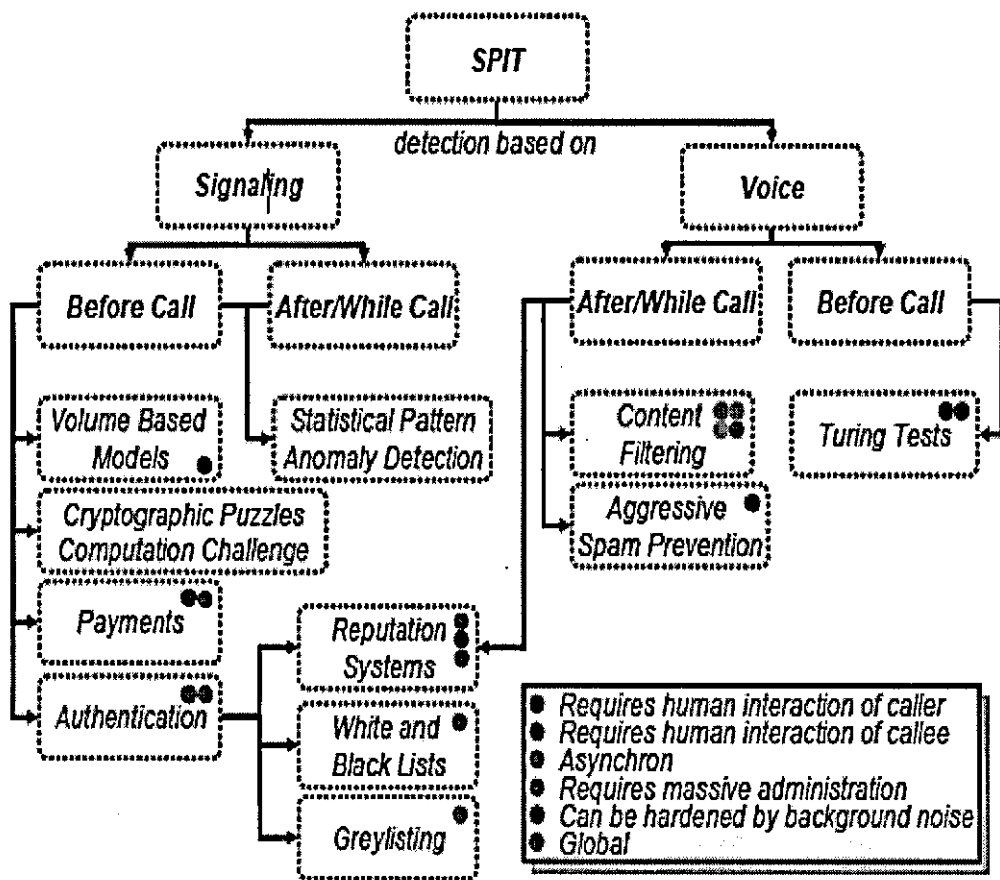


Fig 10, overview and classification of SPIT prevention methods [40]

There are a lot of other techniques which have been used for preventing spam but the fact is that spam is still a problem for users. The facts about the performance of these mechanisms can be identified by the following two statements:

Taughannock Networks mentions in their report on Technical responses to spam [56] *"We believe that technical means can be part of a comprehensive approach to fighting spam, but we don't believe that any technical approaches, individually or in combination, can detect and block enough spam to be deemed a solution to the spam problem."*

Rainer Baumann and his colleagues mention in their report on “Voice over IP – security and SPAM” [40], **“There is no panacea for the spam problem”**.

### **3.6 Summary**

A number of measures have been proposed and implemented to solve the problem of spam. These include legislative, collaborative, social awareness and technological approaches. Each measure has certain drawbacks and limitations. Content filtering is the most popular anti-spam technological approach but all the filter-based approaches show false negatives and false positives. False positives are potentially very costly to users, service providers, companies and organisations. None of the approaches individually or in combination with other approaches, can detect sufficient spam to be deemed a solution to the spam problem. All of the anti-spam mechanisms can be part of a comprehensive approach to fighting spam. The limitations of the state-of-the-art mechanisms create a high demand for an efficient anti-spam mechanism.



# 4 *Spam Prevention using Access Code, SPAC*

---

## **4.1 Introduction to the SPAC mechanism**

Based on the current state-of-the-art there exists no antispam technique which can prevent sufficient spam to be deemed a solution for the spam problem. At the same time these mechanisms have some negative impacts on the distribution of the legitimate messages. The severity of the spam problem and the limitations of the state-of-the-art create a strong desire for a spam prevention mechanism with the following features:

- Prevents the different types of spam attacks (including bots, zombies, spam using address spoofing, sybil attacks, dictionary attacks)
- Allows messages (email/calls) of legitimate clients with no false positives
- Doesn't cause inconvenience to a legitimate user
- Must prevent resource misuse and hence avoid adding to the carbon foot print of IT
- As opposed to state-of-the-art, it must prevent spam messages of all forms (text messages, real-time voice calls etc)

In order to achieve these objectives, we have proposed a mechanism which uses a concept of an Access Code. The proposed spam prevention mechanism has been named "Spam prevention using Access Code", SPAC. This combines the idea of an Access Code mechanism with some of the existing approaches such as white and black lists and

CAPTCHA [64]. However, we use these existing approaches in a different way which overcomes their existing drawbacks and makes them more effective) and enables SPAC to give far better results than the existing approaches. This is achieved by accessing the AC code which combines some of the existing approaches with improvements and/or modifications. SPAC prevents spam by using an Access Code (AC) which is required for sending a message (email or call) to a recipient. The AC can be easily accessed by legitimate clients but it is impossible or so unpleasant for spammers to access it that they give up. SPAC targets spam from two angles i.e. to prevent/block spam and to discourage spammers by making the infrastructure environment very unpleasant for them. Stas Bekman [81] reveals that the idea of discouraging the spammers was first introduced by Ken Simpson and Will Whittaker who founded MailChannels to solve the problem of spam. In addition the spammers will have to pay a cost in terms of time, by gaining information about the recipient and/or provide token as discussed later in this chapter. The AC code can easily be changed by the users. Changing the AC code will not affect the legitimate clients but will create problems for the spammers.

Here we discuss the two main elements of our system.

#### **4.1.1 Identity (ID)**

ID is actually the unique identity or username of a user e.g. email address, VoIP phone number, telephone number or mobile phone number etc. (For convenience, here we will take ID for a VoIP phone number). For example the ID of Prof. Parish in the Department of Electronic and Electrical Engineering at Loughborough University is 01509-123456. In the case of email his ID can be abc@lboro.ac.uk. These are open to both spammers and to legitimate clients. Anyone can search for them on the website of the University. This ID is unique and it can not be changed.

#### **4.1.2 Access Code (AC)**

This is an 'n' number digit which is changeable and is not open to spammers. In our experiments we used 5 digits AC codes. The AC codes are selected by the users and

maintained by a server. The AC code can be changed by the user at any time. However, to do so, the user needs his current AC code (relevant to his ID).

If a sender is totally new to the recipient then he can easily access the AC from a trust worthy server by passing through a mechanism discussed in section 4.5. Apart from this, a sender/caller can also obtain the AC code directly from the recipient (i.e. the recipient can give his/her AC along with the ID when he/she wants to give an ID to the user or through alternative means e.g. if the AC for email is needed then it can be asked on the phone from the recipient and vice versa. It can also be accessed from friends of friends. Multiple users can have the same AC as the ID is unique. In order to prevent spam troll (dictionary attacks) for accessing the AC code, the server must have a lock out mechanism. It means if a user tries 3 false AC codes, the server should lock out him for a certain period of time which will temporarily block the caller's access.

## 4.2 Spammer Vs Legitimate Client

From a detailed analysis of different papers and reports, we have concluded the following facts about a spammer and a legitimate client.

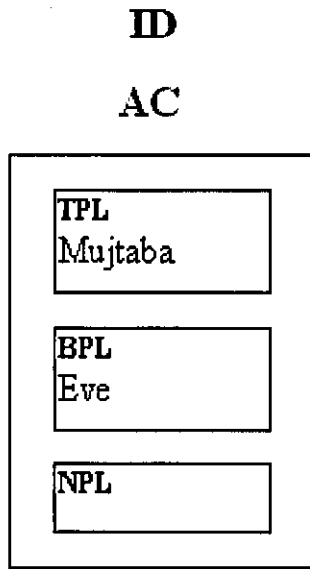
- i. A legitimate client has some knowledge about the recipient whereas a spammer doesn't have such knowledge.
- ii. Sybil attacks, address spoofing and dictionary attacks are the major weapons of spammers to evade state-of-the-art anti-spam mechanisms
- iii. Spammers send spams because the transmission cost of sending spam is almost zero
- iv. It is typically impossible to call a spammer back
- v. Spams are sent to thousands and millions of users within a very short time. Spamming can only profit a spammer if the spammer can send *large number of emails* within a *short time*

- vi. The MailChannels team observed that spammers are impatient. They abort the connection and move on to spam other servers if they can not deliver a message within several seconds [81].

SPAC uses these differences between a spammer and a legitimate client to filter out spammers from the legitimate clients. The AC can be obtained at the cost of information about the recipient and tokens (tokens are required by spammers only and are discussed later in the section on charging mechanism). The mechanism provides the caller/sender with a challenge mechanism where information about the recipient is asked. Information can not be provided by the spammer. The Token mechanism (discussed in section 4.8) reduces the overall attempts of the spammer whereas the lockout mechanism limits the spammer to 3 false attempts per recipient in a particular duration of time. The SPAC server also contacts back the caller for entering the access code. This also reduces spam because in many cases, the spammers can not be contacted back. Introduction of Challenge mechanisms and the CAPTCHA procedures prevent bots or auto-generated spam which defeats the spammer on the time factor. The mechanism as a whole defeats the spammer in all the factors as mentioned above. We will discuss these in detail in the coming sections.

### **4.3 Data Base of a User on the Server**

Fig 11 shows the data base of a user (say Prof. Parish) on a server A.



**Fig 11, data base of Prof. Parish (a VoIP user) on the server, A**

Let us briefly discuss the various functions in Fig 12.

#### **4.3.1 Trusted Persons List (TPL)**

This list contains ID's of those persons who have been approved by the recipient. These users don't need the AC code of the recipient. They simply need the recipient's ID. A user can add persons manually to his TPL list in which case the added users will not need the AC at all or a user (caller/sender) can be added to the TPL list after he/she accesses the AC code and make a first connection to the recipient.

#### **4.3.2 Blocked Persons List (BPL)**

This list maintains addresses of spammers. A user in this list can never contact the recipient who has blocked him/her by using the blocked address/identity. Even if he/she gets the correct ID and AC they can not communicate.

The TPL and BPL lists are modified forms of white and black lists. The state-of-the-art white and black lists maintain IPs, URLs or DNS names of the legitimate users and the

spammers respectively and they are treated globally. If we assume that XYZ is a URL, IP or DNS name which is in the white list then it will be legitimate for all the users of the email service provider (ESP) and vice versa. For this reason such white and black lists are subjected to false praise and address spoofing. However, the TPL and BPL lists in our case are relevant to each individual separately. So, a person with ID xyz may be in the TPL list of a user (say ABC) but he may be in the BPL list of another user (say KLM). This feature of treating the TPL list and the BPL list of each person separately makes SPAC effective against false praise and address spoofing as discussed in section 4.6.

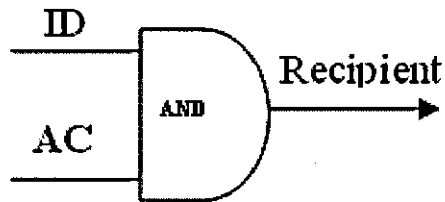
### **4.3.3 New Persons List (NPL)**

Introduction of legitimate strangers (new legitimate clients) is really a case of concern for most state-of-the-art techniques (like white and black lists) which has been solved by SPAC. Any user who wants to send a message for the first time to a particular recipient will get an entry in his/her NPL. He will be added (by the receiver) either to the TPL or to the BPL (depending on whether he is a legitimate user or a spammer). The advantage of this list is in those cases where a new legitimate user is unable to get a connection to the recipient. Suppose, in the case of VoIP calls, when the caller successfully obtains the AC, the server puts him in the NPL list of the recipient. If the connection cannot be established (for example if the recipient is busy) then this caller will not be required to repeat the whole process for obtaining the AC. He will just dial the ID. This new comer will remain in the NPL list until and unless the called person has changed his AC code.

## **4.4 Working of the System**

In the SPAC mechanism, if a caller/sender is already been authenticated by a recipient then he/she would need only the ID (identity or username) of the recipient. However, if the client is totally new to the recipient then he/she would need to know AC (Access Code) of the recipient in addition to his/her ID. For the first message, the SPAC mechanism applies an AND operation on the ID and AC (ID and AC have been discussed

in section 4.1) of the recipient as shown in fig 12. This means that in order to establish a connection or send a message the two entities of the recipient must match.



**Fig 12, the basic operation of the SPAM prevention system**

A user can add a person directly to his TPL list in which case the added person doesn't need the AC even for his/her first contact. In the worst case, a legitimate client will need the AC only for the first contact with the recipient. For sending further messages (to the same recipient), the AC code is not needed. This provides a degree of convenience. The pseudo code for the SPAC application is given below:

### ***Notations***

The notations used in the algorithm are listed below:

A: The trusted server on which SPAC has been installed

S: Sender of email or caller (in VoIP)

R: Recipient

AC: Access Code

ID: Identity or username of a user

DB: Database

n: Number of consecutive false attempts (by S) of any one of the followings:

1. Entering AC
2. Answering questions in the challenge phase

m: Number of tokens in the account of S

k: Threshold level of correct answers set by R to provide AC to S

**Module 1**

A listens for connection request(s) from users

If

A receives a request from S to call R

Then

Check R in the DB of A and Go to module 2

Else

Keep listening

**Module 2**

If

R exists in the DB of A

Then

Go to Module 3

Else

Give error message to S

**Module 3**

If

R has disabled SPAC

Then

Establish the connection

Else

Go to module 4



**Module 4**

If

S is in the database of R

Then

Go to Module 5

Else

Go to Module 6

**Module 5**

If

S is in the TPL or NPL of R

Then

Establish the connection

Else

Reject the connection by giving an error message

**Module 6**

Ask S to enter both ID and AC of R

If

AC and ID of R match

Then

Establish the connection

Else

Go to Module 7

**Module 7**

Provide the distorted text developed through CAPTCHA program

If

The code given in the image is entered correctly then go to Module 8

Else

Give another image of distorted text for entering

### Module 8

If

$m > 0$  and  $n < 4$

Then

Provide questions to be answered by S and listen for the user to submit the answers

Else

Give an error message E3<sup>1</sup>

### Module 9

Event: When the user submits the answers

If

Number of correct answers  $> k$

Then

Give AC in the form of distorted text

Else

Change the sequence of the questions and the options, and go to Module 8

## 4.5 Accessing the Access Code (AC) from the server

At the time of first contact if the sender/caller doesn't know the AC code of the recipient, he can access it easily from a trust worthy server by passing through a mechanism which involves the following steps:

- i. While making a contact, the user will first be connected to a server. If the user is a spammer it will be difficult for the server to contact him back
- ii. It asks the user to enter the distorted text given in an image in a given field (e.g. CAPTCHA code). This will filter out bots and will limit the resources of the spammer.

---

<sup>1</sup> E3: If  $m < 1$ , then E3 = You do not have enough tokens in your account  
If  $n > 3$ , You have made 3 false attempts. Please try again later

- iii. It asks the user for specific information about the recipient. Here the spammer will fail. Third party such as porn website viewers (as discussed in section 3.5.4) will also not be able to fulfil this job for the spammer. A sample of such questions is shown in Figure 28.
- iv. The lock out mechanism will put a limit on the number of false attempts made by the spammer. Also the sequence of questions and their multiple choices (answers) as shown in Figure 28 changes on each new attempt which adds to the inconvenience for the spammer.
- v. The multiple choice answers and the final AC code on passing all the steps is also given in the form of distorted texts in the image to make it more immune to bots.
- vi. The spammer will receive charges for each spam (section 4.8 on the charging mechanism)
- vii. Even if a spammer has overcome the above issues then he will not be able to send multiple spams in parallel to a number of recipients because information regarding each person is different and random. So, he can only provide information and AC codes in series. This will not be efficient for him (see facts 3, 5 and 6 in section 4.2). In addition the use of the CAPTCHA program and providing access code in the form of distorted text in an image further prevents the chances of auto-generated spam (bots).

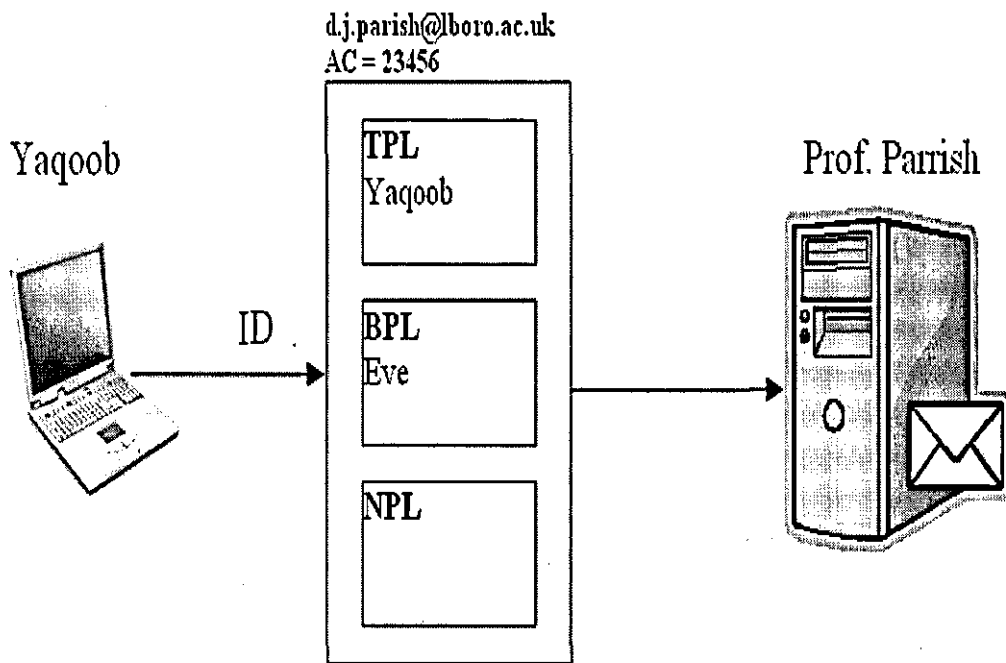
## 4.6 Different Case Studies

Now we consider the response of the system for different cases. We imagine that the recipient (of emails) in our case studies is Prof. David J Parish who is in the Department of Electronic and Electrical Engineering at Loughborough University, UK. For the sake of simplicity we are considering spam in email in these case studies. Similar cases can be used for spam in VoIP (SPIT). The email address of Prof. Parish (say `d.j.parish@lboro.ac.uk`) is represented by the ID in these cases. A user while sending an email to Prof. Parish can come across any one of the following cases:

1. An email can be from a legitimate person who is in regular contact with Prof. Parish
2. The sender of the email can be a legitimate client who is a stranger to Prof. Parish
3. The sender could be a spammer who has no knowledge about the AC of Prof. Parish
4. A spammer with knowledge of the AC code of Prof. Parish who would like to send him an email
5. The sender could be a spammer who is already in the BPL list
6. A spammer might want to send spam emails to Prof. Parish with changed identity (Sybil Attacks)
7. A spammer may want to send an email to Prof. Parish using a spoofed address

#### ***Case 1 – Approved Caller***

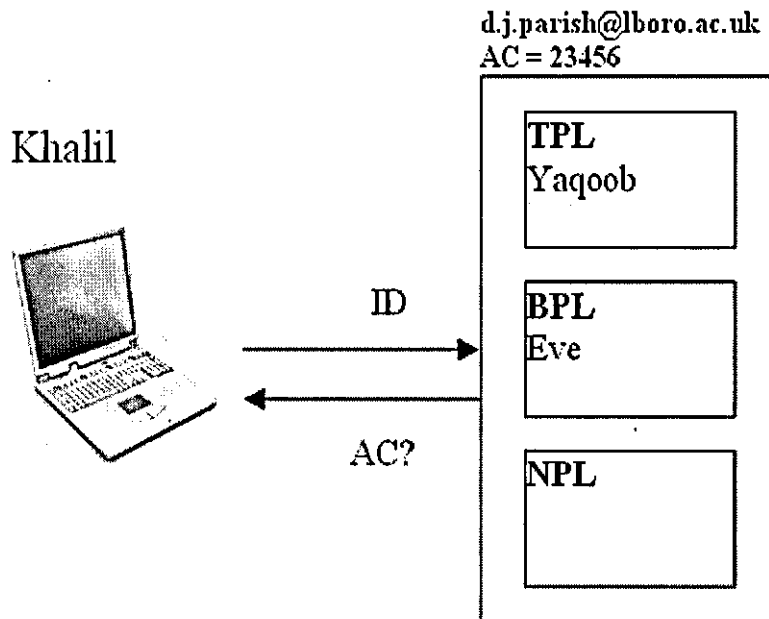
In this case the user doesn't need to pass through the challenge step (collecting information about the recipient). We consider Yaqoob who is a student of Prof. Parish in the High Speed Networks (HSN) research group at Loughborough University, UK. When Yaqoob is given the ID of Prof. Parish, he is also given his AC code. For the first time Yaqoob will use the AC to send an email to the recipient. After that Prof. Parish would add the sender (i.e. Yaqoob) to his TPL list. Prof. Parish can also add the ID of Yaqoob manually (directly) into his TPL list in which case Yaqoob would never need the AC of Prof. Parish. Once Yaqoob is added to the TPL list of Prof. Parish (either manually or by using AC for the first call), he would not need the AC code for any future calls to Prof. Parish because he has been approved by Prof. Parish. The following figure shows that any person in the TPL list (in this case Yaqoob) would need only the ID of the recipient (here Prof. Parish) to contact him.



**Fig 13, A sender in the TPL list requires only email ID**

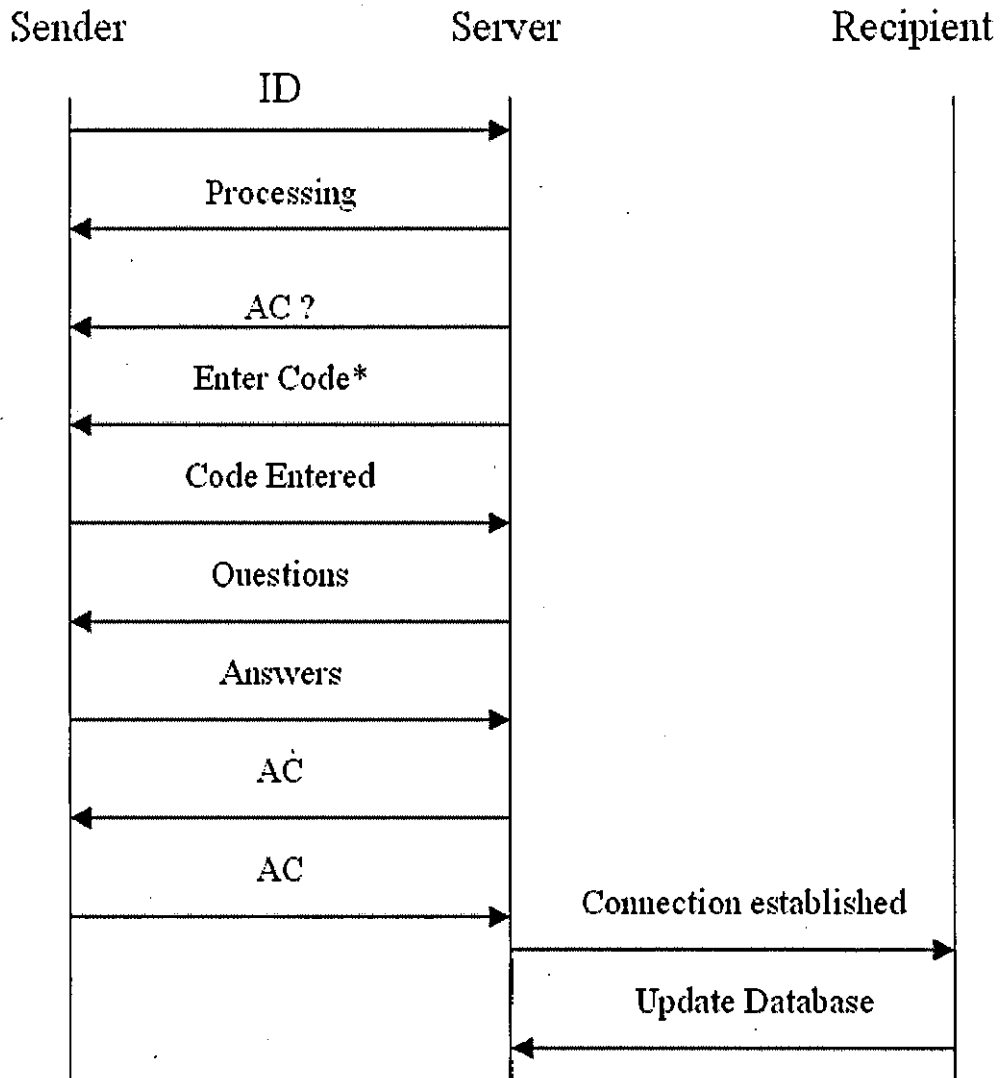
***Case 2 – Legitimate Stranger***

Khalil is a new student and he wants to contact Prof. Parrish. On the university website Khalil searched for the email address (ID) of Prof. Parrish. The figure below shows that Khalil is in none of the lists of Prof. Parrish. This means that he needs the AC code.



**Fig 14, Khalil (a legitimate stranger) needs AC**

In order to obtain the AC from the server, Khalil will need to pass the different steps as mentioned in section 4.5. These steps are shown in Figure 15:



**Fig 15, accessing the AC from the server by a legitimate client**

*\*CAPTCHA code*

Khalil is contacted by the server, the server first checks if Khalil is in any of the three lists. The server finds that Khalil is in none of the lists. Now the server calls Khalil back and asks him to enter the AC code of Prof. Parish but Khalil doesn't know the AC. (A spammer can not be called back as mentioned in section 4.2, fact no iv). The server realizes that Khalil is unknown to Prof. Parish and he wants to send an email to Prof. Parish. The AC will be given to Khalil by passing the following steps:

1. In order to confirm that the request is not a bot (auto-generated spam), the server asks the sender to enter a distorted text (given in an image) in a field as shown in Figure 16. This is achieved by using a CAPTCHA Program [64].

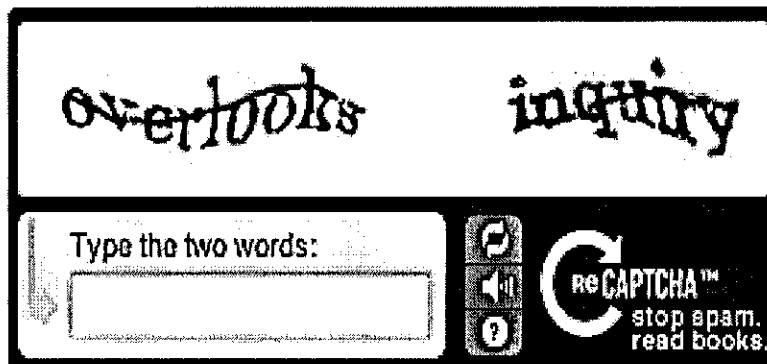


Fig 16, example of CAPTCHA procedures [64]

2. The server will ask Khalil some questions regarding Prof. Parish, such as his full name, department, university etc. If Khalil gives the correct information then he will be provided with the correct AC. Khalil can provide this information because he has found it on the university website.

Note: A spammer would not have such information about Prof. Parish (See Fact no: 1) or it will be costly to obtain.

3. By charging Khalil one token for this service. (Please see the charging mechanism section). Now it's clear that Khalil is not a spammer. If he was a spammer then:
  - It would not be possible for the server to contact him back
  - He would not be able to provide information about Prof. Parish
  - The AC number is going to be lost the moment he sends a SPAM so he would not spend token and time to access it. Spamming could not profit him/her.

When Khalil sends the email, he is listed in the NPL of Prof. Parish. Since Khalil is not a spammer, Prof. Parish will not add him to the BPL list. If the sender were a spammer



then adding him/her to the BPL list would cost the sender 1 token. Prof. Parish can either add him into his TPL list or he can leave him in the NPL list.

- If Khalil is added into his TPL list then the token is returned into his account and for future contacts he would be treated as a sender in Case 1.
- If Khalil is left in the NPL list then until Prof. Parish changes his AC, he will not need the AC code for future emails. But if Prof. Parish changes his AC then Khalil will need to repeat all the steps of Case 2 to obtain the new AC.

### ***Case 3 – Dictionary Attack***

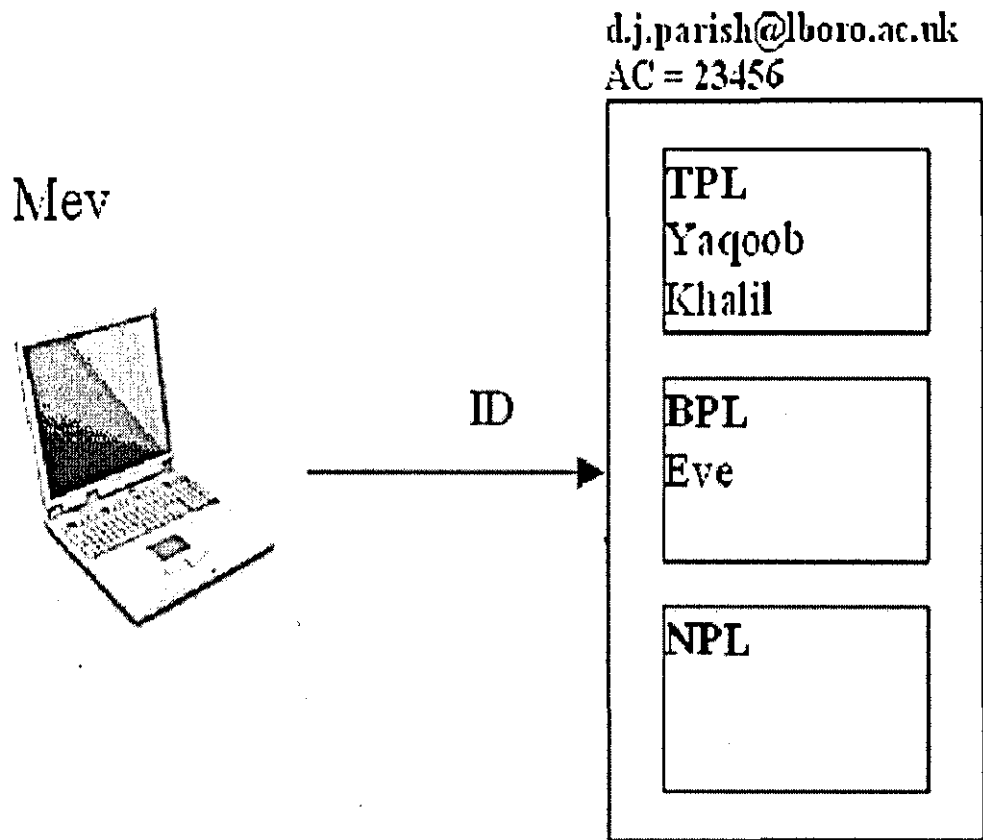
Here the sender is a spammer who wants to send a spam to Prof. Parish by randomly selecting or guessing the email address of a user (the spam troll/dictionary attack). The spammer's address is in none of the lists of Prof. Parish. So, the server will function as for an unknown person (or stranger). The spammer will totally fail because he needs the AC code for which he needs to pass all the steps as discussed in Case 2. This means that our mechanism is totally effective against spam troll/dictionary attacks.

### ***Case 4 – Spammer who accesses the AC***

In this case a spammer (say Mev) accesses the AC of Prof. Parish by fulfilling all the requirements of the server which involves call back, time, information about the recipient and cost.

*Note: It will be very difficult (if not impossible) for the spammer to fulfill all these requirements. At the same time spamming in such a case would not profit the spammer. But we consider a worst case in which a spammer accesses the AC.*

The figure below shows that Mev is in none of the lists of Prof. Parish.



**Fig 17, Mev, the spammer can't be connected to Prof. Parish**

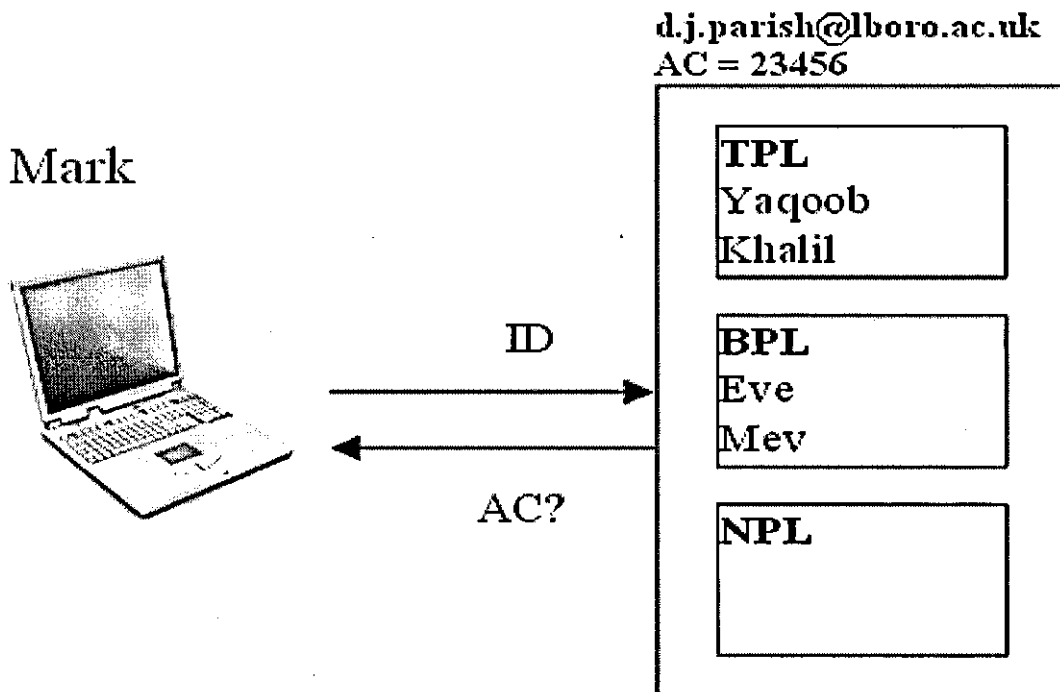
When Mev sends an email to Prof. Parish, she remains in the NPL list of Prof. Parish who is going to put her in his BPL list after concluding that Mev is a spammer (as shown in fig 19). At the same time Prof. Parish will also change his AC. In order to send a spam to Prof. Parish again, the spammer would need to repeat all the steps as discussed in Case 2 and at the cost of one token. This would defeat the spammer on time, tokens, patience and the loss of his spamming business. In fact it would be impossible for the spammer to send emails in bulk. He/she would not be able to get the financial gain that is expected from the spam business. Thus the infrastructure environment has been made so unpleasant for the spammer that he gives up.

**Case 5 – Spammer in the BPL list**

In this case a spammer (say Eve) who is already in the BPL list of Prof. Parish wants to send an email to Prof. Parish. The server finds that he is in the BPL list so it will simply respond to her with a sorry message.

**Case 6 – Sybil Attack**

Here Eve, the spammer wants to change her identity (Sybil Attack). For example instead of eve@domain.com she wants to use the ID of Mark (mark@domain.com) represented by Mark in the figure below.



**Fig 18, In case of Sybil attack the server functions as for unknown person**

In such a case the server would function as for unknown person and she will need to contact server A for the AC of Prof. Parish. The server would function as for case 2.

### Case 7 – Address Spoofing

In the presence of SPAC, address spoofing can not help the spammer because even with address spoofing, the SPAC server functions as for unknown persons (Case 2). Our proposed mechanism provides three levels of resistance against address spoofing. We consider two situations of address spoofing:

- Spoofing the address of a person who is not in the TPL list of the callee
- Spoofing the address of a person who is in the TPL list of the callee

#### a) Spoofed Address not in the TPL List:

In this case, the spitter wants to use the address of a random person (say Mark) who is in none of the lists of Prof. Parish. The server will function as for case 2 as shown in Figure 19.

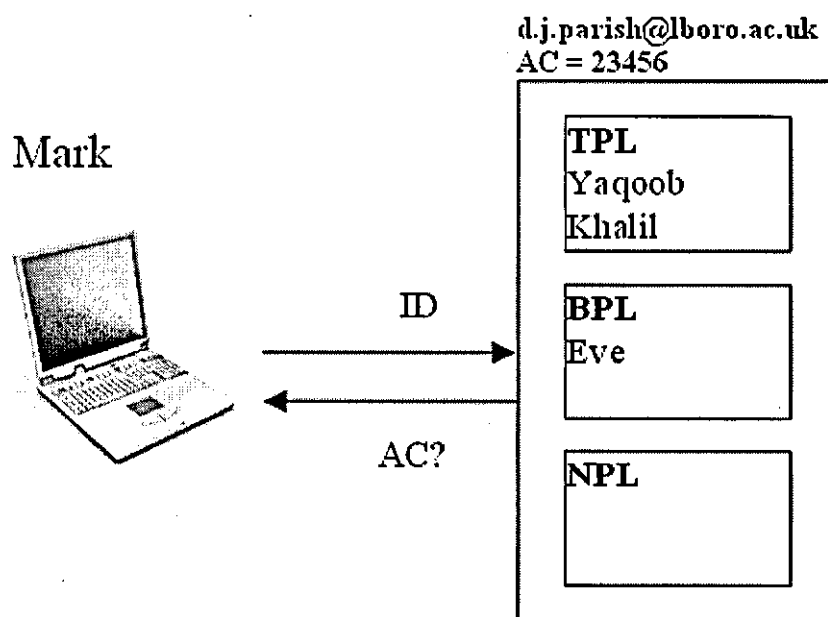


Fig 19, spoofing address of a random person does not help the spammer

The spammer will need to pass all the steps. In the presence of SPAC, this case of address spoofing can not help the spammer.

### b) Spoofed Address in the TPL List:

Spoofing the address of some other person (say XYZ) the spitter must be certain that the person XYZ is in the TPL list of the recipient(s). It is impossible for the spammer to find who is in the TPL list of a certain person unless they hack the server or the host. In addition it is not probable that the same person (XYZ) would be in the TPL list of all the recipients as shown in the figure below (for simplicity we have shown databases of only 3 persons. Spamming can be profitable only if spammers can send spam to a huge number of persons because only a very small percentage of that huge number respond to spam as discussed in Chapter 2. So, in practice we believe that spammers must be certain that the spoofed address is in the TPL list for thousands and millions of recipients).

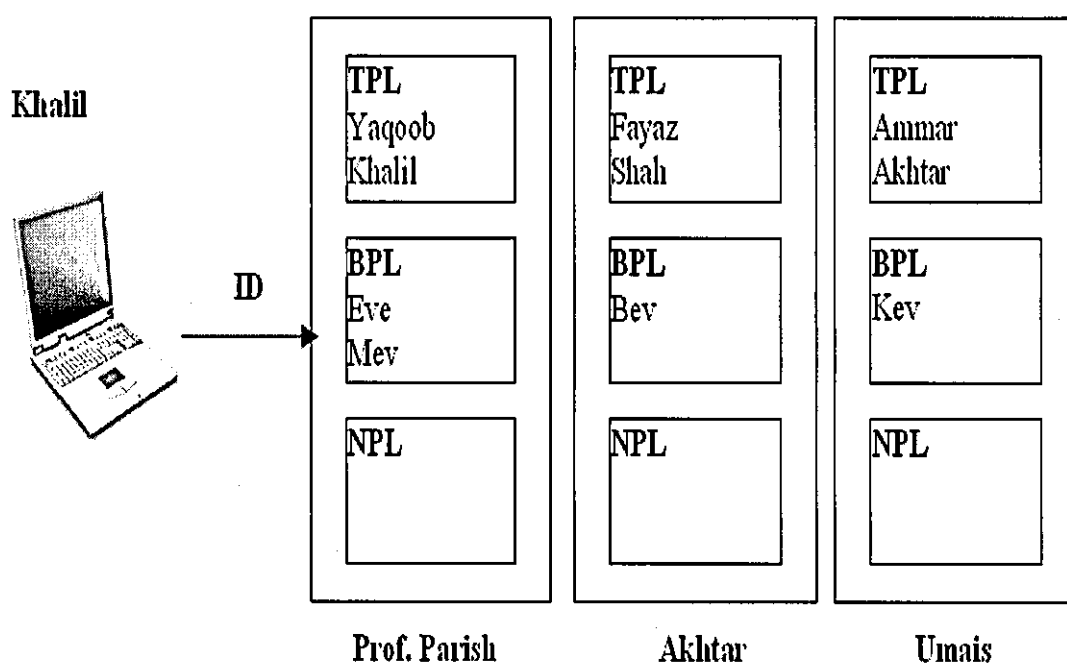


Fig 20, spoofing address of a person in the TPL list

Consider the case where a spammer gets knowledge about the TPL list of a recipient (say Prof. Parish). Spoofing the address of a person who is in the TPL list of the recipient can give little or no help to the spammer. For example, in Figure 20, a spammer spoofs the address of Khalil. But with this address, he would be able to send unwanted email only to

Prof. Parish. He would not be able to send unwanted email to Akhtar or Umair because Khalil is not in the TPL list of Akhtar or Umair. So, we conclude that in the case of address spoofing, the spammer would only be able to send unwanted email to that particular recipient whose TPL list contains the spoofed address (again at the cost of time, patience, tokens and information about the recipient). The spammer would not be able to send spams to different recipients.

SPAC can be further strengthened against address spoofing by using a simple authentication mechanism. With SPAC we can introduce a very simple cryptographic authentication mechanism i.e. by encrypting the ID with the AC. This will help in reducing the chances of spoofing. For example while making a call, the server first authenticates the caller by asking him/her to give his AC. This encrypts his ID with his AC and sends the cyphertext to the server. The server already has the ID and AC of the caller/sender in its database. It encrypts the ID of the caller/sender with his access code and compares the two cyphertexts. After authenticating, the local server will carry out the other steps. However, it should be noted that this authentication mechanism is not a part of this research; rather it is a related area of research. We mentioned it here to show that SPAC provides an easy mechanism for sender/caller authentication.

## **4.7 Charging Mechanism**

One of the main reasons for spam is the fact that the cost of sending spam messages is almost zero. At the same time the key reason for internet popularity is that it is almost free. If we control the cost factor such that it is free for the legitimate clients but not for the spammers then this can help in creating problems for the spammers without creating any difficulty for the legitimate clients. Previously payment methods have been suggested which have been discussed in detail in section 3.5.1. The charging mechanism in our proposed mechanism overcomes the drawbacks/limitations of those proposed mechanisms. In our mechanism an allocation of free tokens per month are provided to every user. One token is subtracted from the sender's/caller's account whenever he/she accesses the server for the AC code of a new recipient. The token is returned to his/her

account when the recipient decides that he/she is not a spammer. In chapter 6 (section 6.2) we will discuss the outcomes of a survey for fixing the number of free tokens for each user. We concluded that 100 tokens per user per month was a reasonable starting point. These tokens are enough for a legitimate caller but not nearly sufficient for a spammer. The reason is that the free tokens are returned to legitimate callers/senders and they don't usually send bulk emails. However, spammers send a large number of spams and they will not be returned the tokens (based on the feedback by the recipient that he/she is a spammer).

Note: The charging mechanism will work only if a global charge structure is introduced.

## **4.8 Summary**

An anti-spam mechanism referred to as “Spam Prevention using Access Code”, SPAC has been proposed to prevent spam. In addition to the ID of the recipient, SPAC requires an Access Code (AC) for sending a message (email or call) to a recipient. SPAC functions equally well for dictionary attacks (spam troll), sybil attacks, address spoofing etc. The AC can be easily accessed by legitimate clients but it is impossible or significantly unpleasant for the spammers to access it. SPAC targets spam from two angles i.e. to prevent/block spam and to discourage spammers by making the infrastructure environment very unpleasant for them. For the first time call from a sender, SPAC treats all these cases as stranger sender/caller and asks for the AC code in addition to the ID which can be obtained by passing through the challenge and charging mechanism of SPAC. Detailed analysis of the results obtained from the tests on the SPAC application (as discussed in chapter 6) and a study of the characteristics of the spammer and a legitimate user show that the challenge and charging mechanisms provided by SPAC are very difficult, unpleasant and costly for the spammer. SPAC provides the user with a degree of convenience because legitimate trusted users will not need the AC code.

# 5 *The SPAC Application*

## 5.1 Introduction

In order to test the SPAC mechanism as discussed in the previous chapter, we have developed a SPAC application. The software application is similar in function to a chat messenger. The application has been developed such that it can be used to perform tests for preventing both types of spams (SPIT and spam in email) under consideration in this thesis. The application was developed using PHP (Hypertext Preprocessor) as our scripting language and MySQL as our database management system. Dreamweaver was used as a web development application. Dreamweaver, PHP and MySQL have been introduced in Appendix A, Appendix B and Appendix C respectively (given at the end of this thesis).

This chapter describes the working of the SPAC application and introduces the different features of the SPAC mechanism (with snapshots) discussed in chapter 4 and explanations of the different functions of the SPAC application. The application has been developed such that it can be used to test the ability of the SPAC mechanism to prevent SPIT and spam emails which are the 2 major types of spams under consideration in this research. The results we obtained from our different tests are given in the following chapter.



## 5.2 Flow Chart

Figure 21 shows the flow chart of the SPAC application.

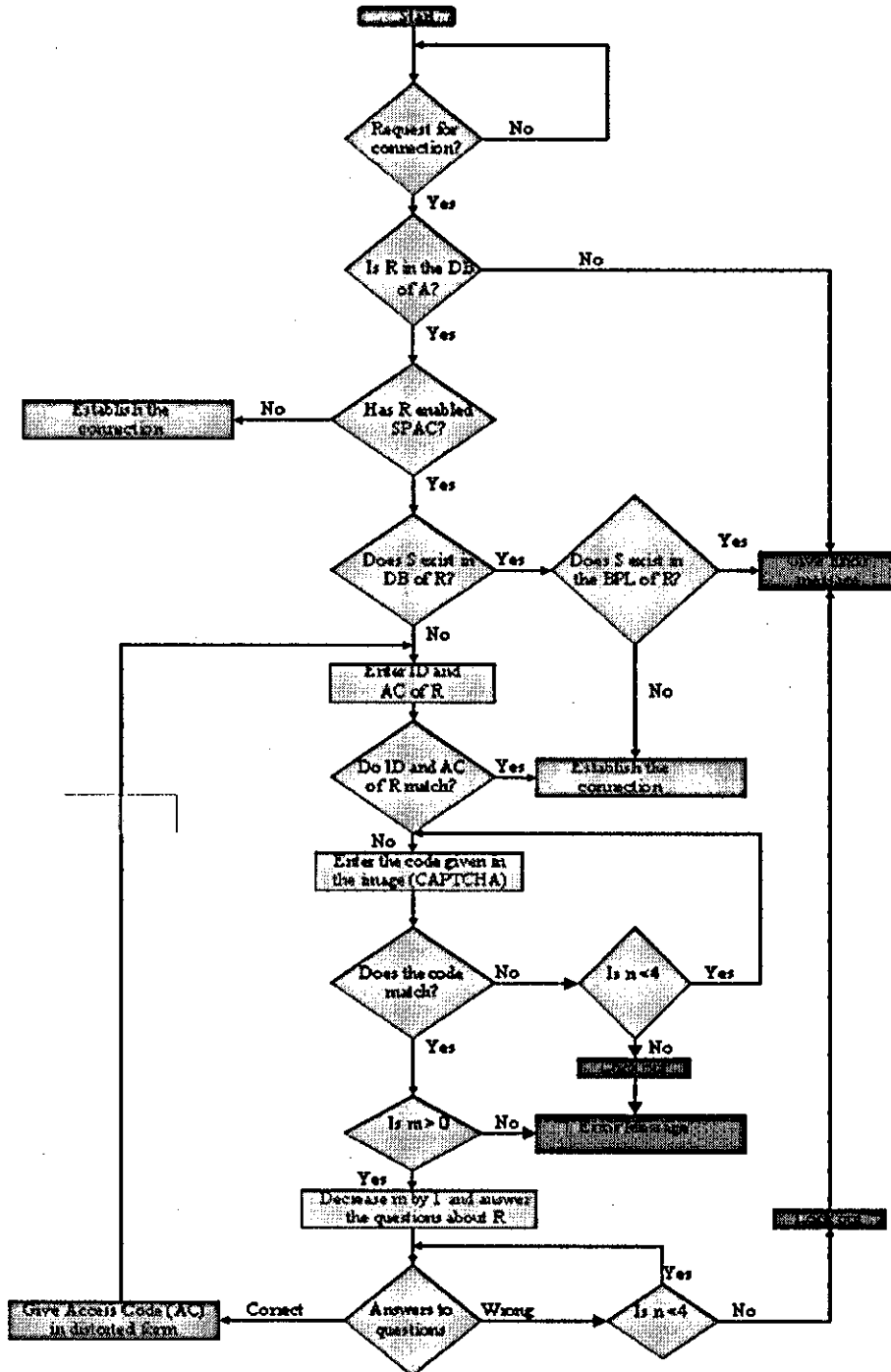


Fig 21, Flow chart of the SPAC application

## 5.3 The SPAC Application

Below we discuss the software application and how it is used.

### 5.3.1 Accessing the SPAC home page:

We consider that the “spac” folder (which contains all the necessary files and database of the SPAC application) is located on a server named “Akhtar” on the local network. The spac server can be accessed from any pc on the local network via a web browser. A web browser is opened and the URL <http://akhtar/spac/index.php> or <http://IPaddressoftheserver/spac/index.php> entered. (where “IPaddressoftheserver” is the IP address of the server on which the spac folder is present). This opens the main home page of the SPAC application which gives the main menu as shown in Figure 22:

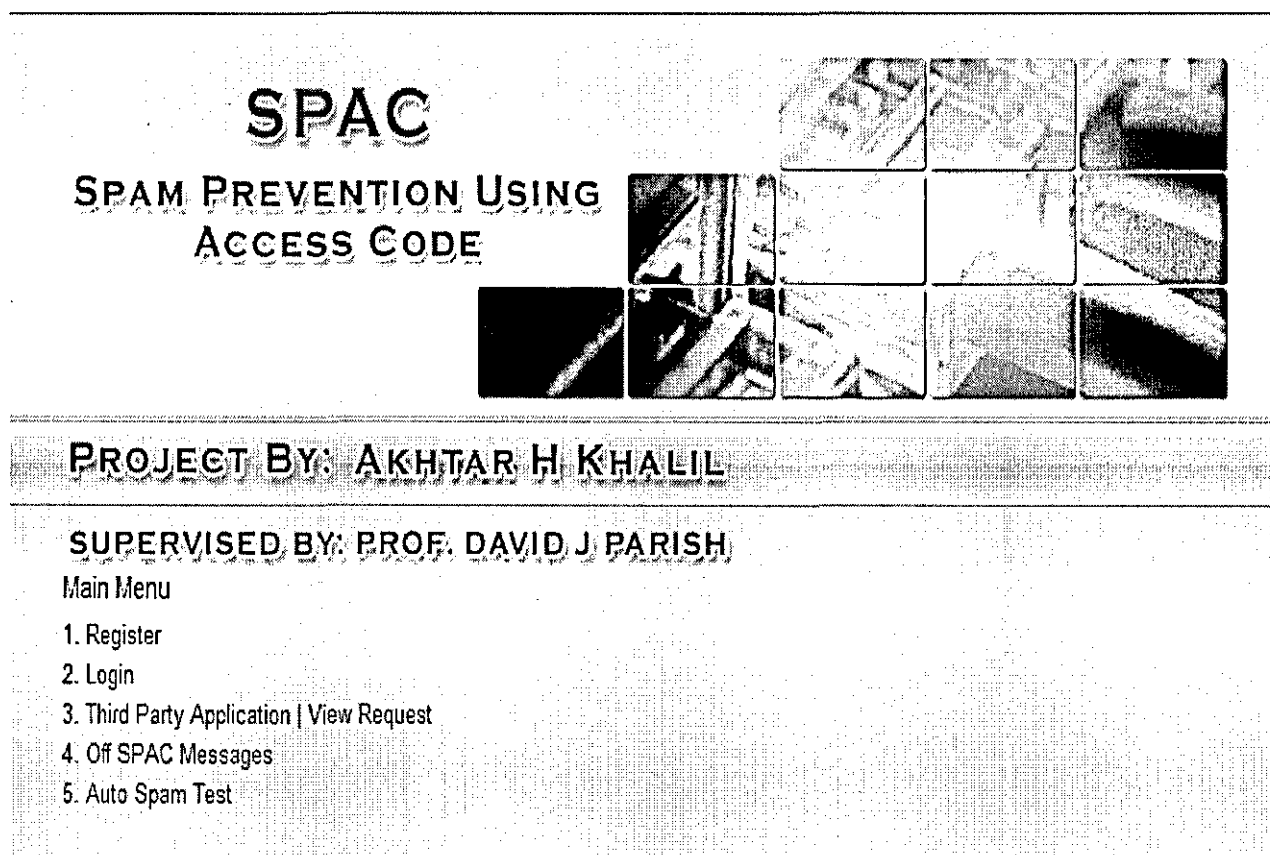


Fig 22, Home page of the SPAC application

The main menu on the SPAC home page shows five options. We discuss the function of each option.

### 5.3.2 Register

Clicking on this link will provide the registration form as given below:

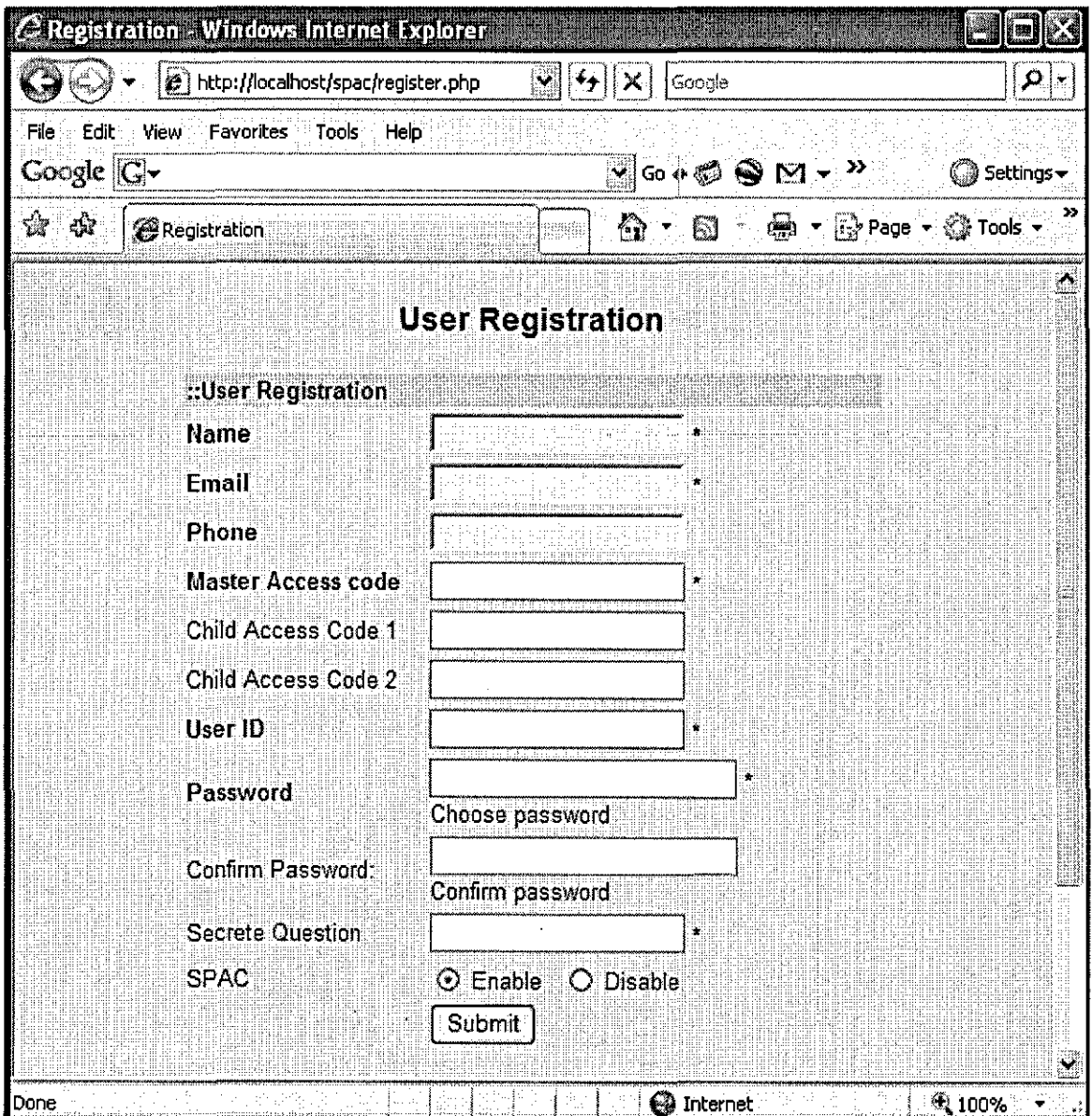
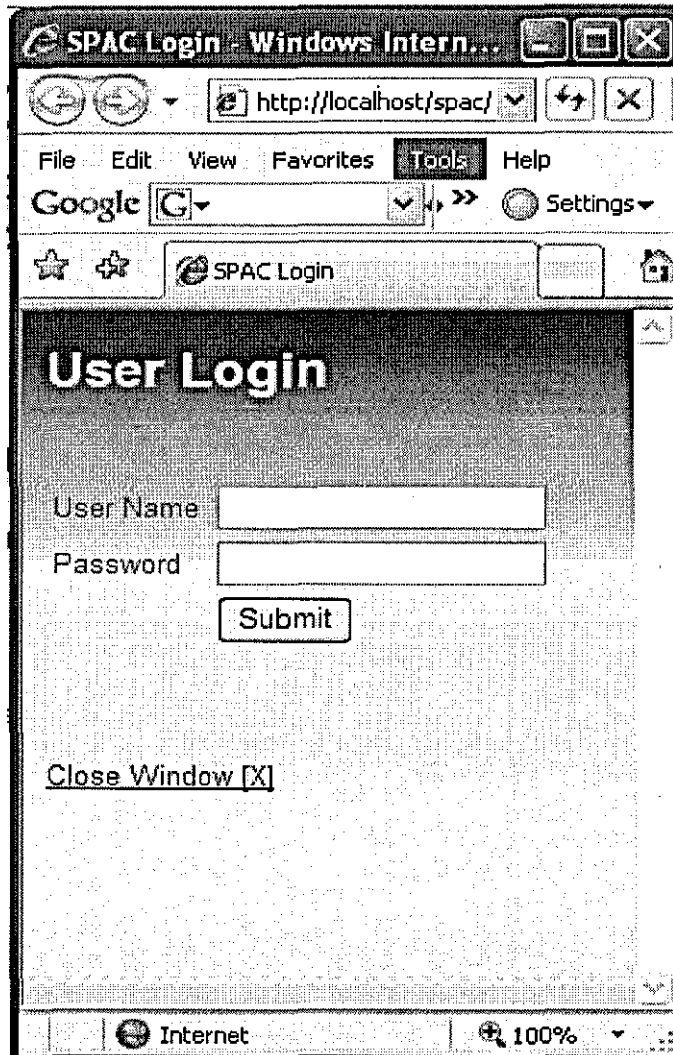


Fig 23, Registration page

At the bottom of the registration form, if the user has selected the “Enable” option, the caller/sender would need to pass through the SPAC mechanism in order to contact this user. If this option is disabled then the recipient is contacted and the SPAC mechanism is not used. A recipient with SPAC disabled will receive messages just as he would in a standard email system. In other words we could say that by disabling the SPAC, the user doesn't mind receiving any spam messages. There exist a large number of Internet users who want to receive advertising, jokes, messages about surveys and/or political messages. In the current opt-in and opt-out procedures the control is in the hands of the sending organisations and companies due to which the users have to first understand and then pass through the opt-out mechanism of each organisation. Also he/she receives spam from unknown organisations. This feature of the SPAC mechanism gives control to the users. He/she can simply enable or disable the SPAC. In other words a user can simply disable or enable (respectively) the delivery of spam messages into his account.

### **5.3.3 Login**

A registered user can login to the SPAC application using an option on the home page. After pressing the login the user needs to give his/her username and password in the following window:



**Fig 24, Login Window**

After giving the correct username and password the client gets the following welcome window which is used to activate the SPAC application.

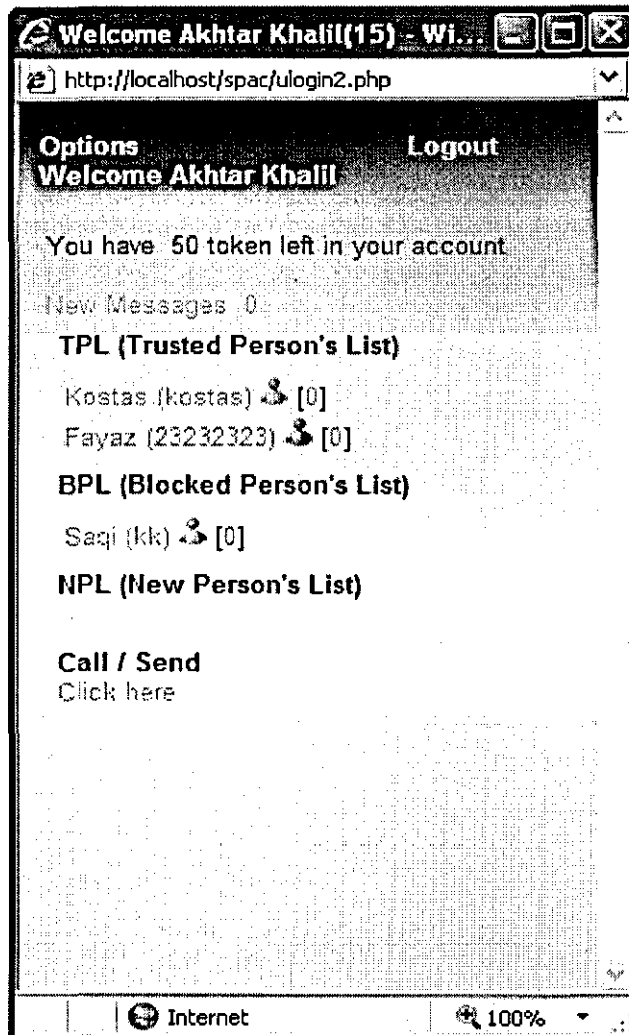


Fig 25, welcome window

Figure 25 shows the welcome window for a user named “Akhtar Khalil” which is displayed in the welcome message.

The figure shows that the ***TPL list*** of the user has two users:

Kostas with ID “kostas”

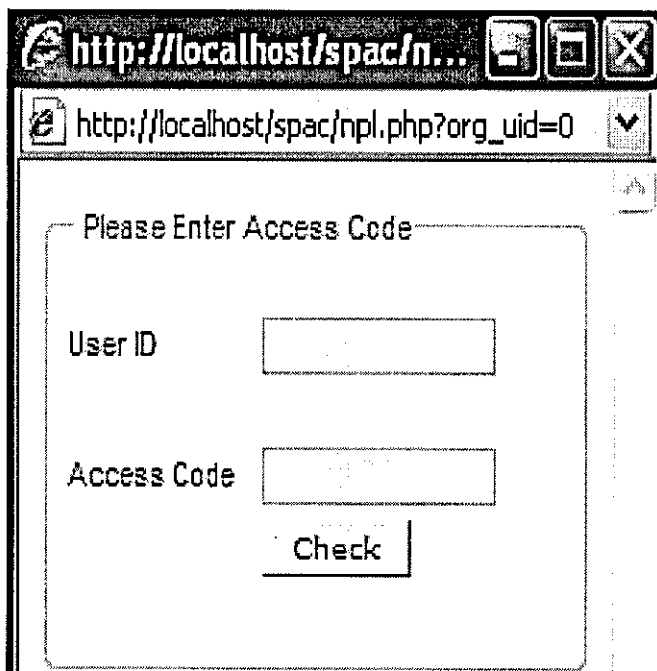
Fayaz with ID “23232323”

These two users can send messages to “Akhtar Khalil” without the need of his AC.

The ***BPL list*** of the Akhtar Khalil contains one user which is Saqi with the usr ID “kk”. kk can never send messages to Akhtar Khalil because Akhtar Khalil has blocked him.

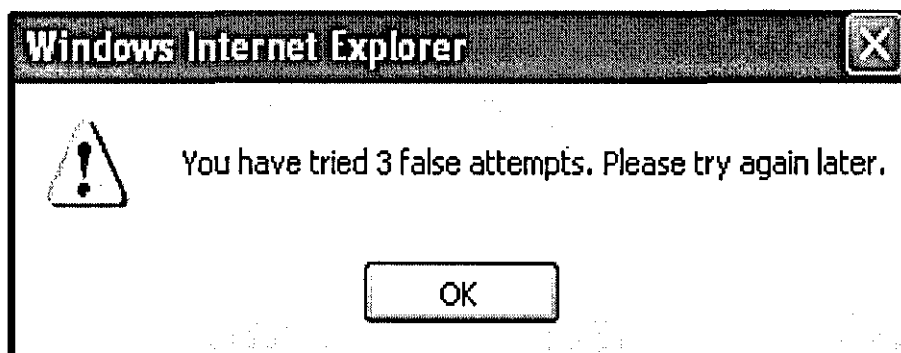
The ***NPL list*** maintains addresses of the new callers/senders. When a caller/sender obtains the AC and would like to make his first call to the recipient then his ID for the first message would be placed into the NPL list of the recipient. The advantage of this list is that if a legitimate user obtains the AC (by passing all the steps of SPAC) but couldn't contact the recipient for some reason (for example if a recipient has another call or is busy and can not attend the call) then for the next try the recipient would not need to repeat all the previous steps. From the NPL list a recipient can shift the caller/sender to his TPL or BPL list.

In order to send a message to Kostas or Fayaz, Akhtar Khalil can directly click on Kostas or Fayaz. To send a message to a new person say Ilyas with username/ID “87878787” (Ilyas in this case is a registered SPAC user), Akhtar Khalil needs to click on the “Click here” link in Call / Send giving the window shown in Figure 26.



**Fig 26, A caller needs ID and AC of the recipient for the first contact**

Akhtar gives the username (or ID) of Ilyas but he also needs to give the access code as well because Akhtar is not in any of the Ilyas's lists. If a caller/sender makes 3 false attempts in providing AC (e.g. if a caller/sender enters a wrong AC code 3 times) then the server locks him out for a certain period of time (1 minute in our test case) as shown in Figure 27.



**Fig 27, message received after making 3 false attempts of AC code**

This feature is used to restrict the number of attempts by a spammer. If Ilyas has not given his access code to Akhtar then Akhtar can access the code from the server. For this he clicks on the "check" button. As he clicks on the check button the server will first check if he has a token in his account. If yes then the server will provide him with some distorted text by using the CAPTCHA program as shown in Figure 7.

The reason for this distorted text in the image is to filter out bots from human users. (Note: In the experimental SPAC application we have not used the CAPTCHA program because the development of this program is not a part of this thesis. However it is a well known mechanism to prevent bots). We have used CAPTCHA with the proposed mechanism to provide further resistance against bots. Information about each recipient is random which will also stop automated spam including spams generated by human beings but combination with CAPTCHA program will further strengthen it against bots. After entering the correct text as given in the distorted text in the image, Akhtar (caller/sender) is provided with the following questions about Ilyas (the recipient).



④ To obtain accesscode answer the following!

What is my country name?

- Germany
- Spain
- UK
- France

What is my name?

- Shafi
- Ilyas
- Jamil
- Fahim

What is my profession?

- Business
- Administration
- Research
- Teaching

What is my department?

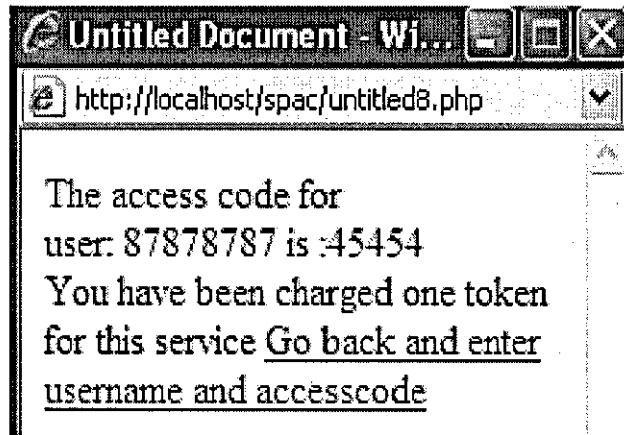
- Mechanical
- Computer science
- Civil
- Electrical

Submit

**Fig 28, challenge step's questions presented to the caller/sender for obtaining AC of the recipient**

If the caller/sender gives correct answers to the questions provided, he will receive the AC of the recipient. It is worth mentioning that the order of questions and the order of

possible answers for a particular recipient (Ilyas in this case) is changed by the server every time the server displays questions. A lock out mechanism has been introduced here. If a user makes 3 false attempts then he/she is locked out for a certain period of time (say 1 minute). After submitting the correct information about Ilyas, the application gives the message shown in Figure 29.



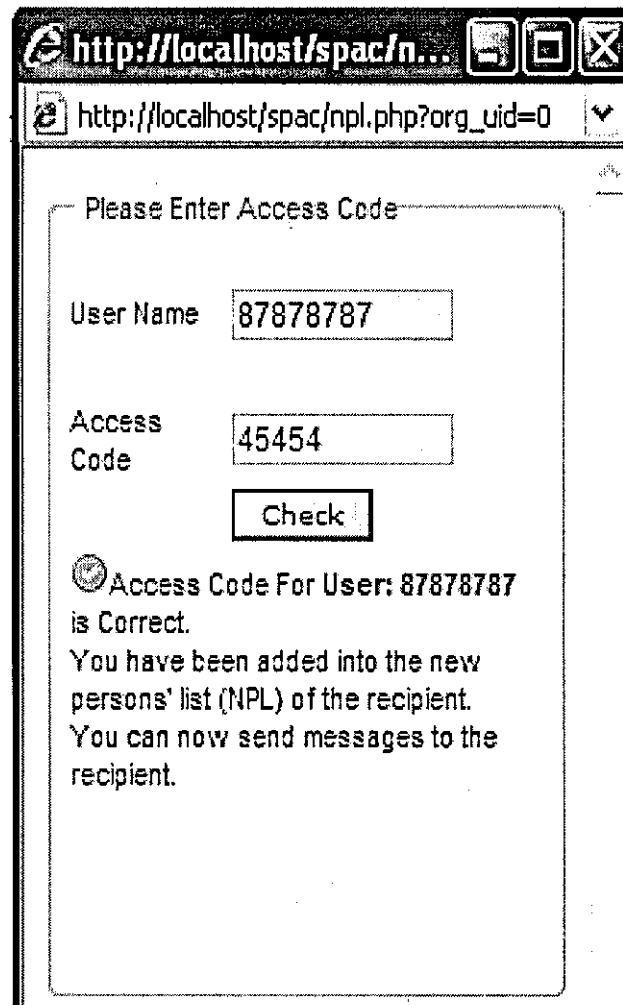
**Fig 29, Access Code provided after passing the challenge step**

Clicking on the Go back link takes the user to the following frame again:

A form titled "Please Enter Access Code" with two input fields: "User ID" and "Access Code". Below the "Access Code" field is a "Check" button.

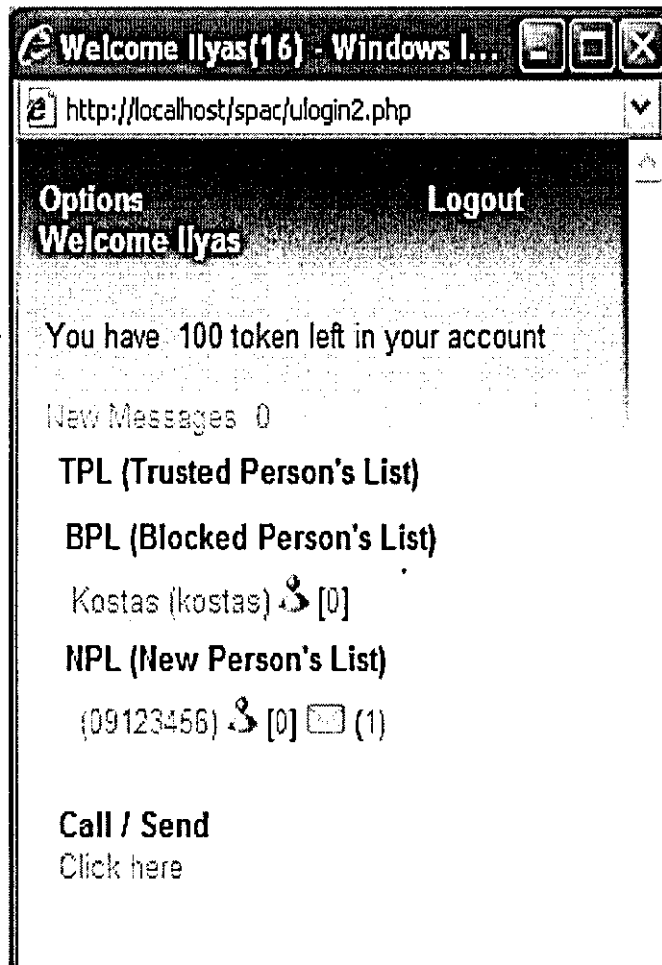
**Fig 30, caller/sender needs to re-enter the AC**

Now the user can give the AC code for the recipient. Submitting this information will add the caller/sender to the NPL list of the recipient (Ilyas). Akhtar is given the following message shown in Figure:



**Fig 31, message showing that the caller/sender is added in the NPL list of the recipient**

Now Akhtar can send messages to Ilyas. Here Akhtar sends a message, "Hello! Ilyas" to Ilyas. We now see the account of Ilyas. Figure 32 shows the welcome window of Ilyas (the recipient):



**Fig 32, welcome window of Ilyas (the recipient)**

The figure shows that he has received one message from the user ID “09123456” which is the ID of Akhtar Khalil. Ilyas can read this message and can update his database by shifting Akhtar to his TPL list (if Akhtar is not a spammer).

### **Options**

The options link is given on the main welcome window (as in figure 33). It gives options as shown in the Figure 33:

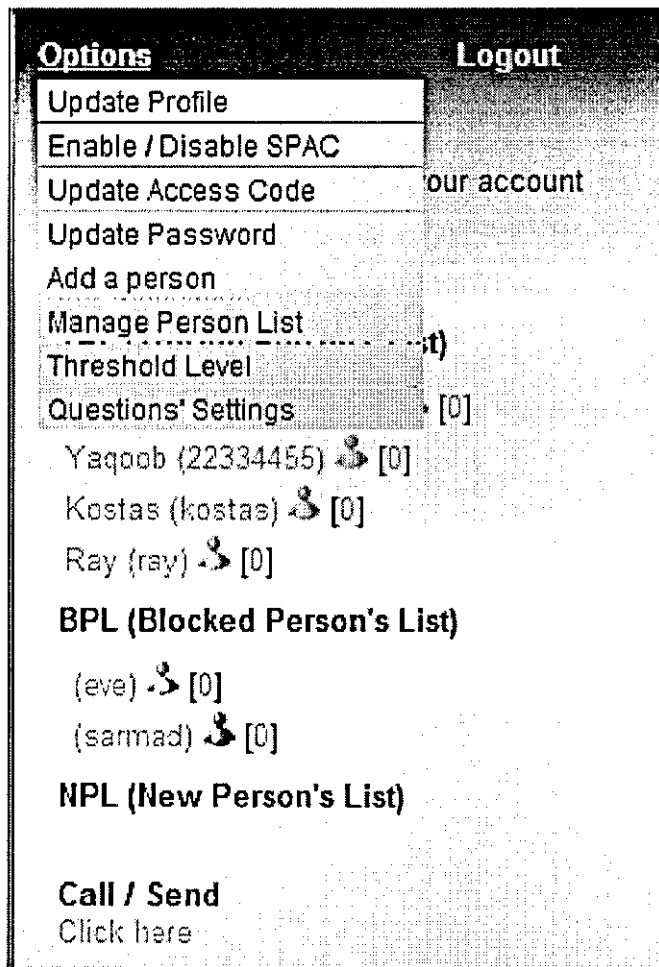


Fig 33, option menu in the welcome window

We discuss these options briefly:

***Update Profile***

Used to update the profile of a user

***Enable / Disable SPAC***

Used to enable or disable the SPAC mechanism

***Update Access Code***

This is used to change the Access Code(s). There are 3 types of Access Codes:

- i. Master Access Code – the normal access code as discussed earlier

- ii. Child Access Code(s) – Access code(s) used on a temporary basis. Its importance is manifested in section 6.4 on Third Party Applications.

#### ***Update Password***

Used to change the password

#### ***Add a person***

This is used to add a person manually to any of the three lists. So, if a user (say David) adds Akhtar to his TPL list then Akhtar would not need the AC of David even for his first contact with David.

#### ***Manage Person list***

Used to give names to different IDs and shifts persons from one list to another

#### ***Threshold Level***

Used to change the threshold level of correct answers to be given for obtaining the correct AC code from the server. A threshold level of 3 means that a caller must provide correct answers to at least 3 questions regarding the recipient in order to obtain his/her access code (AC).

#### ***Questions' Settings***

Used to add, delete or edit questions and answers

Note: It should be noted that the remaining options (Option 3, 4 and 5) of the SPAC home page are not related directly to the SPAC mechanism. However they have been developed to simulate and test the efficiency of the SPAC system in different practical scenarios.

### **5.3.4 Extension for Third Party Application / View Request**

This option on the home page of the SPAC application is used as a third party application in order to create a simulation of an e-commerce web site (e.g. e-bay). For example

David wants to sell his laptop, so he posted the following message on an online auction website or some other site (e.g. his university site for “Sale and Wanted”):

**Put Your Request**

User ID:

Accesscode:

Message:

**Fig 34, example of a message to be posted on an e-commerce website**

It is clear that any user can easily contact him. It should be noticed that here David has given a secondary AC code (or a child AC code) which is used for display in public places and can be changed frequently. Actually this means that until David has sold his laptop, he will receive emails/calls from others (both spammers and legitimate users). The callers/senders (related to the purchase of the laptop) would be using his secondary code to access him. But as soon as David sells the laptop and he wants no more calls about it, he would change his secondary AC code or will deactivate it and with this callers would not be able to access him rather they would need to pass through the challenge mechanism as discussed earlier. This will also indicate all respondents that the author of the message is no longer interested in selling his laptop or he has sold it. This again benefits the seller. Instead of receiving a number of calls after sale, he could simply

change his secondary access code to inform all the callers that the laptop is no more available.

The “View Request” link can be used to view the requests posted by different users. This is a simulation of viewing the online e-commerce website. For example the message posted by David above using the Third Party Application could be viewed by users as:

### Third Application Request

User ID	Access Code	Message
23456	76543	Hello, P4 Laptop (with one year warranty) for sale.
23456	87654	Nokia N95 wanted. Expected range £300 to £350.
87878787	89898	I want to sell my BMW car for £5000.
kostas	88888	I would like to sell my Mac for £400.

Fig 35, simulation of an online e-commerce website

### 5.3.5 Off SPAC Messages

This link is used to send messages to any user of SPAC message. It is a simulation of other applications (e.g. hotmail, yahoo, Skype etc). It is used in relation to the previous link for “Third Party Applications”. For example a user after seeing a message from the view request link is interested in buying David’s Laptop. By clicking on the “Off SPAC Messages” link, the user is provided with the window shown in Figure 36:



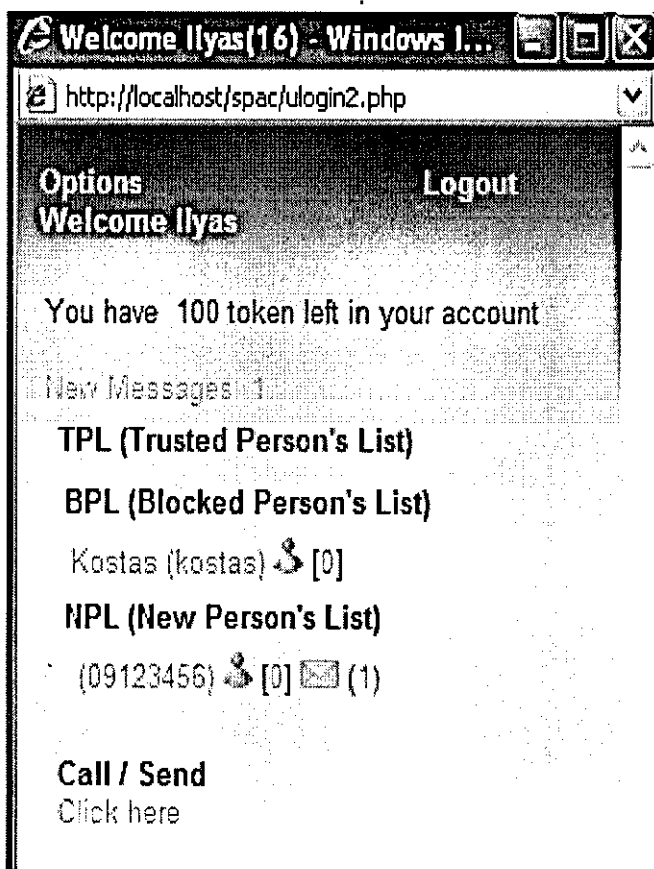
**Third Party Application**

User ID

Accesscode

**Fig 36, window for sending Off-SPAC messages to a user**

The customer can easily access David's ID and his secondary AC code (from Fig 36) in order to send him a message which is received in his account as shown by the "New Messages" link in his Welcome Window; see Figure 37.



**Fig 37, the recipient's welcome window showing new messages**

### 5.3.6 Auto Spam Test

This link is used in the SPAC application home page to test the bots scenario. When the "Auto Spam Test" link is clicked, the application picks up users from a database to send them emails as shown in figure:

Username	<input checked="" type="checkbox"/>	k
	<input checked="" type="checkbox"/>	s
	<input checked="" type="checkbox"/>	kk
	<input checked="" type="checkbox"/>	ss
	<input checked="" type="checkbox"/>	eva
	<input checked="" type="checkbox"/>	mev
	<input checked="" type="checkbox"/>	hussain
	<input checked="" type="checkbox"/>	sadiq
	<input checked="" type="checkbox"/>	kashifm
	<input checked="" type="checkbox"/>	kostas
	<input checked="" type="checkbox"/>	riz
	<input checked="" type="checkbox"/>	23232323
	<input checked="" type="checkbox"/>	09123456
	<input checked="" type="checkbox"/>	87878787
	<input checked="" type="checkbox"/>	4567890
Message	<p>Hello! To help you enjoy your summer holidays, XYZ Airlines is offering 90% discount on all of its International flights. Apply online on <a href="http://www.xyz.com">www.xyz.com</a>. Offer runs till end of July.</p>	
	<input type="button" value="Submit"/>	

Fig 38, example of an auto-spam test

As soon as the spammer wants to send spam messages, all the messages come across the first step of the SPAC mechanism that is the CAPTCHA procedures (as shown in Figure 39) where the bots will fail.

Username					
<input checked="" type="checkbox"/>	<input type="text" value="k"/>	Enter the text given in the image		<input type="text"/>	
<input checked="" type="checkbox"/>	<input type="text" value="s"/>	Enter the text given in the image		<input type="text"/>	
<input checked="" type="checkbox"/>	<input type="text" value="kk"/>	Enter the text given in the image		<input type="text"/>	
<input checked="" type="checkbox"/>	<input type="text" value="ss"/>	Enter the text given in the image		<input type="text"/>	
<input checked="" type="checkbox"/>	<input type="text" value="t"/>	Enter the text given in the image		<input type="text"/>	
<input checked="" type="checkbox"/>	<input type="text" value="m"/>	Enter the text given in the image		<input type="text"/>	
<input checked="" type="checkbox"/>	<input type="text" value="hussain"/>	Enter the text given in the image		<input type="text"/>	
<input checked="" type="checkbox"/>	<input type="text" value="khalil"/>	Enter the text given in the image		<input type="text"/>	

**Fig 39, example of the hindrances provided by SPAC for bots**

This step will filter out the bots from the human users. Again human spammers or machines/computers (used by a spammer in a bot attack) do not know the Access Code (AC) for these users, so they would fail. The only way to access the AC from the server is to pass all the steps of the SPAC mechanism as discussed in section 4.6 which is possible neither for zombies (infected machines controlled by spammers to send bots) nor for the human spammers.

## 5.4 Summary

An application has been developed to test the performance of the SPAC mechanism, the “SPAC application”. This allows the user to register, login and send messages to different

users. Users can also update/edit their databases (including their passwords, questions, profile, Access Code(s), manage lists, and enable/disable SPAC). The SPAC application also provides a mechanism to test the proposed mechanism against bots and to support third-party activities. The SPAC application has all the features of the SPAC mechanism.

# 6 *Experiments and Results*

---

## 6.1 Introduction

This chapter focuses on evaluating the performance of the SPAC mechanism. For this various tests were performed on the previously discussed SPAC application. The aim of these tests was to evaluate the performance of the SPAC mechanism. The SPAC mechanism was tested both for spam messages and also for legitimate messages. In these tests we assumed that if a sender wanted to send message(s) to recipient(s) without any valid reason then it would be considered as a spam message and if a sender wanted to send message(s) to recipient(s) with some valid reason then it would be considered as a legitimate message. We also considered bots where a sender is interested in sending auto-generated spam. Initially we identify the number of free tokens for the charging mechanism. 20 people participated in these tests whereby people on different computers on a LAN were asked to access the SPAC application from the SPAC server (the laptop on which the SPAC application was installed). Each participant performed a total of 55 tests. The distribution of the tests was such that each participant acted as a legitimate sender in 10 tests (5 for legitimate strangers and 5 for approved legitimate senders) and as a spammer in 45 tests (5 as a stranger, 5 as a spammer who is blocked and 35 as simulated bots). We performed a total of 1100 tests. Out of these tests, the senders acted as spammers in 900 tests and as legitimate users in 200 tests. Out of the 900 spam tests, 700 were simulated bots (auto-spam) tests. We found interesting results and features of

the SPAC mechanism. The results obtained showed that SPAC has a clear edge over the existing anti-spam technological approaches. In the following sections we discuss the number of free tokens that we selected for our tests and the results obtained from the practical tests on the SPAC mechanism.

## 6.2 Number of free tokens

We discussed the token charging mechanism in section 4.8. In order to decide the number of free tokens for the charging mechanism, we carried out a survey on 100 phone users to determine an acceptable number of free tokens. Figure 5 shows the number of calls that 100 phone users made to new persons in one month.

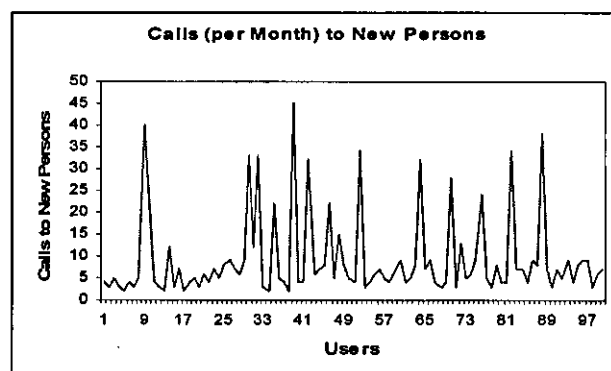


Fig 40, number of calls made to new persons

We assume a similar number for email users. The graph shows that a normal user calls 4 to 10 new persons in one month. Persons dealing in some businesses contacted 25 to 45 new persons per month. In addition to that free tokens are also returned to legitimate callers. From the graph we can clearly observe that 100 tokens per week are enough for a legitimate caller but not sufficient for a spammer because of his interest in generating a large number of spams where the token will not be returned (based on the feedback by the recipient that he/she is a spammer). It should be noticed that this survey is a very informal survey and is meant only for understanding the charging mechanism. However a

detailed study (by a certain ESP or VoIP service provider) of the number of emails/calls to new persons is needed to fix accurately the number of these free tokens.

### **6.3 Results of different tests**

In our tests we considered both legitimate messages and spam messages. Whenever a sender wanted to send a message to a recipient with valid reason, it was considered to be a legitimate message. If the sender wanted to send messages to the recipients without any reason and without any knowledge about the recipient, the messages were assumed as spam messages. The sections below discuss the results obtained from the different tests on the SPAC application.

#### ***Legitimate stranger (first time contact)***

The details of this case have been discussed in section 4.7 (Case 2). In the very first test for the legitimate users, the sender (Kostas) was acting as legitimate user to send a message to a recipient (David). He was provided with the ID of David (23456) and was asked to send a message to David. David could give his AC code along with the ID to Kostas. But this test was meant to check if Kostas can obtain AC code from the server. Kostas came across the following questions about David.

What is my profession?

Engineering

Arts

Business

Medicine

What is my university?

Leicester

UCL

Loughborough

Manchester

What is my name?

John

Ian

Mark

David

What is my location?

Germany

Australia

UK

Spain

What is my Gender?

Female

Dont know

Dont know

Male

**Fig 41, challenge step to get AC of David**

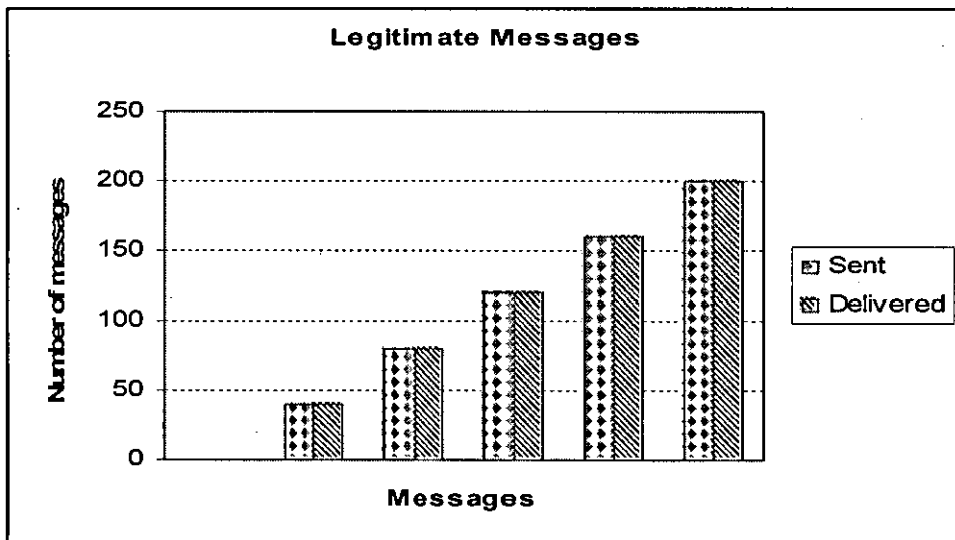
Kostas could easily answer these questions because he knows David and in this case Kostas was a legitimate user. Kostas successfully obtained the AC code from the server which was used for sending a message to David. David added Kostas to his TPL list and Kostas didn't need the AC code for any future messages. We performed 100 such tests to check the performance of the system for the legitimate messages and none of them showed any false positives.



### *Approved (Trusted) Caller*

Case 1 of section 4.7 (chapter 4) discusses the approved caller. We performed 100 tests where different users added different persons manually in their TPL lists. The persons in the TPL list could call the recipients without any need for the AC code. This category also included cases in which the sender is provided with both the ID and AC code of the recipient. The system showed no false positive in this case.

The following graph shows the performance of the SPAC application for legitimate messages.



**Fig 42, graph showing the performance of SPAC for legitimate messages**

The graph above shows that when using the SPAC application all the legitimate messages were delivered with zero false positives. This means that SPAC mechanism doesn't impact the distribution of legitimate messages.

### *Spammer (Stranger)*

We performed 100 tests in which the sender acted as a spammer. These tests included sybil attacks, dictionary attacks and address spoofing. An example of the first spam test that we carried out is as follows:

In this test the participant (say Mujtaba) wanted to send a message to a stranger XYZ. In this test Mujtaba was acting as a spammer because he wanted to send a message to XYZ without any reason and he has no knowledge about XYZ. After passing the CAPTCHA procedure, he came across the following questions:

What is my profession?

Academics

Business

Driving

Management

What is my organisation?

DERA

ERA

PERA

ESPRC

What is my country name?

Germany

Spain

UK

France

What is my name?

John

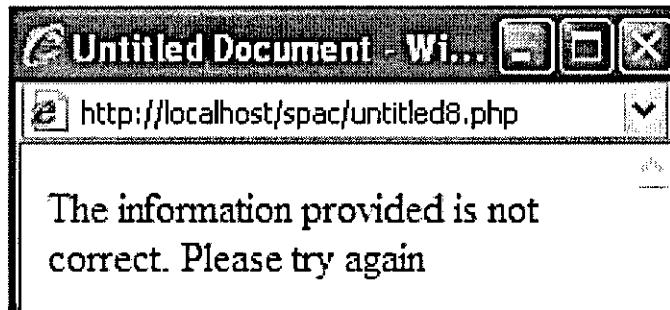
Patrice

James

Carlos

**Fig 43, challenge step of SPAC**

Mujtaba was unable to answer these questions. He made a rough guess and received the message shown in Figure 44:



**Fig 44, message for providing incorrect information**

He tried three false attempts and then he was locked out for a certain amount of time (in this case for 1 minute). Also on every new attempt the sequence of the questions and their possible answers change. For example in the second attempt he received the questions and their possible answers shown in Figure 45:

What is my name?

James

Carlos

John

Patrice

What is my country name?

UK

France

Germany

Spain

What is my profession?

Driving

Management

Academics

Business

What is my organisation?

PERA

ESPRC

DERA

ERA

**Fig 45, challenge step's questions and multiple choices with changed sequence**

In a second test we tried to spam a user with ID Abeera and we obtained the same results. We performed 100 such tests and obtained a false negative in only 2 tests. For 4 questions (with 4 multiple choice answers to each questions) about a recipient, the probability of getting an access code by guessing all the answers correctly would be 0.0039 (1/256) and a probability of 0.00097 (1/1024) for 5 questions. Again remember that due to the lock out mechanism a spammer cannot make more than 3 false attempts for a particular recipient. This means that the spammer could deliver spam at the cost of effort and time which is not acceptable in the spamming business. The beauty of the SPAC mechanism is that it prevents spam and if a spammer does get through (in negligible cases) then it provides such an unpleasant infrastructure environment for the spammer that he gives up and goes away with no profit. Another interesting thing that we would like to mention here is that we asked a number of participants to act as spammers (including the colleagues in my research lab). While acting as spammers all of them showed their impatience with the SPAC mechanism. Their general expression was, *“we will not be able to succeed then why are you asking to try it for so many users”*. Their expression while acting as a spammer in the presence of SPAC mechanism reveals that the SPAC mechanism is very unpleasant for spammers. This also supports the findings of MailChannels (fact iv, section 4.2) that spammers are impatient and they abort the connection if they can not deliver a message within several seconds.

SPAC treated sybil attacks, dictionary attacks and address spoofing the same way because in each case the system function as for unknown persons. The sender had to pass through the CAPTCHA procedure, challenge mechanism and charging mechanism.

### ***Blocked spammers***

Persons with addresses in the BPL (Blocked Persons lists) list of a recipient are blocked spammers for that recipient. Figure 46 shows the welcome window of Prof. Parish and the persons in his BPL list.

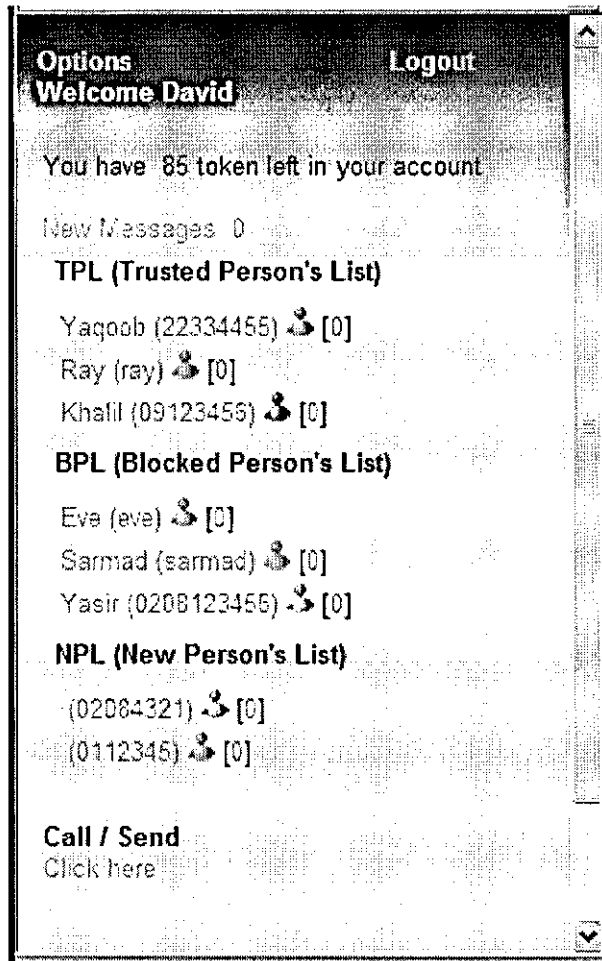


Fig 46, The different lists in the welcome window

We performed 100 tests where persons in the BPL list of a recipient failed to send spam to the recipient. SPAC didn't show any false negative for these tests. For example in the very first such test when the spammer with ID sarmad wanted to send spam to Prof. Parish, he received the message shown in Figure 47.

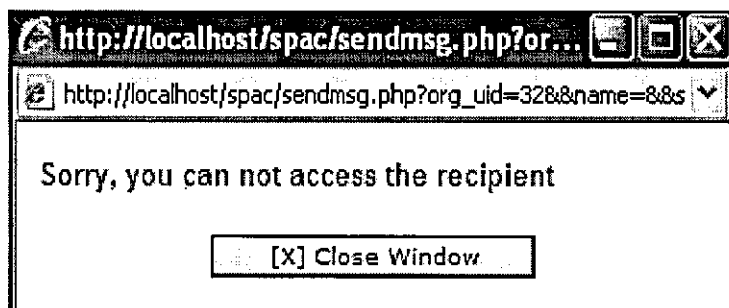










Fig 47, Error message received by a spammer

**Simulated bots (Auto-spam)**

The 20 participants in our test also tested the efficiency of the SPAC mechanism against bots whereby the users were asked to click on a link given on the SPAC home page. After clicking on the link the application selected usernames (IDs) from a list in order to send messages to the recipients with the selected IDs. After clicking the auto-spam test link on the SPAC home page, CAPTCHA procedures codes were introduced to differentiate between a human spammer and auto generated spam machine. This is the first step for obtaining AC code of the SPAC mechanism. This is shown in Figure 48.

Username				
<input checked="" type="checkbox"/>	<input type="text" value="k"/>	Enter the text given in the image		<input type="text"/>
<input checked="" type="checkbox"/>	<input type="text" value="s"/>	Enter the text given in the image		<input type="text"/>
<input checked="" type="checkbox"/>	<input type="text" value="kk"/>	Enter the text given in the image		<input type="text"/>
<input checked="" type="checkbox"/>	<input type="text" value="ss"/>	Enter the text given in the image		<input type="text"/>
<input checked="" type="checkbox"/>	<input type="text" value="l"/>	Enter the text given in the image		<input type="text"/>
<input checked="" type="checkbox"/>	<input type="text" value="m"/>	Enter the text given in the image		<input type="text"/>
<input checked="" type="checkbox"/>	<input type="text" value="hussain"/>	Enter the text given in the image		<input type="text"/>
<input checked="" type="checkbox"/>	<input type="text" value="khalil"/>	Enter the text given in the image		<input type="text"/>

**Fig 48, CAPTCHA codes to stop bots**

Apart from the CAPTCHA codes in the first step, SPAC provides the possible answers and the Access code in the form of distorted text which make it more immune to bots.

Keeping in view the fact that modern computers can not bypass the CAPTCHA procedures, the auto-spam tests failed.

Table 2 summarises the results obtained from different tests on the SPAC application.

Message Type	Sending Technique	Total Tests	False Positive	False Negative	Delivered	Efficiency (%)
Legitimate	Stranger	100	0	—	100	100
	Trusted callers/senders	100	0	—	100	
	<b>Total</b>	<b>200</b>	<b>0</b>	<b>—</b>	<b>200</b>	
Spam	Stranger	100	—	2	2	99.78
	Blocked Spammers	100	—	0	0	
	Simulated Bots	700	—	0	0	
	<b>Total</b>	<b>900</b>	<b>—</b>	<b>2</b>	<b>2</b>	
<b>Total Messages</b>		<b>1100</b>	<b>0</b>	<b>2</b>		<b>99.82</b>

**Table 2, results obtained from tests on the SPAC application**

These results show some interesting facts about the SPAC mechanism which are as follows:

- SPAC doesn't show any false positive (A very interesting and important feature of SPAC)
- No introduction problem for new contacts
- An overall efficiency of 99.8%
- No impact on the distribution of legitimate messages (no false positive)
- The performance of SPAC is not affected by sybil attacks, dictionary attacks and address spoofing which are the common types of attacks used by spammers. In fact, SPAC functions as for unknown persons for all these sorts of attacks.

Figure 49 shows a graph of the performance of the SPAC application against spam messages.

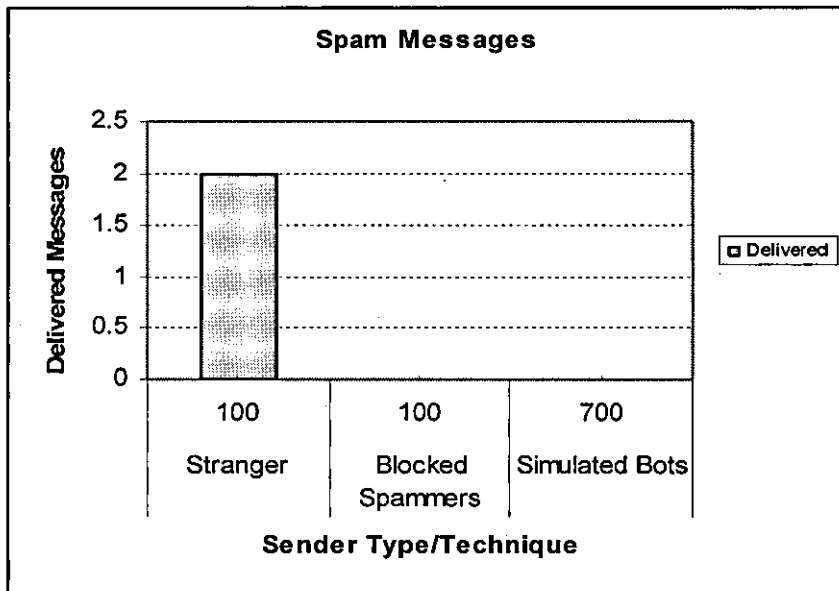


Fig 49, graph showing the performance of SPAC against spam messages

The graph shows that only two out of the 900 spam messages were classified as false negatives.

### 6.4 Third party application (E-commerce websites)

The initial results dug out that there can be some concerns in cases where a person wants to purchase or sell a certain item e.g. a laptop. In such a case the interested customers may or may not be strangers but are still legitimate users. For this we proposed the concept of multiple access codes. Apart from the master access code (also called the primary access code), SPAC offers two other child access codes. We performed 100 such tests in which the sender wanted to sell or purchase something from some e-commerce website such as e-bay.



We could not of course put our fake requests (for testing purposes) for sale or purchase on the real e-bay website or similar. So, we created a simulation of such a website in the SPAC application. The third option, “Third Party Application” on the SPAC home page was used for this purpose. The user can click on it and can post his message along with his child access code. The “View Request” link on the SPAC home page is the analogy of a webpage open to users for viewing posts by different users. A sample snapshot is given in Figure 50:

User ID	Access Code	Message
23456	76543	Hello, P4 Laptop (with one year warranty) for sale.
23456	87654	Nokia N95 wanted. Expected range £300 to £350.
87878787	89898	I want to sell my BMW car for £5000.
kostas	88888	I would like to sell my Mac for £400.

**Fig 50, snapshot of the simulation of e-commerce website**

Note that the secondary access code of the user with ID “2345” (ID of David) is 76543. All the viewers (interested in buying David’s laptop) can send David a message by using his username (ID) and his secondary access code (a temporary access code). David will not change this temporary access code until he doesn’t want to receive messages (regarding his laptop) from all strangers. In this interval even a spammer would be able to send him a spam message. This is a limitation of the SPAC mechanism. David changes his access code after he has sold his laptop. This means that even if David displays his secondary access code for a short time (until he sells his laptop) and the spammers obtain his secondary access code, it will not benefit the spammer because spamming in such a small amount will not profit the spammer as discussed in detail in chapter 4.

## 6.5 Summary

The performance of the SPAC application has been evaluated by performing different tests on the SPAC application. The results obtained from the SPAC application verifies that the SPAC mechanism is targeting spam from two angles i.e. preventing spam and

discouraging spammers. The overall efficiency of the SPAC mechanism was 99.8% with no false positives. Those who were testing it found it so frustrating to spam. This shows that someone who is spamming for money would find it more frustrating. The tests show that SPAC has no impact on the distribution of legitimate messages. At the same time it prevents spam and makes the infrastructure environment so unpleasant for the spammers that they give up. In fact SPAC makes the spamming business non-profitable for the spammers because it wastes a lot of their time, resources, effort and money. SPAC treats sybil attacks, dictionary attacks and address spoofing the same way. SPAC functions as for unknown persons for these sorts of attacks. Based on the results obtained from the various tests on the SPAC application, the SPAC mechanism shows better performance than the existing anti-spam techniques.

# 7 *Performance Comparison*

---

## **7.1 Introduction**

In addition to the idea of Access Code, SPAC uses a combination of technological anti-spam measures. The difference is that SPAC uses those techniques in a different way and combines them in a unique way which enables SPAC to acquire the good features of a number of technological anti-spam approaches without showing the drawbacks of these approaches. This chapter compares the performance of SPAC with existing technological anti-spam approaches.

## **7.2 Comparison with the state-of-the-art anti-spam technological approaches**

Detailed descriptions of the working and drawbacks of state-of-the-art technological anti-spam techniques have been discussed in chapter 3. In order to have a better understanding of the comparison, we will review the drawbacks of state-of-the-art techniques and compare the performance of SPAC with these technologies.

### **7.2.1 Payment**

The payment mechanism is based on charging the sender of an email with a small payment. The idea is to keep the payment so small that it remains negligible for a legitimate user but potentially high for the spammers. However as discussed in section 3.5.1 (chapter 3) this is a difficult trade off and the implementation of such a payment infrastructure can be an ambitious endeavour [40].

There is a lot of risk involved in this mechanism especially in the presence of address spoofing and zombies. Also these payment mechanisms are subject to false praise and create a significant problem in delivering legitimate/solicited bulk emails. The mechanism introduced by the “Penny Black Project” can cause loss to legitimate users in the presence of bots and zombies. The proposers of the Zmail protocol (discussed in section 3.5.1) concluded in [58] that there remain numerous questions about all the payment mechanisms and deployment of the service raises several tricky issues. In all cases, deployment is difficult. In addition the email and VoIP users are reluctant to pay anything extra for emails and VoIP services. The payment approach is against the open flavour of Internet and is expected to punish legitimate users for the misbehaviours of others.

The charging mechanism provided by SPAC doesn't charge any money, rather it provides free tokens on regular intervals (say on per month basis). This is not in contradiction with the open flavour of the Internet. Also this mechanism doesn't charge for each message. Rather it charges only for accessing the access code from the server and returns the token when the recipient doesn't declare the caller/sender, a spammer. Each message of a spammer is to a new or stranger recipient so a spammer will be charged a token for each message. In contrast very few (nearly negligible) messages of legitimate persons are to strangers. Based on the average number of messages to new persons or strangers per month (in our informal survey) we fixed the number of free tokens for the legitimate users (for our tests and experiments). Fixing the number of free tokens for each user becomes easy as compared to the difficult trade off in the existing payment mechanisms. All the drawbacks of the existing charging mechanisms are directly or indirectly related to: sybil attacks, address spoofing and the fact that money is involved in it. As discussed

in section 4.7 that these factors cannot affect the performance of the SPAC mechanism. This enables SPAC to overcome the drawbacks of the existing payment mechanisms.

The payment mechanism in SPAC doesn't involve money so there remains no issue of tricks and money theft as associated with the existing payment mechanisms. At the same time, the charging mechanism in SPAC is not affected by these attacks. In other words the charging mechanism in SPAC doesn't show the drawbacks of the existing payment mechanisms. The tokens are given to each user on per week (or per month) basis so loss of a few tokens (in any case) would not affect the user.

A user can use his account even if he has no tokens in his account. The only limitation is that a user would not be able to obtain the access code from the server if he/she doesn't have the AC in his accounts. Recharging of tokens can also be requested from a server by providing information about the account including password and Access code.

### **7.2.2 White and black lists**

Our mechanism uses a modified form of this technique in the form of TPL and BPL lists. The state-of-the-art white and black lists maintain IPs, URLs or DNS names of the legitimate users and the spammers respectively and they are treated globally. The TPL list and BPL list are similar to the white and black lists in the sense that they maintain addresses of legitimate users and spammers respectively. However, instead of comparing the sender's ID with a global database, SPAC compares the sender's ID with the recipient's database only which is very small (nearly negligible) as compared to the global database used by the state-of-the-art white and black lists. This not only reduces processing load on the server and the wastage of precious resources but also results in a more efficient mechanism. If we assume that XYZ is a URL, IP or DNS which is in the current white list then it will be legitimate for all the users of the ESP and vice versa. This is a drawback because XYZ may be legitimate (approved) for a certain recipient whereas XYZ may not be trusted/approved for some other recipient. Due to this reason such white and black lists are subjected to false praise. In addition, spoofing addresses in the white

lists is not only easy for spammers but also of interest for them because the spammers know that they would be able to deliver the message using the spoofed addresses in the white lists. However, the TPL and BPL lists in our case are relevant to each individual separately. So, a person (say John) may be in the TPL list of a user (say ABC) but it is impossible that John would be in the TPL list of all the users. So, if a spammer wants to use the spoofed address of John for spamming, it would not benefit him because John is not in the TPL list of other recipients. Also the spammer doesn't know the usernames of those accounts which contain John in their TPL lists. Even if the spammer has certain knowledge (which is of course impossible) then he would be able to send spam only to a few persons, in which case spamming is not profitable. This feature of treating the TPL list and the BPL list of each person separately makes SPAC effective against false praise and address spoofing which are amongst the major limitations of state-of-the-art white and black lists as discussed in section 3.5.2. Apart from that, one of the major drawbacks of white and black lists is the introduction problem of the legitimate strangers (or first time callers/senders). With the ability to obtain access code from the server, SPAC overcomes this problem.

Since the spammers keep on changing their identities and addresses, the existing white and black lists also show some issues regarding updating the lists. For example in case of sybil attacks a spammer changes his identity and this needs the black lists to be updated. On the other hand the spammers use false praise techniques to put their addresses in the white lists which creates the need to update these lists as well. SPAC has no such problem because in this mechanism, the TPL and BPL lists are updated by the users individually and these are not maintained by the organisations or ISPs nor are they done publicly. So, this eliminates the possibility of false praise which in turn overcomes the chances of false positives or false negatives). Also SPAC doesn't show any introduction problem. Furthermore, the use of the NPL (as a 3<sup>rd</sup> list) adds to the convenience of the legitimate users as discussed in section 4.4.3.

### 7.2.3 Greylisting

In Greylisting, the Greylisting email server looks for the triplets of the sender. This triplet can be of any legitimate user. A spammer can deliver a spam by spoofing the triplet of any legitimate user. However, spoofing doesn't help the spammer in SPAC as discussed previously.

Spammers bypass the Greylisting mechanism by using software packages that retry delivery to other MX hosts for a domain if delivery through one MX fails. In such a scenario the purpose of Greylisting fails. Our mechanism also bears the call back feature of Greylisting but it is only in cases where a caller/sender is new to the recipients account and the spammers cannot bypass the SPAC mechanism because information about each recipient is different and unknown to the spammers. In such a case, for every new message the caller/sender is called back by the server by providing him with the challenge mechanism.

Greylisting can also result in false positive in cases where an email never passes the triplet rule (for example in cases where the IP address of a host changes). In most of the cases, the sending and receiving clients are unaware of the failure of the delivery. The results discussed in chapter 6 show that SPAC doesn't show any false positive.

Considering the fact as discussed previously that the spam traffic is 4 to 5 times more than the legitimate email traffic, Greylisting results in an increase of email traffic because most of the emails will be resent. This also leads to the conclusion that this technique does not take into account the network resources' abuse. In addition the users may experience unwanted delays due to congestion on the network which results due to the huge amount of resent emails. As opposed to it, SPAC functions as a standard email process for majority of the cases. It calls back a user only in cases where a sender/caller wants to obtain access code from the server. In fact SPAC takes into account the network resources' abuse because this mechanism prevents spam on the connection establishment phase. That is before spam traffic (message or information data) accesses the network, a

sender needs to pass the SPAC mechanism which is not possible for the spammers. As opposed to Greylisting, SPAC doesn't create delay due to congestion.

#### **7.2.4 Challenge/response**

It is discussed in section 3.5.4 that the challenge/response mechanism provides inconvenience to the users. The reason is that that the distorted text or the enriched (with noise or music) audio signal should not be too cluttered to be understood by a legitimate user. Especially in the case of VoIP the problem of different accents and different languages can complicate the communication process and it would be very difficult for a caller to understand the noise enriched audio voice. At the same time if it is made too easy then the intelligent recognition software used by spammers could identify the letters and numbers (in the distorted text) or voice (in case of VoIP).

Instead of providing the caller/sender with random distorted text in images as for CAPTCHA, SPAC asks information about the recipient in the form of distorted text. Since the sender/caller knows the information about the recipient so introduction of noise in voice or distorted text will not create a problem for the user and the relevant text can easily be identified. We can increase the noise level in the case of SPAC to make it more cluttered. This can be achieved by introducing CAPTCHA procedures at the challenge step (i.e. questions and answers). The distorted text and the noise enriched audio signal is not a problem in this case. The reason is that in the case of SPAC, the caller/sender already has information about the answers of the questions provided to the sender/caller by the SPAC mechanism. The multiple answers to different questions are provided to the user in the form of distorted text and/or noise enriched signals. If a user has advance information about the actual answers then it means that the words or answers would not be totally random for the legitimate sender/caller but it would be totally random for a spammer because the spammer doesn't have any such information about the recipient. The final access code is also given to the user in distorted form in an image.



Apart from bots, challenge/response mechanisms like CAPTCHA can not prevent other forms of spam especially those where the spammers send spam manually to a large number of recipients. The spammers take benefit of this weakness of the CAPTCHA procedures and overcome it by using social engineering attacks as discussed in section 3.5.4. In the case of SPAC it would not be possible for a spammer or a third party (like porn website viewer) to pass the challenge phase on behalf of the spammer.

State-of-the-art challenge/response mechanisms also complicate the sending of solicited bulk messages (like newsletter) which will become impossible or costly in the presence of challenges. This is because each challenge will need a human resource. SPAC doesn't show such problems because the sender doesn't need to pass through the SPAC mechanism for persons who are already in the TPL list which are the majority of cases in case of SPAC. It means that SPAC has all the good features of CAPTCHA but overcomes its drawbacks by providing an access code mechanism which can not be bypassed even by spammers (or third party like porn websites viewers) due to the lack of information about the recipient.

### **7.2.5 Cryptographic puzzles**

In the presence of bots, the idea of Cryptographic puzzles not only fails but also causes a significant loss of resources to legitimate users. At the same time with such an approach, a legitimate user with a slow machine can experience unacceptable delays due to the meaningless challenges. SPAC doesn't show the negative responses to the cryptographic puzzles because address spoofing, sybil attacks and/or gaining unlimited computational power of innocent users will not help the spammer in overcoming the challenge steps of information and CAPTCHA. Above all the use of tokens for an account will limit the use of a certain account used by spammers for sending spams.

### **7.2.6 Reputation Mechanisms**

As discussed in section 3.5.6, reputation mechanisms are susceptible to false praise and they don't provide any protection against address spoofing. For these reasons such

mechanisms can result both in false positives and false negatives. Such mechanisms can also cause significant loss to innocent users. As opposed to these mechanisms, SPAC provides the ability for users to declare the reputation of each of their contact persons individually and locally. Therefore the control is with the users and this overcomes the problem of false praise because a user will not classify a trusted person as a spammer in his database otherwise he will lose contact with that person. At the same time the good or bad praise (classifying legitimate or spammer respectively) about a sender/caller by a certain user doesn't affect the sender's status for other users. That is the reputation in the case of SPAC is not global rather it is local. Reputation systems are also ineffective against Sybil attacks whereas SPAC doesn't show any inefficiency in the case of Sybil attacks.

### 7.2.7 Content Filtering

Since content filtering is the most popular and widely deployed anti-spam mechanism so we will compare our proposed SPAC mechanism in detail with the state-of-the-art content filtering mechanism. As mentioned by Enrico Blanzieri and Anton Bryl [13], filtering solves the problems caused by spam only *partially* which prevents end-users from wasting their time on junk messages. But it should be noticed that since all the messages are delivered nevertheless, this mechanism does not prevent resource misuse. Apart from the resource misuse the statement by Enrico Blanzieri and Anton Bryl shows that filtering mechanisms are unable to solve the problem of spam to an acceptable level. The probabilistic nature of the content filtering techniques also encourages the spammers to send a large amount of messages so as to increase the percentage of their delivered messages and benefit from the spamming business. This means that content filtering techniques do not help in preventing the abuse of network/Internet resources. Since SPAC filters spam at the connection phase, spam messages do not access the network resources and are not even given a connection.

Content filtering mechanisms also waste huge amount of resources on processing while checking the contents of the message. Since SPAC doesn't work on content filtering, it doesn't show such drawbacks.

Many filtering research works claim that the accuracy of their filtering algorithms (obtained from the experimental evaluation of their algorithms) is above 90%. However the experimental evaluation of these filtering algorithms can not show the exact real behaviour or efficiency of these algorithms. The reason is that all these experiments (and hence their results) are based on empirical testing and their data sets are small as compared to the global spam data. These results don't take into account all the techniques used by spammers to evade these filters. At the same time their data sets may not be up to date. Also filter-based approaches use different definitions of spam. Due to these reasons we believe that in practice the actual accuracy of these filtering algorithms is less than what the results of the empirical tests show (mostly 90%). At the same time all these filtering algorithms show a considerable amount of false positives. As opposed to this, the results obtained from our different tests on the SPAC mechanism (discussed in chapter 6) show that the accuracy of the SPAC mechanism is 99.8%. In addition, SPAC provides such an unpleasant infrastructure environment for the spammer that the spamming business becomes non-profitable for him. This shows a clear edge of SPAC over the currently most widely deployed anti-spam filtering techniques.

Another edge of SPAC over filter-based approaches is that SPAC doesn't show any false positives (where a ham message is classified as spam). All the filter-based approaches show false positives which are potentially very costly for individuals, companies and organisations. Currently there is no filtering mechanism that can give 100% accurate results (i.e. to find an accurate and precise difference between legitimate and spam messages) and in fact we do not think that such a filtering mechanism is possible due to the lack of a single technical definition of spam (which can be used by filters) and for the reasons given below.

State-of-the-art filtering mechanisms are also costly in terms of maintenance and training the IT staff on using these filters. Configuring the parameters of the spam filters for a trade off between false positives and false negatives is one of the most difficult aspects of the filtering mechanisms. That is to set some linear parameters in spam filters which will stop a good percentage of spam emails but still allowing annoying amounts of spam in order to get ensure a certain level of probability that legitimate emails are not filtered out. Since SPAC doesn't depend on the contents of the message so it doesn't show any drawbacks related to the lack of a single technical definition of spam and/or the cost of IT training and maintenance.

Another unique feature of the performance of the SPAC mechanism is that it creates a distinction about the two types of users; those who like to receive advertising messages and those who do not like to receive such messages. This is achieved by providing a feature for disabling or enabling the SPAC mechanism. In the case of disabling SPAC, the standard operation of email will work and the recipient will receive all messages. This allows the email advertisers to advertise their messages in the best possible way without any regard to evade spam filters. This will reduce the legal issues associated with advertising emails and will increase the return on investment for the advertising companies. SPAC clearly informs the sender about any recipient who has disabled SPAC and who would like to receive advertising emails. This also reduces traffic on the Internet. The reason is that with the current filtering techniques, in order to increase the number of positive responses from the recipient, spammers send large amounts of email because they know that only a small percentage of it will respond or will be interested in the advertisement.

Another benefit of SPAC is that it gives convenience to the users by making them free of all the opt-in and opt-out mechanisms of different companies and organisations. In opt-in and opt-out mechanisms, the control is in the hands of advertising companies (email senders). The email advertising companies are very large in number. So, in these mechanisms it becomes very difficult for the users to understand all the opt-out mechanisms especially as spam messages are more than four times ham messages and

when the senders are not from reputed organisations. However, SPAC transfers this control to the users who can enable and disable reception or rejection of advertising emails by an action as simple as a click of mouse.

### **7.3 Summary**

State-of-the-art techniques show a number of negative impacts. Comparison of the performance of SPAC with existing technological anti-spam approaches reveals that SPAC bears the good features of these techniques without showing any drawbacks. It gives better results than the current filtering techniques in which sometimes the filters filter out legitimate messages. The comparison shows that SPAC has a clear edge over the existing anti-spam techniques. As compared to other technological approaches, SPAC needs less management and reduces the need for IT staff. It also prevents the network resources misuse.

# 8

# *Conclusion*

---

## **8.1 Conclusion**

This thesis has addressed the spam problem and has developed an anti-spam mechanism known as SPAC (Spam Prevention using Access Codes). The thesis has discussed the losses caused by spam which were underestimated for many years and has showed that state-of-the-art anti-spam techniques, policies or training have helped the Internet from collapsing due to spam but they likely to have very limited positive impact. None of the existing anti-spam solutions can prevent sufficient spam to be considered a solution for the spam problem. The losses caused by spam are worth billions of dollars per annum and the growth rate of spam is increasing every year. The thesis found that there is a strong desire for an efficient anti-spam mechanism that can stop sufficient amounts of spam with no negative impact on the distribution of legitimate messages. This research has proposed and evaluated the SPAC (Spam Prevention using Access Codes) mechanism which uses AC codes to prevent spam and to combine various anti-spam approaches. There are many contributions of the thesis including the use of Access Codes for preventing spam and for combining some of the popular anti-spam mechanisms to achieve better results, no impact on the distribution of legitimate messages, ability of SPAC to prevent spam in its various forms (voice, audio, text and image), prevention of the consumption of precious network resources and the little or no effect of the different types of spam attacks

(including address spoofing attacks, sybil attacks, dictionary attacks and bots) on SPAC. This work found a limitation of the charging mechanism of SPAC which can work only if a global charge structure is introduced. However, this limitation does not affect the overall performance of the SPAC mechanism. The results obtained from the tests on the SPAC application reveal that the objectives as mentioned in section 4.1 have been achieved by the SPAC mechanism. These tests show that SPAC is very unpleasant for spammers. It wastes a lot of their time, resources, effort and money. SPAC doesn't show any false positives. Apart from spam email and SPIT, SPAC can be used to prevent spam in other applications as well such as cellular telephony, traditional telephony and instant messaging services. It gives better results than the current filtering techniques in which sometimes the filters filter out legitimate messages and in most of the cases neither the recipient nor the sender knows that the message has been filtered or blocked. Detailed analysis of the results obtained from the tests on the SPAC application (as discussed in chapter 6) and study of the characteristics of the spammer and a legitimate user show that the challenge and charging mechanisms provided by SPAC are very difficult, unpleasant and costly for the spammer. As compared to other technological approaches, SPAC needs less management and reduces the need for IT staff and saves the network's resources misuse. SPAC provides the user with a degree of convenience because legitimate trusted users (which are the majority of the cases for accounts of legitimate recipients) will not need the AC code and provides them with a better alternative to the opt-out mechanism. The overall performance of SPAC is better than the existing anti-spam techniques.

## **8.2 Areas of Application**

The beauty of the SPAC mechanism is that it prevents spam on the connection establishment phase. After the connection is established, no processing is done on the contents of the message. So, there is no issue of SPAC related to the QoS of various technologies like VoIP. Because of this reason, SPAC is applicable to a number of other technologies such as cellular telephony, traditional telephony and instant messaging service. In all these technologies, SPAC prevents spam on the connection establishment phase which involves text data in the connection establishment phase.

## 8.3 Future Work

We will continue to explore interesting features and properties of the SPAC mechanism. Future research would include fixing the number of free tokens. This would require collaboration from some reputed ESP or VoIP service provider. The reason is that they already have the records of all the values that are needed for fixing the number of these tokens. Specifically we will be interested in information regarding the average number of new contacts per month that each user makes. We would also look into improving the database structure for the SPAC mechanism. This part will focus on developing a database mechanism which generates automatic questions based on the information given by the user at the time of registration and will also suggest the answers for those questions.

In case of successful implementation of spam where it would be difficult or non-profitable to send spam, the spammers might attempt to hack the server or the user account with the aim of accessing the TPL list or tokens. In such a case the spammer would be able to send spam to the recipient by spoofing the IDs in their TPL list. However, the spammer would not be able to send spam in bulk because an ID in the TPL list of a person X is not necessary to be in the TPL lists of other users. Preventing the spammers from hacking the server or users' accounts is not a part of this research. However, our future work will also focus on making SPAC more immune to cases where a spammer accesses the TPL lists by hacking the server and/or users' accounts.

The reduction in the number of tokens of a legitimate user (as a result of address spoofing) can show him/her that his address is being used illegitimately. Thus he would be able to take further actions. At the same time it can also help in research about address spoofing which is not a part of this research. When provided with a feedback mechanism, the owner of the account could contact the email service provider (ESP) about the illegitimate use of his/her account which will help the ESP in collecting information and data about such abuse. This data and information from the user would help in the research about address spoofing. Our future work would focus on designing a feedback



mechanism with a well-designed database that updates itself every time the user gives a feedback. The database would give a number of features/attributes of address spoofing and spoofers and would help in finding a way to overcome it.

## References

- [1] Frost & Sullivan, "*World Content Filtering Markets, Market Engineering Research*", 25 July 2007
- [2] "*Networking Basics CCNA1 Companion Guide*", Cisco Press, 30<sup>th</sup> March 2006, ISBN -10: 1587131641
- [3] Behrouz A. Forouzan, "*Data Communications and Networking*", Mc Graw Hill, 2003, ISBN: 007-251584-8
- [4] Peter Thermos, Ari Takanen, "Securing VoIP Networks, Threats, Vulnerabilities, and Countermeasures", Addison-Wesley, August 2007, ISBN: 0321-43734-9
- [5] RFC3067. J. Arvidsson, A. Cormack, Y. Demchenko, J. Meijer., "*TERENA'S Incident Object Description and Exchange Format Requirements*", 2001.
- [6] Satya Bhan, Jonathan Clark, Joshua Cuneo, Jorge Mejia-Ramirez, "*Information Security Issues in Voice Over Internet Protocol*", 2006, pp 10
- [7] Alan B. Johnston and David M. Piscitello, "*Understanding Voice Over IP Security*", 1st ed., vol. 1, London: Artech House, 2006, ISBN: 1-59693-050-0
- [8] Wikipedia, "*Man-in-the-middle Attack*", Retrieved on 28<sup>th</sup> Sept, 2007 from [http://en.wikipedia.org/wiki/Man-in-the-middle\\_attack](http://en.wikipedia.org/wiki/Man-in-the-middle_attack)
- [9] David Piscitello, "*Anatomy of a Phishing Attack*", Core Competence Inc. Retrieved 16<sup>th</sup> June 2008 from: <http://hhi.corecom.com/phishingexpedition.htm>
- [10] The Nucleus Research, "*Spam: The Silent ROI Killer*", Inc. D59, June 2003
- [11] Organisation for Economic Co-operation and Development, "*Background Paper for the OECD Workshop on Spam*", Jan 2004, 57 pp
- [12] Guido Schryen, "A formal approach towards assessing the effectiveness of anti-spam procedures", 39<sup>th</sup> Hawaii International conference on system Sciences, 2006, 10 pp
- [13] Enrico Blanzieri and Anton Bryl, "*A Survey of Learning-Based Techniques of Email Spam Filtering*", Technical Report # DIT-06-056, University of Trento, Italy, January 2008, 30 pp
- [14] Andy Walker, "*Absolute beginner's guide to security, spam, spyware & viruses*", Que Publishing, 2006, ISBN: 0-7897-3459-1
- [15] Wikipedia, "*Spam*", Retrieved 22<sup>nd</sup> July 2008 from: <http://en.wikipedia.org/wiki/Spam>
- [16] Homel Foods Corporation, "*SPAM family of products*", Retrieved 9<sup>th</sup> August from: <http://www.hormelfoods.com/brands/spam/default.aspx>

- [17] SpamHaus, "*The Definition of Spam*", Retrieved on 12<sup>th</sup> July 2008 from: <http://www.spamhaus.org/definition.html>
- [18] Ion Androutsopoulos, John Koutsias, Konstantinos V. Chandrinou, and Constantine D. Spyropoulos, "*An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with personal e-mail messages*", *23rd annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, ACM Press, 2000, pp 160–167
- [19] SpamDefined, "*Spam defined*", Retrieved 16<sup>th</sup> July 2008 from: <http://www.monkeys.com/spam-defined/>
- [20] Gordon Cormack and Thomas Lynam, "*Spam corpus creation for TREC*", Second Conference on Email and Anti-Spam, CEAS 2005 proceedings, Stanford University, USA 21-22 July 2005, 2 pp
- [21] ITU, "*ITU Survey on Anti-spam Legislation Worldwide*", 28 June – 1 July 2005, Document: CYB/06, 62 pp
- [22] Enrico Blanzieri and Anton Bryl, "*A Survey of Anti-Spam Techniques*", University of Trento, Italy, Technical Report # DIT-06-056, September 20, 2006, 19 pp
- [23] Guido Schryen, "*Anti-Spam Measures Analysis and Design*", New York, Springer-Verlag Berlin Heidelberg, 2007, ISBN 978-3-540-71748-5
- [24] Jon Postel, "*On the Junk Mail Problem*", RFC 706, NIC # 33861, IETF Network Working Group, Nov 1975.
- [25] Peter J. Denning, "*Electronic junk*", *Communications of the ACM*, Vol. 25, Issue 3, 1982, pp 163-165
- [26] Michael Specter, "*Damn Spam, The losing war on junk e-mail*", *The New Yorker*, 6<sup>th</sup> August 2007, Retrieved 24<sup>th</sup> July 2008 from: [http://www.newyorker.com/reporting/2007/08/06/070806fa\\_fact\\_specter](http://www.newyorker.com/reporting/2007/08/06/070806fa_fact_specter)
- [27] Brad Templeton, "*Reaction to the DEC Spam of 1978*", Retrieved 22<sup>nd</sup> July 2008 from: <http://www.templetons.com/brad/spamreact.html>
- [28] Templeton, B.: n.d.b, "*Origin of the term 'spam' to mean net abuse*", Retrieved 28<sup>th</sup> August 2008 from: <http://www.templetons.com/brad/spamterm.html>
- [29] Andy Walker, "*Absolute beginner's guide to security, spam, spyware & viruses*", Que Publishing, 2006, ISBN: 0-7897-3459-1
- [30] Stas Bekman, "*High-performance asynchronous IO for SMTP Multiplexing*", MIT Spam Conference 2007, USA
- [31] Spam-o-meter, "*Spam stats*", Retrieved 8<sup>th</sup> August 2008 from: <http://www.junk-o-meter.com/stats/index.php>

- [32] Christopher Lueg, "Spam and Anti-Spam Measures – A Look at Potential Impact" Informing Science InSITE - "Where Parallels Intersect" June 2003, 10 pp
- [33] FerrisResearch, "The global economic impact of spam", Report #409
- [34] CopiaTech Articles, Antispam, "Top 5 Impacts of Spam to a Business", July 2007
- [35] BBC News, (2004, 20<sup>th</sup> September), "Net security threats growing fast", Retrieved on 16<sup>th</sup> July 2008 from: <http://news.bbc.co.uk/1/hi/technology/3666978.stm>
- [36] Zulfikar Ramzan and Candid Wuest, "Phishing Attacks: Analyzing Trends in 2006", Proceedings of CEAS 2007 Fourth Conference on Email and AntiSpam, Mountain View, CA, USA 2-3 August 2007, 8 pp
- [37] Mikko Siponen and Carl Stucke, "Effective antispam strategies in companies: An international study", Proceedings of the 39th Annual Hawaii International Conference on System Sciences, HICSS '06, volume 6, 2006, 10 pp
- [38] Evangelos Moustakas, C. Ranganathan, and Penny Duqueno, "Combating spam through legislation: A comparative analysis of US and European approaches", proceedings of second conference on Email and Anti-Spam, CEAS'2005, Stanford University, USA, 21-22 July 2005, 8 pp
- [39] Gordon Worley, "The negative impacts of spam", 2001, Retrieved 12 June 2008 from: [http://homepage.mac.com/redbird/doc/neg\\_impacts\\_spam.html](http://homepage.mac.com/redbird/doc/neg_impacts_spam.html)
- [40] Rainer Baumann, St'ephane Cavin and Stefan Schmid, "Voice over IP - security and SPAM", University of Berne, pages 9, 10, 11-14, August 24 - September 8, 2006, 34 pp
- [41] Techfaq, "What is SPIT?", Retrieved 4<sup>th</sup> August 2008 from: <http://www.tech-faq.com/spit.shtml/What%20is%20SPIT?>
- [42] David Endler and Mark Collier, "Hacking Exposed VoIP: Voice Over IP Security Secrets & Solutions", McGraw-Hill/Osborne © 2007, ISBN: 9780072263640
- [43] Enrico Blanzieri and Anton Bryl, "A Survey of Anti-Spam Techniques", University of Trento, Italy, September 20, 2006
- [44] Official Journal of the European Communities, "Directive 2002/58/EC of the European Parliament and of the Council", L 201/37, July 2002, 11 pp.
- [45] Nicola Lugaresi, "European Union vs. Spam: A legal response", First Conference on Email and Anti-Spam, CEAS'2004 proceedings, Mountain View, CA, USA, 30-31 July, 2004, 8 pp
- [46] Wikipedia, "CAN-SPAM Act of 2003", Retrieved 24<sup>th</sup> July 2008 from: [http://en.wikipedia.org/wiki/CAN-SPAM\\_Act\\_of\\_2003](http://en.wikipedia.org/wiki/CAN-SPAM_Act_of_2003)
- [47] Petr Piškula and Jana Klaschková, "Report on non-OECD Countries' Spam Legislation", OECD, 30<sup>th</sup> April 2004, 16 pp

- [48] John P. Mello Jr., "Malware 101: University offers course on spyware", TechNewsWorld, 10<sup>th</sup> February 2005, Retrieved 13<sup>th</sup> August 2008 from: <http://www.technewsworld.com/story/40479.html>
- [49] John Aycock, "Teaching Spam and Spyware at the University of Calgary", *Proceedings of the 3<sup>rd</sup> Conference on Email and Anti-Spam, CEAS'2006* proceedings, Mountain View, California, USA, 27-28 July 2006, pp 137-140
- [50] Lorenzo Lazzari, Marco Mari, and Agostino Poggi, "Cafe - collaborative agents for filtering e-mails", 14th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprise, WETICE'05 proceedings, 2005, pp 356-361
- [51] Feng Zhou, Li Zhuang, Ben Zhao, Ling Huang, Anthony Joseph, and John Kubiawicz, "Approximate object location and spam filtering on peer-to-peer systems", ACM/IFIP/USENIX International Middleware Conference proceedings, Middleware 2003, 2003, pp 1-20
- [52] HoneyPot. Project honey pot, "Distributed spam harvester tracking network", Retrieved on 14<sup>th</sup> July 2008 from: <http://www.projecthoneypot.org/?rf=41074>
- [53] SpotSpam, "The European Spambot Project", Retrieved 11<sup>th</sup> August 2008 from: <http://www.spotspam.net/>
- [54] "Memorandum of understanding on mutual enforcement assistance in commercial email matters among the agencies of the United States, the United Kingdom and Australia", 2004, Retrieved 11<sup>th</sup> August 2008 from: <http://www.ftc.gov/os/2004/07/040630spammoutext.pdf>
- [55] International spam enforcement network, "London Action Plan On International Spam Enforcement Cooperation", 11<sup>th</sup> October 2004, Retrieved from: <http://www.londonactionplan.org/>
- [56] Taughannock Networks, "Technical responses to spam", November 2003, pp 11
- [57] Larry Seltzer, "Should senders pay for the mess we call e-mail?" eWeek, 18<sup>th</sup> October 2003, Retrieved 12<sup>th</sup> July 2008 from: <http://www.eweek.com/article2/0,4149,1273186,00.asp>
- [58] Benjamen Kuipers, Alex Liu, Aashin Gautam and Mohamed Gouda, "Zmail: zerosum free market control of spam", IEEE Fourth International Workshop on Assurance in Distributed Computing Systems and Networks (ADS-N), ICDCSW 2005, IEEE Computer Society, 2005, pp. 20-26
- [59] Martin Abadi, Andrew Birrell, Mike Burrows, Frank Dabek and Ted Wobber, "Bankable Postage for Network Services", 8th Asian Computing Science Conference, 2003 Proceedings, 20 pp
- [60] Microsoft Research, "The Penny Black Project", Microsoft, Retrieved 31<sup>st</sup> July 2008 from: <http://research.microsoft.com/research/sv/PennyBlack/>
- [61] Brad Templeton, "E-Stamps", Retrieved 31<sup>st</sup> July 2008 from: <http://www.templetons.com/brad/spam/estamps.html>
- [62] Anirudh Ramachandran, David Dagon and Nick Feamster, "Can DNS-Based Blacklists Keep Up with Bots?", Third Conference on Email and Anti-Spam, CEAS'2006 Proceeding, Mountain View, California, USA, 27-28 July 2006, pp 55-56

- [63] Harris, E., *"The Next Step in the Spam Control War: Greylisting"*, 2003.
- [64] Captcha, *"The CAPTCHA project"*, Retrieved 13<sup>th</sup> July 2008 from: <http://www.captcha.net/>
- [65] C. Jennings, *"Computational Puzzles for SPAM Reduction in SIP"*, Cisco Systems, Network Working Group, Internet-draft, 2007
- [66] Martin Abadi, Mike Burrows, Mark Manasse and Ted Wobber, *"Moderately Hard, Memory Bound Functions"*, NDSS proceedings, 2003, pp 25-39
- [67] The Honey net Project & Research Alliance, *"Know your Enemy: Tracking Botnets"*, 2005, Technical report.
- [68] Bradley Taylor, *"Sender Reputation in a Large Webmail Service"*, Third Conference on Email and Anti-Spam, CEAS' 2006 proceedings, Mountain View, CA, USA, 27-28 July 2006, pp 116-121
- [69] Joshua Goodman, *"IP addresses in email clients"*, First Conference on Email and Anti-Spam, CEAS'2004 Proceedings, Mountain View, CA, USA, 30-31 July 2004, 8 pp
- [70] Mikko Siponen and Carl Stucke, *"Effective antispam strategies in companies: An international study"*, HICSS '06 proceedings, volume 6, 2006, 10 pp
- [71] Alexandru Catalin COSOI, *"Methods for dynamically combining relevancies of different antispam filters"*, MIT 2007 Spam conference, 2007, 10 pp
- [72] Hrishikesh Aradhye, Gregory Myers, and James Herson, *"Image analysis for efficient categorization of image-based spam e-mail"*, Eighth International Conference on Document Analysis and Recognition, ICDAR 2005 proceedings, IEEE Computer Society, volume 2, 2005, pp 914-918.
- [73] Ching-Tung Wu, Kwang-Ting Cheng, Qiang Zhu, and Yi-Leh Wu, *"Using visual features for anti-spam filtering"*, IEEE International Conference on Image Processing, ICIP 2005 Proceedings, volume 3, 2005, pp 509-512.
- [74] Cara Garreston, *"Anti-spam market braces for shakeout"*, Network World, 2003
- [75] Ion Androutsopoulos, Evangelos Magirou, and Dimitrios Vassilakis, *"A game theoretic model of spam e-mailing"*, Second Conference on Email and Anti-Spam, CEAS'2005 Proceeding, Stanford University, USA, 21-22 July 2005, 8 pp
- [76] Wen-tau Yih, Joshua Goodman, and Geoff Hulten, *"Learning at low positive rates"*, Third Conference on Email and Anti-Spam, CEAS'2006 proceedings, Mountain View, CA, USA, 27-28 July 2006, pp 87-94
- [77] Ion Androutsopoulos, Georgios Paliouras, and Eirinaios Michelakis, *"Learning to filter unsolicited commercial e-mail"*, Technical Report 2004/2, National Center for Scientific Research, Demokritos, 2004, 52 pp

- [78] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz. "A Bayesian approach to filtering junk email", Proceedings of the AAAI Workshop on Learning for Text Categorization, 1998, pp 55 – 62
- [79] S. Webb, S. Chitti, and C. Pu. "An experimental evaluation of spam filter performance and robustness against attack", Proceedings of the 1<sup>st</sup> International Conference on Collaborative Computing (CollaborateCom '05), December 2005, 8 pp
- [80] Eirinaios Michelakis, Ion Androutsopoulos, Georgios Paliouras, George Sakkis, and Panagiotis Stamatopoulos. "Filtron : a learning-based anti-spam filter", Proceedings of the 1<sup>st</sup> Conference on Email and Anti-Spam, CEAS'2004, Mountain View, CA, USA, 30-31 July 2004, 8 pp
- [81] Stas Bekman, "High-performance asynchronous IO for SMTP Multiplexing", MIT Spam Conference 2007, USA
- [82] Wikipedia, "Adobe Dreamweaver", Retrieved 18<sup>th</sup> July 2008 from:  
<http://en.wikipedia.org/wiki/Dreamweaver>
- [83] Mehdi Achour, Friedhelm Betz, Antony Dovgal, Nuno Lopes, Philip Olson, Georg Richter, Damien Seguy, Jakub Vrana, Gabor Hojtsy, "PHP Manual", The PHP Documentation Group, March 2007
- [84] "MySQL 3.23, 4.0, 4.1 Reference Manual", MySQL AB, 27<sup>th</sup> March 2007

## **Appendix A – Dreamweaver**

To create web pages easily we have used Macromedia Dreamweaver 8 as our visual editor. It can hide the details of the HTML code of pages from the user. This makes it easy for non-coders to create web pages and sites.

With Dreamweaver (as discussed in [82]), users can preview websites in many browsers, provided they are installed on their computer. It has some site management tools including the ability to find and replace lines of text or code by whatever parameters specified across the entire site, and a templating feature for creating multiple pages with similar structures. The behaviors panel in Dreamweaver enables use of basic JavaScript without any coding knowledge.

With the use of "Extensions" - small programs in Dreamweaver, any web developer can write (usually in HTML and JavaScript). In addition, extensions provide added functionality to the software for whomever wants to download and install them. A large community of extension developers support Dreamweaver who make extensions available (both commercial and free) for most web development tasks from simple rollover effects to full-featured shopping carts.

Like other HTML editors, Dreamweaver edits files locally, then uploads all edited files to the remote web server using FTP, SFTP, or WebDAV.



## Appendix B – PHP (Hypertext Preprocessor [83])

PHP stands for "Hypertext Preprocessor" which is a widely-used Open Source general-purpose scripting language that is especially suited for Web development and can be embedded into HTML. We have used PHP as our server technology to build web applications because PHP server technology is supported by Dreamweaver. Scripting or tag-based language to be used is selected based on the server technology available on the server. The most popular languages for the five server technologies supported by Dreamweaver are as follows:

<b>Server technology</b>	<b>Language</b>
ColdFusion	ColdFusion Markup Language (CFML)
ASP.NET	Visual Basic C#
Active Server Pages (ASP)	VBScript JavaScript
Java Server Pages (JSP)	Java
PHP	PHP

Since we use PHP as our server technology so our scripting language for the SPAC application is PHP. PHP development is focused on server-side scripting. It means that the code is executed on the server. The client receives the results of running the script but no way of determining the underlying code. PHP can do anything that other CGI program can do, such as collect form data, generate dynamic page content, or send and receive cookies.

As described in [83], PHP scripts are used in three main areas:

### *Server-side scripting*

This is the most traditional field for PHP and is used in our SPAC application. Three things are required to make this work. The PHP parser (CGI or server module), a web server and a web browser. We need to run the web server, with a connected PHP installation. The PHP program output can be accessed with a web browser, viewing the

PHP page through the server. All these can also run on a home machine for experimenting with PHP programming.

### ***Command line scripting***

You can make a PHP script to run it without any server or browser. You only need the PHP parser to use it this way. This type of usage is ideal for scripts regularly executed using cron (on \*nix or Linux) or Task Scheduler (on Windows). These scripts can also be used for simple text processing tasks.

### ***Writing desktop applications***

PHP is not the very best language to create a desktop application with a graphical user interface but you can use some advanced PHP features in your client-side applications by using PHP-GTK to write such programs. You also have the ability to write cross-platform applications this way. PHP-GTK is an extension to PHP, not available in the main distribution. It greatly simplifies writing client-side cross platform GUI applications.

PHP can be used on all major operating systems, including Microsoft Windows, Mac OS, Linux, many Unix variants (including HP-UX, Solaris and OpenBSD), RISC OS, and probably others. In addition, PHP has support for most of the web servers today. This includes Apache, Microsoft Internet Information Server, Personal Web Server, Netscape and iPlanet servers, O'Reilly Website Pro server, Caudium, Xitami, OmniHTTPd, and many others. It means that with PHP, a user has the freedom of choosing an operating system and a web server. Furthermore, a user has the choice of using procedural programming or object oriented programming, or a mixture of them. Besides outputting HTML, PHP outputs include images, PDF files and even Flash movies (using libswf and Ming) generated on the fly.

PHP can support a wide range of databases which is one of the strongest and most significant features in PHP. Writing a database-enabled web page is incredibly simple. Currently PHP supports the following databases:

Adabas D  
InterBase  
PostgreSQL  
dBase  
FrontBase  
SQLite  
Empress  
mSQL  
Solid  
FilePro (read-only)  
Direct MS-SQL  
Sybase  
Hyperwave  
MySQL  
Velocis  
IBM DB2  
ODBC  
Unix dbm  
Informix  
Oracle (OCI7 and OCI8)  
Ingres  
Ovrimos

PHP also uses a database abstraction extension (named PDO) allowing the users to transparently use any database supported by that extension. Additionally PHP supports ODBC, the Open Database Connection standard, so the user can connect to any other database supporting this world standard.

## **Appendix C - MySQL**

MySQL is one of the most popular Open Source SQL database management systems which is developed, distributed, and supported by MySQL AB. MySQL AB is a commercial company, founded by the MySQL developers. The essence about the importance and features of MySQL as taken from [84] are as follows:

### ***MySQL is a database management system***

A database is a structured collection of data. It may be anything from a simple shopping list to a picture gallery or the vast amounts of information in a corporate network. To add, access, and process data stored in a computer database, you need a database management system such as MySQL Server. Since computers are very good at handling large amounts of data, database management systems play a central role in computing, as standalone utilities, or as parts of other applications.

### ***MySQL is a relational database management system***

A relational database stores data in separate tables rather than putting all the data in one big storeroom. This adds speed and flexibility. The SQL part of “MySQL” stands for “Structured Query Language.” SQL is the most common standardized language used to access databases and is defined by the ANSI/ISO SQL Standard. The SQL standard has been evolving since 1986 and several versions exist. “SQL-92” refers to the standard released in 1992, “SQL:1999” refers to the standard released in 1999, and “SQL:2003” refers to the current version of the standard. The phrase “the SQL standard” is used for the current version of the SQL Standard at any time.

### ***MySQL software is Open Source***

Open Source means that it is possible for anyone to use and modify the software. Anybody can download the MySQL software from the Internet and use it without paying anything.

***The MySQL Database Server is very fast, reliable, and easy to use***

MySQL Server was originally developed to handle large databases much faster than existing solutions and has been successfully used in highly demanding production environments for several years. Although under constant development, MySQL Server today offers a rich and useful set of functions. Its connectivity, speed, and security make MySQL Server highly suited for accessing databases on the Internet.

***MySQL Server works in client/server or embedded systems***

The MySQL Database Software is a client/server system that consists of a multi-threaded SQL server that supports different backends, several different client programs and libraries, administrative tools, and a wide range of application programming interfaces (APIs).

***A large amount of contributed MySQL software is available***

It is very likely that your favorite application or language supports the MySQL Database Server.

The official way to pronounce “MySQL” is “My Ess Que Ell” (not “my sequel”), but it can be pronounced as “my sequel” or in some other localized way.



