On Privacy and the Prevention of Unsolicited Sessions in the IP Multimedia Subsystem

David Waiting

Supervisor:
Neco Ventura

Thesis Presented for the Degree of
Doctor of Philosophy
in the Department of Electrical Engineering
University of Cape Town
March 2008

Declaration

I declare that the above thesis is my own unaided work, both in concept and execution, and that apart from the normal guidance from my supervisor, I have received no assistance except as stated in the text of this document.

This work is being submitted for the Doctor of Philosophy Degree in Electrical Engineering at the University of Cape Town. Neither the substance nor any part of the above thesis has been submitted in the past, or is being, or is to be submitted for a degree at this university or at any other university.

_____ _____

David Waiting Date

Acknowledgements

I would like to acknowledge the following people for their assistance with this work:

- Neco Ventura, my supervisor. Thank-you for your guidance and unwavering commitment to the Centre of Excellence.
- Richard Good, proof reader and golf partner. One day QoS will experience the glory of the IPtv superstars.
- Natalie Burls. Thanks for the support and encouragement.
- Saverio Niccolini, NEC Europe. I thoroughly appreciate the feedback on my work and the access to your resources.
- Peter Weik, Fraunhofer FOKUS. Thank-you for the frank discussions on IMS in Berlin and Washington DC. Best of luck in your new marriage.
- Dragos Vingarzan, Fraunhofer FOKUS. Well done on the Open IMS Core project. I believe most of its success can be attributed to your hard work.
- Simon Woodford, Forschungszentrum Jülich. Thank-you for the help with the tricky parts.
- Thomas Magedanz and Yacine Rebahi, Fraunhofer FOKUS. Thank-you for the excellent feedback.
- Aymeric Moizard, antiSIP. Your libraries and your quick replies to my many queries over the years have been invaluable.
- Past and present members of the Communications Research Group.

Synopsis

The Internet has evolved from a platform serving static web pages and email to an all-encompassing communications environment that supports real-time voice, instant messaging, presence, video conferencing and a growing number of rich interactive services. This is in stark contrast to the circuit-switched telephony world that has seen little innovation since its inception. The IMS (IP Multimedia Subsystem) has been hailed as the technology that will merge traditional telephony with the Internet.

The IMS is an all-IP based architectural framework that provides multimedia services, inter-operator roaming, differentiated quality of service and charging capabilities. It is founded on the technologies and protocols of the world's leading Internet standards organisation, the Internet Engineering Task Force. Unfortunately, by emulating an Internet environment the IMS stands to inherit many of its problems.

One such problem is the huge volume of unsolicited bulk email, commonly known as spam. Spam has many different purposes including marketing, fraud, identity theft and electioneering, and the volume of spam in the Internet far outweighs the volume of legitimate communications. Due to the reduced costs that can be realised by an IMS architecture and its similarity with the Internet, it is expected that the problem of spam will manifest itself within IMS networks soon after deployment.

The thesis investigates this problem of multimedia spam using a bottom-up approach. The work begins by identifying the origins of multimedia spam and how it will affect next generation telecommunication networks such as the IMS. A thorough literature review is performed that summarises the current state of the art in spam prevention techniques and several case studies of previous attacks are investigated.

A limited-scale survey is performed to gauge the current prevalence of multimedia spam in South Africa and to determine consumer attitudes towards spam. The survey finds that voice and text-message spam is commonly received by the survey participants and that there is a general negative attitude towards receiving spam in any form. The survey also finds that participants hold the network service providers responsible for preventing spam.

The 3GPP has identified spam as a potential problem in the IMS but as yet has not developed any solutions. Therefore, in order to aid this process the

thesis proposes a novel spam prevention architecture that mitigates the spam problem with a multi-layered solution, including call pattern analysis, Turing tests and computational puzzles. The thesis also introduces a novel hypothesis that abnormal calling behaviour can be detected by examining the durations of a user's past calls. In order to validate this hypothesis an experiment is performed in which automated calls are made to random recipients and the duration of these calls is measured. It is found that the experimental results differ from the calling patterns of legitimate callers and as such the hypothesis is not disproved.

The proposed spam prevention architecture is implemented in a fully standards compliant practical network test-bed. The test-bed demonstrates conclusive proof-of-concept and provides a platform for several evaluations. These evaluations include measuring the overheads introduced by the proposed architecture and testing the various modules and mechanisms in a variety of different scenarios. It is found that while the proposed solution adds overhead to both the IMS-level registration and session setup procedures it is able to correctly identify abnormal calling behaviour and slow the rate of session setup of identified spammers.

This study covers a new area of IMS research and as such there is little comparison to be made with existing literature. However, the work performed provides a relevant and timely point of departure for future work in this field and highlights one of the most concerning problems that will be faced by IMS networks in the near future.

Contents

Declaration i

Acknowledgements ii

Synopsis iii

List of Figures xii

List of Tables xvi

I Preliminaries 1

1 Introduction 2

1.1 VoIP Revolution . . . 3
1.2 Migration to All-IP and the NGN . . . 4
1.3 Birth of IP Multimedia . . . 6
1.4 Standardisation of the 3GPP IMS . . . 7
1.4.1 IMS as a Service Delivery Platform . . . 8
1.4.2 IMS Challenges . . . 9
1.5 A History of Unsolicited Communications . . . 11
1.5.1 The Economies of Multimedia Spam . . . 12
1.5.2 Address Harvesting . . . 13
1.6 The Manifestation of Spam into the NGN . . . 14
1.6.1 Multimedia Spam Types . . . 15
1.6.2 Spam Threats in the IMS / NGN . . . 16
1.7 Thesis Objectives . . . 19
1.8 Thesis Scope and Limitations . . . 20
1.9 Thesis Outline . . . 22

2 Literature Review **24**
2.1 IETF Spam Prevention Literature 25
2.1.1 The Session Initiation Protocol and Spam 25
2.1.2 Computational Puzzles for Spam Reduction in SIP . . . 31
2.1.3 Spam Score for SIP 33
2.1.4 A Framework to Tackle Spam 34
2.1.5 A Document Format for Expressing Authorisation Policies 37
2.2 Detection of Voice Spam 38
2.2.1 Progressive Multi Gray-Levelling 39
2.2.2 Detecting SPIT Calls by Checking Human Communication Patterns 40
2.3 Detection of Instant Message Spam 42
2.3.1 A Bayesian Approach to Filtering Junk E-Mail 42
2.3.2 Image Spam 43
2.4 Spam Legislation 45
2.4.1 Electronic Communications and Transactions Act of 2002 45
2.4.2 CAN Spam Act of 2003 45
2.4.3 Do Not Call Registry 46
2.5 3GPP Working Group 47
2.6 Discussion 48

3 Case Studies **49**
3.1 Russian Business Network Harbours Online Crime 49
3.2 Pump and Dump Scams Cause Havoc on Stock Market 50
3.3 Users Defrauded with the Unknown Missed Call 52
3.4 Robo-call Flood in 2006 Mid-Term Elections 52
3.5 Vodacom MMS Campaign 54
3.6 Discussion 54

II Data Collection **56**

4 Consumer Perceptions on Spam - A Survey **57**
4.1 Motivation 57
4.2 Item Construction 58
4.2.1 Section A 58
4.2.2 Section B 58

4.2.3 Section C . . . 59
4.3 Validity and Reliability of the Questionnaire . . . 59
4.4 Data Collection . . . 59
4.4.1 Pilot Survey . . . 59
4.4.2 Pilot Analysis and Revisions . . . 60
4.4.3 Final Survey . . . 61
4.4.4 Data Analysis . . . 61
4.5 Current State of Spam (Items 1-4) . . . 62
4.5.1 Email Spam . . . 62
4.5.2 Unsolicited Voice Calls . . . 62
4.5.3 Cellular Telephone Text Messaging Spam . . . 64
4.5.4 Multimedia Messaging Spam . . . 64
4.6 Spam Tolerance (Items 5 and 6) . . . 65
4.6.1 Tolerance For Unsolicited Voice Calls . . . 65
4.6.2 Tolerance For Unsolicited Text and Multimedia Messages 66
4.7 Attitudes Towards Spam (Items 7-16) . . . 67
4.8 Accountability (Item 17) . . . 69
4.9 Discussion . . . 70

5 Robo-call Experiment 71
5.1 Hypothesis . . . 72
5.2 Motivation for Experiment . . . 72
5.3 Limitations . . . 73
5.4 Methodology . . . 73
5.4.1 Pre-recorded Message . . . 73
5.4.2 Website . . . 74
5.4.3 Robo-call . . . 74
5.5 Results . . . 75
5.5.1 Response Ratio . . . 75
5.5.2 Call Durations . . . 76
5.6 Analysis . . . 77
5.6.1 Histogram Analysis . . . 77
5.6.2 Bimodal Probability Density Function . . . 78
5.7 Comparison with Legitimate Callers . . . 78
5.7.1 Data Source . . . 79
5.7.2 Outliers . . . 80

5.7.3 Call Duration Histograms 80
5.8 Discussion 80

III Proposed Solution 84

6 Spam Prevention Architecture 85
6.1 Design Considerations 85
6.1.1 Discourage Spam and Ensure Privacy 86
6.1.2 Unobtrusive 86
6.1.3 Conformance to Standards 87
6.1.4 Resistance to False Positives 88
6.1.5 Scalability 89
6.1.6 User Configurability 89
6.2 Proposed Multilayered Solution 90
6.2.1 Legislation 90
6.2.2 Whitelist / Blacklist 91
6.2.3 Call Pattern Analysis 91
6.2.4 Turing Test / Computational Puzzle 93
6.3 Solution Architecture 93
6.3.1 IMS Core Control 95
6.3.2 Rule Database 96
6.3.3 Call Pattern Analyser 97
6.3.4 Authorisation Engine 98
6.4 Discussion 100

7 Test-bed Implementation 101
7.1 IP Connectivity Access Networks 102
7.1.1 IPv4 vs IPv6 102
7.1.2 Local Area Network 103
7.1.3 IEEE 802.11 103
7.1.4 EDGE 103
7.1.5 HSDPA 104
7.2 IMS Core Network, HSS and Registration 104
7.2.1 Call Session Control Functions 104
7.2.2 Home Subscriber Server 105
7.2.3 IMS Registration 106

7.2.4 Application Server Invocation . . . 108
7.3 User Equipment . . . 109
7.3.1 SIPp . . . 109
7.3.2 UCT IMS Client . . . 110
7.4 Call Pattern Analyser . . . 111
7.4.1 SIP Proxy Behaviour . . . 111
7.4.2 Data Collection . . . 112
7.4.3 Session Duration Analysis Module . . . 113
7.4.4 Session Volume Analysis Module . . . 116
7.4.5 Concurrent Session Analysis Module . . . 117
7.4.6 Overall Spam Probability Score Calculation . . . 117
7.5 Rule Database XDMS . . . 118
7.5.1 XCAP URIs . . . 118
7.5.2 XCAP Signalling . . . 118
7.5.3 Default Behaviour . . . 119
7.6 Authorisation Engine . . . 119
7.6.1 Server Tasks . . . 120
7.6.2 Proxy Behaviour vs. UAS Behaviour . . . 120
7.6.3 Decision Engine . . . 121
7.6.4 Computational Puzzle . . . 124
7.6.5 Turing Test . . . 126
7.7 Summary . . . 127

8 Evaluations and Results 128
8.1 Overheads . . . 129
8.1.1 Method . . . 129
8.1.2 Registration Traffic Overhead . . . 131
8.1.3 Session Setup Traffic Overhead . . . 131
8.1.4 Session Setup Delay Overhead . . . 133
8.1.5 Discussion . . . 135
8.2 Session Volume Analysis Module . . . 135
8.2.1 Method . . . 136
8.2.2 Scenarios . . . 136
8.2.3 Results . . . 138
8.2.4 Discussion . . . 138
8.3 Concurrent Session Analysis Module . . . 139

8.3.1 Method . . . 139
8.3.2 Scenarios . . . 139
8.3.3 Results . . . 140
8.3.4 Discussion . . . 140
8.4 Session Duration Analysis Module Function Minimiser . . . 142
8.4.1 Method . . . 142
8.4.2 Scenarios . . . 143
8.4.3 Results . . . 143
8.4.4 Discussion . . . 147
8.5 Session Duration Analysis Module Correlation . . . 147
8.5.1 Method . . . 147
8.5.2 Results . . . 148
8.5.3 Discussion . . . 148
8.6 Computational Puzzle . . . 150
8.6.1 Method . . . 150
8.6.2 Scenarios . . . 151
8.6.3 Results . . . 151
8.6.4 Discussion . . . 153
8.7 Turing Test . . . 153
8.7.1 Method . . . 153
8.7.2 Results . . . 154
8.7.3 Discussion . . . 154
8.8 Summary . . . 155

9 Conclusions and Further Study Items 156
9.1 Conclusions . . . 156
9.1.1 IMS Adoption . . . 156
9.1.2 IMS Challenges . . . 157
9.1.3 Spam No Longer Just Text . . . 157
9.1.4 State of the Art . . . 158
9.1.5 Types of Spam . . . 158
9.1.6 Spam Legislation . . . 159
9.1.7 Perceptions on Spam . . . 159
9.1.8 Response to Robo-calling . . . 160
9.1.9 Test-bed Implementation . . . 160
9.1.10 Solution Overheads . . . 161

9.1.11 Session Volume Analysis . . . 162
9.1.12 Concurrent Session Analysis . . . 162
9.1.13 Session Duration Analysis . . . 163
9.1.14 Computational Puzzle . . . 163
9.1.15 Turing Test . . . 164
9.2 Further Study Items . . . 164
9.2.1 Survey Data . . . 164
9.2.2 Spam Data . . . 165
9.2.3 Text Message Filters . . . 165
9.2.4 Call Duration Analysis . . . 165
9.2.5 Computational Puzzles . . . 166
9.2.6 CAPTCHAs . . . 166
9.2.7 3GPP Standardisation Efforts . . . 167

Bibliography **168**

A Survey Documents **176**
A.1 Pilot Survey . . . 176
A.1.1 Section A . . . 176
A.1.2 Section B . . . 178
A.1.3 Section C . . . 179
A.2 Final Survey . . . 180
A.2.1 Section A . . . 180
A.2.2 Section B . . . 182
A.2.3 Section C . . . 183

B Robo-call Pre-recorded Message **184**

C Accompanying CD-ROM **185**

D List of Abbreviations **186**

List of Figures

1.1 An example of a stock touting email. This message was flooded over the Internet as a PDF attachment on the 8th August 2007. The stock in question trades for fractions of a cent. 17

2.1 SIP Puzzle header proposed by Jennings - the image and pre-image are base64 encoded. 32
2.2 SIP message with spam scores appended to the via headers. . . 33
2.3 The framework proposed by Tschofenig *et al.* for the interaction between the end user and the authorisation proxy. 35
2.4 An authorisation policy document that forwards all callers who fail a Turing test to voice-mail. 38
2.5 A four-state two-way telephone conversation model. 40
2.6 An image spam advertising generic drugs. The text geometry is skewed and random noise is added making optical character recognition problematic. 44

4.1 Item 1 - How often do you receive unsolicited email? 63
4.2 Item 2 - How often do you receive unsolicited phone calls on your cellphone? 63
4.3 Item 3 - How often do you receive unsolicited SMSs? 63
4.4 Item 4 - How often do you receive unsolicited MMSs? 63
4.5 Item 5 - How many unsolicited phone calls would make you stop using your cellphone altogether? 65
4.6 Item 6 - How many unsolicited SMSs or MMSs would make you stop using your cellphone altogether? 66
4.7 Item 9 - Receiving unsolicited cellphone calls bothers me. 68
4.8 Item 10 - I feel bothered by unsolicited SMSs. 68

4.9 Item 12 - Unsolicited calls and SMSs reduce the enjoyment I receive from using my cellphone. 68
4.10 Negative spam attitude score frequencies. 68
4.11 Item 17 - Who, in your opinion, is responsible for managing cellphone spam? 69

5.1 Robo-call experiment - response ratio 75
5.2 Robo-call experiment - call duration results. 76
5.3 Robo-call experiment - histogram of call durations. 77
5.4 The original histogram normalised and overlaid with the bimodal probability density function f. 79
5.5 The call duration histograms of random legitimate callers (excluding outliers) overlaid with an exponential probability density function (Part I). 81
5.6 The call duration histograms of random legitimate callers (excluding outliers) overlaid with an exponential probability density function (Part II). 82

6.1 Proposed multilayered spam prevention architecture. 90
6.2 The solution architecture depicted with an IMS core network, user terminal and spam source. 94
6.3 Structure of the Rule Database. 96
6.4 Modular architecture of the Call Pattern Analyser. 97
6.5 The Authorisation Engine makes decisions based on several sources of information. 99

7.1 A trigger point is configured using the FHoSS web interface. . . 105
7.2 The UCT IMS Client registers with the FOKUS Open IMS Core.106
7.3 User profile containing the initial filter criteria for the Call Pattern Analyser. 107
7.4 The Call Pattern Analyser and the Authorisation Engine are traversed in a serial fashion. 108
7.5 Session setup between two SIP clients (left) and two IMS clients (right). 110
7.6 Users are allowed to initiate a maximum number of sessions during a sliding time window. 116
7.7 Storing and retrieving documents from the Rule Database (XDMS).119

7.8 The Authorisation Engine proxies a request back to itself so that it can be answered by the user agent server. 121
7.9 An authorisation document illustrating the use of the spam-score element. 123
7.10 Signalling flows for the computational puzzle (left) and the Turing test (right) between the UAC and the Authorisation Engine UAS 125
7.11 An example of the *Refer-To* header field. 126

8.1 The three evaluation cases. 130
8.2 Session setup signalling (Case III). 132
8.3 Session volume analysis module evaluation results - Scenario 1. . 137
8.4 Session volume analysis module evaluation results - Scenario 2. . 137
8.5 Session volume analysis module evaluation results - Scenario 3. . 137
8.6 Session volume analysis module evaluation results - Scenario 4. 138
8.7 Concurrent session analysis module evaluation results - Scenario 1. 141
8.8 Concurrent session analysis module evaluation results - Scenario 2. 141
8.9 Concurrent session analysis module evaluation results - Scenario 3. 141
8.10 Concurrent session analysis module evaluation results - Scenario 4. 141
8.11 Scenario 1 - A good fit of the function f is found after only eight iterations. 144
8.12 Scenario 2 - After twelve iterations shown the minimiser has still not found a good fit, but is successful after 44 iterations. 145
8.13 After seven iterations the minimiser finds the incorrect minimum and quits. 146
8.14 The spammer call data correlates well with the model function as expected. 148
8.15 Correlation results for fourteen legitimate callers. 149
8.16 Average computation times for the hashcash puzzle. 150
8.17 Results of computational puzzle evaluation - Scenario 1. 152
8.18 Results of computational puzzle evaluation - Scenario 2. 152
8.19 Results of computational puzzle evaluation - Scenario 3. 152

8.20 Results of computational puzzle evaluation - Scenario 4. 152
8.21 Media traffic throughput at the Turing test server. 154

List of Tables

5.1 Robo-call experiment - interesting figures. 76

7.1 An example of the session record table stored by the CPA. . . . 113

8.1 IMS-level registration traffic overheads. 131
8.2 Session setup traffic overheads. 133
8.3 Session setup delay overheads for LAN. 133
8.4 Session setup delay overheads for WLAN. 133
8.5 Session setup delay overheads for HSDPA. 134
8.6 Initial parameter values for the session duration analysis module evaluation scenarios. 143
8.7 Results of call duration analysis module evaluation scenarios. . . 143
8.8 Final parameter values for a correct fitting. 143

Part I

Preliminaries

Chapter 1

Introduction

In the contemporary age of electronic communications end users are offered unheralded access to a wide range of voice, video and data services over a multitude of access technologies. The wide choice of digital communication mediums and user terminals presents several advantages to modern society, but most notably, ubiquitous access to the global network that millions of people rely upon for business, shopping, social networking and leisure - the Internet.

The Internet has revolutionised modern living with global inter-connectivity that provides a common-ground for humans to work, debate and interact with complete strangers in a virtual world. This virtual world gives its inhabitants the choice to reveal a great deal of personal information about themselves or to be completely anonymous. Unfortunately this scenario presents a serious security problem that threatens to erode the utility of ordinary Internet users and disrupt the correct functioning of the Internet's components. Hackers, fraudsters and marketers acting with complete impunity and anonymity have thus far invested a great deal of money and resources into polluting the Internet with advertising, financial scams and propaganda mostly with the goal of making short-term profit. The amount of undesirable content on the Internet will only increase in the future while there are further profits to be made.

Despite the dangers and inherent lack of security, there is an increasing trend from businesses and end-users alike to use the Internet for new applications and services, evidenced by the fact that over 1.2 billion adults worldwide use the Internet on a regular basis [32]. Additionally, as of December 2005 43% of Americans used the Internet for online banking, 27% downloaded movies and

songs, and over 40% used instant messaging services [26, 76, 45]. It is not surprising therefore, that the same architecture that allows users to send email, check their bank balances and experience audio and video entertainment, is now being adopted by end users to make voice and video calls and by the telecommunications sector to replace their ageing technologies.

1.1 VoIP Revolution

In legacy wired and wireless architectures voice calls are made over a circuit switched network. Large telecommunications giants have for many years made huge profits from circuit-switched networks by charging consumers for the amount of time that they make calls and the distance over which these calls must travel. In the past these corporations were protected legally and financially by their respective governments because it was seen as important to have a good communications infrastructure, and the high cost of deploying telephone networks made it unfeasible for many companies to enter the market. Recently, however, such networks have started losing favour with consumers who have begun to embrace VoIP (Voice over IP).

VoIP carries digitised voice over packet-switched networks such as the Internet. The voice is encoded using a low-bandwidth codec and transmitted over the user's regular Internet connection. There are several advantages to using VoIP in place of the telco-provided telephones but the over-riding benefit of VoIP is its low cost. Advances in technology have resulted in the marginal cost of data bits to plummet and it is not uncommon for both home and mobile users to have access to several megabits per second of bandwidth. Thus the relatively small bandwidth that a voice conversation consumes (approximately 16 kbps) is of little concern to the user. In essence if the consumer is already paying for Internet connectivity they might consider VoIP to be essentially free. This radical shift in the communications landscape has seen the emergence of several new companies that embrace this technology.

Having started in 2003, VoIP-provider Vonage currently provides 2.45 million subscriber lines. Most consumers find their service almost indistinguishable from their legacy circuit switched telephones except that they experience significant cost savings. Skype, recently purchased by online auction giant eBay, was also founded in 2003. This VoIP-provider has become a household name by offering free Skype-to-Skype calls and offering significant cost saving

to consumers calling legacy networks. This innovative business model accounts for the fact that currently 4% of worldwide long-distance calls are made over the Skype network [22].

Unfortunately there are several disadvantages to making calls over the Internet. Since IP is a best-effort protocol there are no guarantees that the voice packets will arrive reliably, in order, or in a reasonable time frame. Higher layer protocols can add reliability and re-order packets but cannot guarantee the bandwidth or delay of an IP connection. The variation in end-to-end delay, otherwise known as jitter, is also a major concern for real-time traffic. These factors can often result in poor voice quality and is the primary reason for the slow uptake of VoIP services by big business.

Furthermore, the lack of VoIP standardisation makes many VoIP terminals incompatible with each other, and the lack of security in the Internet leaves VoIP users susceptible to fraud, eves-dropping and identity theft. Telcos are addressing many of these drawbacks in their next-generation network architectures in order to protect their voice revenue and prevent themselves becoming low margin bit pipes.

1.2 Migration to All-IP and the NGN

The NGN (Next Generation Network) is a generic term that describes the current trend for network operators to move to an all-IP architecture that supports many types of services including the traditional telephony service. These networks should be accessible by a multitude of wireline and wireless QoS-enabled (Quality of Service) access networks and support generalised mobility, hence providing ubiquitous service access to the end user. One of the major goals of the NGN is to integrate the multitude of services available on the Internet, such as instant messaging, presence, file sharing, distance learning and presence, into a single unified architecture. Advocates of NGN claim that it will facilitate convergence between fixed and mobile networks to provide a common user experience, network architecture and billing infrastructure.

There are several benefits for both the telcos and their customers for adopting the NGN architecture. From the customers point of view the benefits include reduced cost, simplified billing, a range of new services and a greater choice of service provider. The telcos largest benefit from the NGN architecture is a reduction in both capital expenditure, due to owning a single network

with unified AAA (Authentication, Authorisation and Accounting) facilities, as well as a reduction in operational expenditure. The telcos also benefit from the reduced complexity, cost and time for developing new services, which will ultimately help them generate revenue and distinguish themselves from their competition.

Most large network operators have developed some kind of NGN strategy and some have even begun deployment. British Telecom (BT) have marked 2008 as the date when their network will contain only IP switches [81]. With their new network infrastructure they aim to offer broadband speeds of up to 24 Mbps to home customers. Other operators in the Netherlands, Bulgaria and Libya have also begun deployment of all-IP networks and it is likely that more will follow suit shortly.

While NGN architecture is based on large capacity core optical fibre networks, the best-effort nature of IP still presents QoS problems for network operators. In order to mitigate these problems the network must implement optimisations to the generic IP routing procedure such as DiffServ (Differentiated Services) and MPLS (Multi-Protocol Mabel Switching). These mechanisms handle different QoS classes with statistical guarantees of packet delivery - as opposed to absolute guarantees that are almost impossible to provide in IP networks. With DiffServ and MPLS in place the core IP networks are able to handle real-time services, but the access networks still provide challenges, especially for mobile operators.

Mobile network bodies have recognised the need for NGN and IP-based services. The second generation GSM (Global System for Mobile communications) standards includes provisioning of packet-switched data over GPRS (General Packet Radio Services), EDGE (Enhanced Data Rates for GSM Evolution) and UMTS (Universal Mobile Telecommunications Systems). These systems are offering increasingly higher speed packet access to the Internet, and run in parallel to their circuit switched narrow-band voice links. So while the the fixed and mobile networks provide converged data services, convergence has not yet been realised for real-time services such as voice. The solution that has been proposed to solve these problems and harmonise telephony with data communications is known as IP multimedia.

1.3 Birth of IP Multimedia

IP multimedia promises to offers a rich communication experience complete with real-time voice, video and text. Users will be able to transmit files, images and video clips to each other. This new communication experience has been driven by the desire of consumers to communicate in new ways yet still have the same quality, reliability and security that they have come to expect from the POTS (Plain Old Telephone Service).

Enterprises also stand to benefit from the new services offered by IP multimedia by integrating email, conferencing, push-to-talk, collaboration and remote support features into a single unified network. Travelling workers while away from the office will have access to all these features via secure and reliable channels. Moreover, the flexibility of the IP multimedia platform allows for enterprises to develop new solutions for their communications needs rapidly and inexpensively.

In order to provide the diverse services of IP multimedia, many different protocols that run over IP must be utilised and standardised to work in a predictable and repeatable manner. This ensures competing vendors products are interoperable and allows new players to enter the market. Two protocols that have gained widespread acceptance for setting up real-time communication sessions are the H.323 standard from the ITU (International Telecommunication Union) and SIP (Session Initiation Protocol) from the IETF (Internet Engineering Task Force). Both protocols run over IP networks and separate the signalling part of the call from the media. This design ensures that the signalling switches are not burdened with media traffic which generally constitutes the majority of the bandwidth required for a call. It also takes advantage of IP routing principles by allowing the media to take the shortest route to the destination, thus reducing media delays. It appears, despite the initial popularity of H.323, that SIP has become the protocol of choice for new IP multimedia applications.

The popularity of SIP can probably be attributed to its ability to traverse firewall and NAT (Network Address Translation) servers. Furthermore, SIP messages are human readable as they are text-based. SIP is also flexible enough to run over several different transport layers including UDP, TCP and ATM. As the number of applications delivered by IP multimedia has grown so has the SIP protocol. The most recent SIP standard as well as extensions to

the protocol means that not only can SIP set up and tear down all types of multimedia sessions but also provides event subscriptions and notifications, and supports terminal mobility.

The standards bodies responsible for the world's telecommunications needs have begun to recognise the power, flexibility and extensibility of Internet protocols, specifically those designed by the IETF. Consequently, the 3GPP (3rd Generation Partnership Project), as part of Release 5 of their standards introduced the IMS (IP Multimedia Subsystem) powered exclusively by Internet protocols such as SIP.

1.4 Standardisation of the 3GPP IMS

The 3GPP recognised the need to evolve their networks to an NGN architecture early on and in 2002 IMS first emerged as a small part of their monolithic standards. But these early standards did not discuss voice or any other service in detail; as such there was a large amount of confusion as to the realisation and interoperability of these services. Large corporations, such as Ericsson, took it upon themselves to define the IP multimedia telephony for inclusion in Release 7 of the standards and now the most recent standards define the minimum set of capabilities required for a multi-vendor and multi-operator telephony service.

Other standardisation bodies had a similar vision to the 3GPP for an all-IP network. 3GPP2, TISPAN and PacketCable have all adopted the 3GPP IMS into their own standards albeit in slightly modified forms. 3GPP2's version has been adapted to fit into their CDMA2000 MMD (Multimedia Domain), while the TISPAN version is destined for fixed networks. PacketCable focuses on the cable environment, and their PacketCable 2.0 specifications are based on Release 6 of the 3GPP IMS.

Another form of IMS currently being touted by Verizon Wireless and several of their vendors including Lucent, Cisco, Motorola, Nortel and Qualcomm is known as A-IMS (Advances to IMS). The A-IMS is designed to be an evolution of the IMS standards with the most notable enhancement being the support for both SIP and non-SIP session control. The companies behind A-IMS are hopeful that the 3GPP will in the future include A-IMS as an addendum to their current standards as there is a growing consensus that not all applications will become SIP-based anytime in the near future.

This thesis will focus only on the 3GPP version of the IMS standards due to the fact that all the IMS standards are similar in most aspects and that the 3GPP standards have thus far gathered the majority of industry acceptance. Where the term IMS is used the reader can assume that this refers to the 3GPP IMS. Still, most of the concepts and systems discussed in this body of work will be equally applicable to the other flavours of IMS.

1.4.1 IMS as a Service Delivery Platform

The provisioning of integrated, scalable and chargeable services is the main draw-card for operators considering implementing IMS solutions in their networks. Historically only large equipment vendors and operators had the knowledge and resources to develop multimedia services for the telephony environment. However, with the advent of Web 2.0 and the plethora of rich multimedia services available on the Internet, network operators are under increasing pressure to provide novel interactive services to their customers, often created by third-party developers who have only recently entered the market. Consequently the operator needs the assurance that these services can be incorporated into their networks in a safe and predictable manner. The operator also requires that these services be compatible with existing services from other vendors in order to create combinational services.

Combinational services refers to the principle that each basic service is simply a building block to a rich integrated experience for the customer. For example the voice-mail server may interact with the presence server to identify when the customer is available to receive voice-mail notifications via a text message. These basic services are known as *enablers* and form part of a the SOA (Service Oriented Architecture) principle. SOA services act as building blocks and can be added and removed from the network architecture quickly and robustly. These services must also provide a common charging interface so that the service provider can charge based on any number of variables, such as flat-rate, time-based, QoS-based or data usage. IMS does not dictate a particular business model, therefore it is up to the operator to choose how best to generate revenue from their services.

The 3GPP itself does not standardise the services that run over the basic core architecture apart from basic voice and video telephony and conferencing. Rather this task is left to other standardisation bodies that work with the

3GPP to provide the reference interfaces for vendors to create their own products. The primary standards body responsible for IMS services is the OMA (Open Mobile Alliance), a consortium of vendors, service providers and content providers. The general agreement between 3GPP and OMA is that the OMA generates requirements and the 3GPP extends the IMS to meet these requirements hence maintaining one single IMS architecture. Unfortunately there are several situations where the two standards bodies work in parallelled on services, including presence, availability and messaging, amongst others.

Regardless of where services originate, it is clear that the IMS is simply a base on which these revenue generating services will operate and as such the IMS can be considered to be a SDP (Service Delivery Platform). While the exact definition of an SDP is open for interpretation, in its generalised form it is a platform that provides control, creation, orchestration and execution tools for any number of services. In an SDP the control and service are separated logically and services will run on on their own AS (Application Server).

There are several service creation and execution environments available to the modern telecoms network engineer. JAVA, Parlay/OSA and JAINSLEE have all been touted as robust and scalable service creation environments yet it remains to be seen which will gather the most support from vendors. A common feature of these environments is the ability to provide generic resource adaptors in order to interact with several different networks simultaneously. As such a service might have adaptors for SIP, HTTP and even SS7 (Signalling System No. 7), in which case the service could be accessible from almost any terminal connected to the network. It will be shown in this work how this feature of modern service creation environments provides the power and flexibility required to solve some of the current and future challenges of the IMS.

1.4.2 IMS Challenges

It seems that network operators have no choice but to adapt their current monopolistic, circuit-switched business models, but it is still unclear whether or not IMS is the solution to counteract dwindling voice revenues and increased customer churn. There are several problems with the IMS architecture and standardisation process that may cripple the take-up of IMS both by network operators and their customers.

The first hurdle facing adopters of IMS technology is that IMS is only a reference architecture. The standards developed by the 3GPP do not provide clear-cut definitions of what an actual IMS implementation entails because a reference architecture only defines functional elements rather than physical network elements. This architecture provides a huge amount of flexibility to the technology and equipment vendors but also creates serious problems for the network operators who must somehow integrate the solutions from these different companies [73]. This situation causes the costs of product research, integration and testing to increase dramatically.

Moreover, network operators will initially struggle to find a business case for implementing IMS. Operators worldwide have already spent a considerable amount on their packet-switched hardware and will be still be looking to recover costs from these investments for years to come. Often the business cases are made on an application-by-application basis and it is hard to say at which point a full IMS implementation becomes more cost effective then the traditional stovepipe approach [47].

Apart from justifying the cost and complexity of installing IMS, many operators and consumers will be concerned that there may be no need for it at all. IMS places substantial emphasis on QoS management so that users will perceive good quality voice and video streams to whichever handset they happen to be using. But with recent advancements to both wireline and wireless connectivity access networks the need for QoS management may have been over emphasised [88].

One of the most pressing issues within IMS is the security that the architecture provides. It is well-known that the Internet has cost companies and individuals alike huge amounts in terms of attacks on machines and the resources required to prevent future attacks. The IMS employs comprehensive authentication, authorisation and encryption mechanisms to prevent attackers from compromising the network. But the issue of security is far more complex than simply preventing malicious attacks on the physical network elements.

In modern society the Internet is frequently used as a platform not only for communications but also for generating monetary profit. And wherever there are profits to be made there will also be nefarious individuals that seek to gain profit through questionable practises such as unsolicited marketing, spamming, fraud and phishing. The current transformation of the telecommunications landscape towards the Internet architecture now subjects telephony users to

these same threats [92]. If network operators hope to generate revenue from IMS services then they have a responsibility to protect them from security threats and to guarantee their privacy.

1.5 A History of Unsolicited Communications

Unsolicited electronic communications, otherwise known as spam[1], have been around since the 19th century when Western Union allowed telegraphic messages on its network to be sent to multiple destinations. These messages were mostly directed at wealthy investors that were subjected to questionable business offers. Over one hundred years later the problem of unsolicited communications has grown exponentially.

The first documented case of email spam was sent by Gary Thuerk over the network of government and university computers known as the ARPAnet [80]. His spam, sent to 600 ARPAnet members, publicised an open-house where his company's latest computers would be unveiled. Gary Thuerk's innovative marketing idea has now grown to a global problem, as recent studies now show that approximately 80% of all email worldwide can be classified as unsolicited bulk email [62]. This equates to roughly 2.5-billion items of spam every day of the year. Modern filters installed by ISPs (Internet Service Providers) are able to trap a large percentage of this spam but unfortunately the amount that gets through is still enough to cause outrage amongst their Internet users and lose them valuable customers. Some ISPs have estimated that they will lose as many as 5% of their customers because they can simply no longer tolerate the amount of junk email [80]. And while ISPs must deal with complaints and losing customers they must also employ significant human and technical resources into combating the problem.

A survey has shown that up to 29% of people have reduced their overall usage of email because of spam, and 63% have said that they are less trusting of email in general because of spam [60]. A full 77% said that the flood of spam made the act of being online unpleasant and annoying and 86% reported some level of distress after receiving spam [60]. These studies have shown that the

[1] The term *spam* comes from the canned meat product, parodied in a Monty Python skit, in which a certain cafe insists that every order come with a helping of spam. The despairing customer asks "Have you got anything without spam?", to which the waitress replies: "Well, there's spam, egg, sausage and spam. That's not got much spam in it."

spam problem is reducing the utility that users experience from the Internet.

Based on the above statistics it is clear that this problem should not be allowed to migrate to the world of NGN and IMS. The costs and inconvenience involved with trying to patch such problems retrospectively is simply untenable. However, questions remain whether the same problems that have plagued the Internet will manifest themselves into modern packet-switched telecommunications architectures. Network operators must consider whether or not this is a credible threat to their future deployment of IMS. This thesis attempts to shed light on this question by studying current consumer perceptions towards multimedia spam on their mobile terminals.

1.5.1 The Economies of Multimedia Spam

The NGN is based on the principle of high capacity fibre links providing users with a great deal of essentially unlimited bandwidth [87]. For the consumer this means that after an initial fee to the service provider their marginal cost per bit reduces to almost zero. This results in the consumer having access to a variety of rich, bandwidth-intensive services that they have never experienced before. Unfortunately this also opens up new avenues for spammers looking to take advantage of the economies of scale that must be maximised in order to make a profit.

Bulk emailers do not need to have a large response ratio in order to make a profit because the costs of spamming are so low. A study has shown that a response ratio as low as 1 per 100,000 emails sent is sufficient for spammers to recover their costs [95]. It has also been shown that as many as one third of users will click on a link in cleverly disguised spam [71]. So while filters are able to block a large amount of spam, the few that are delivered certainly pay dividends to the spammers.

Moreover, there is no significant increase in the cost of SIP spam over email spam [65]. A SIP call spam application can be easily created even by a non-expert programmer. A SIP application is able to generate several concurrent calls, play a pre-recorded message and then terminate the call. A benefit to SIP spam over traditional telemarketing is that the spamming application can be realised entirely in software and run from modest desktop computers. It is estimated that SIP spam only costs in the order of 400 US micro-cents per call and is thus approximately three orders of magnitude cheaper than traditional

telephony [65]. With the ever-decreasing costs of Internet bandwidth and computing hardware it is likely that the marginal cost per SIP spam will drop further.

Email spammers often make use of zombie machines that have been infected with viruses in order to spread messages very rapidly. The cost of sending the spam is passed on to the owner of the infected machine, and the process of blocking the messages becomes all the more complicated as their origin changes frequently. The use of infected machines will also be attractive to SIP spammers and poses a risk to NGN user equipment.

The most important factor when analysing the economics of spam is the issue of proximity. Currently there is very little in the way of telemarketing over international boundaries, for the simple reason that international calls are expensive. Sending an international email, however, is no more expensive than sending locally; and the same applies to SIP spam. This presents a significant problem, especially to those countries were the prevalence of multimedia spam has been low and the inhabitants are less suspicious of scams and fraud. Countries that are notorious for spamming, such as the USA, China, South Korea and Russia, will suddenly have a far larger audience at the same cost [19]. Since it is very difficult to prosecute individuals across international boundaries legislation will have little effect on this type of spam [71].

1.5.2 Address Harvesting

Spammers require a large address database of intended recipients and these can be gathered either by the spammers themselves or list merchants that sell these databases for substantial profits. The most popular method for obtaining addresses is known as address harvesting. This refers to the process of trawling through the world wide web, forums, Usenets and newsgroups searching for the syntax of a valid email address. Corporate staff directories are a prime target for address harvesters. It is also possible to use a domain's DNS and WHOIS records to gather the email addresses of the administrators. These addresses are obtained without consent and without regard for the privacy preferences of their owners. Many of the addresses harvested are malformed, expired, invalid or undeliverable but this is of no concern to the spammer who might send millions of messages per day.

Another option for collecting addresses is via viruses and trojans residing in

infected computers. The computer files are scanned for email addresses that are not necessarily available on the Internet and the addresses are reported back to a repository. An infected machine may also scan the traffic of nearby machines in an attempt to harvest even more addresses. Unscrupulous businesses may make a side profit from selling their customer databases. These addresses are particularly valuable as they are known to work and the recipient is known to be interested in whatever product or service that the business deals in. Therefore even users who are careful not to divulge their address in public places might still find themselves on an address merchant's list.

Courteous bulk emailers provide unsubscribe links at the bottom of the message offering to remove the recipient from their records. However, often when a user visits an unsubscribe web link the spammer will not delete their address of their database but rather consider it far more valuable, as it is now known that the recipient's email is valid and in active use.

The US FTC (Federal Trade Commission) has performed several experiments in order to determine spammers abilities to harvest addresses. In one particular experiment, a newly created AOL email address was used in a religious chat room to post a message. The account received its first spam message only 21 minutes later. In a second experiment an address was hidden on the agency's home page and within seven months the address had received 5,150 spam messages [80].

Unfortunately SIP addresses are just as easy to harvest as email addresses. The format of a SIP URI (Uniform Resource Identifier) is *sip:john.doe@operator.com*. These addresses are very similar to email addresses syntactically and will no doubt be posted on websites, web forums and other public locations, to be harvested by list merchants and spammers alike [65]. In addition, the ENUM system is considered to be a virtual gold mine for SIP address harvesters. ENUM is the Internet system that resolves fully qualified telephone numbers to fully qualified domain names using a DNS-based architecture [24]. A spammer can gain access to millions of URIs by simply traversing the e164.arpa tree and mining it for data.

1.6 The Manifestation of Spam into the NGN

Network security experts need to envisage how spam will make the transition from email to SIP and to other forms of multimedia that make up the NGN

landscape. Telemarketing is the one form of non-email spam that most people have already experienced. While telemarketing is highly frustrating and problematic, it simply cannot compare to the volume of spam that can be generated by automated spammers. Telemarketers are limited by the costs of human resources and the time it takes to make individual calls whereas automated applications can set up many simultaneous sessions at very low cost. Indeed, inventive spammers have already devised several inventive types of multimedia spam.

1.6.1 Multimedia Spam Types

The practise of initiating numerous bulk voice calls over a packet-switched architecture is known as SPIT (Spam over IP Telephony) [65]. These calls can be of several different media types, including voice, video or even data files streamed over the media plane of the session. SPIT relies on the recipient either answering the call or some proxy answering the call on behalf of the recipient, such as a voice-mail server. Many consider SPIT to be the most intrusive type of spam because it requires the immediate attention of the recipient [92]. If these calls happen after hours or during some other inconvenient time then they can be particularly frustrating. It is also very hard to quickly identify whether or not a call is spam as there is a certain period of time at the beginning of a call when an unknown caller would typically introduce themselves. Email users nowadays have become particularly savvy at detecting and deleting email spam very quickly but this is not possible for a voice or video call, or when the content must be downloaded before it can be inspected.

Unsolicited multimedia using instant text messaging is known as SPIM (Spam over Instant Messaging). SPIM is a very attractive option for the spammer as they achieve much higher click-through rates, several orders higher than achieved using email [13]. SPIM saw massive growth around the year 2000 but has tailed off in recent years due to the policy of modern IM clients to allow only messages from individuals on the user's whitelist, otherwise known as a buddy-list. While the buddy-list has curtailed the SPIM problem it has obvious drawbacks in the telecommunications environment where traditionally any person can call or send a message to anyone else. IM can be sent using either pager-mode or session-mode. Pager-mode IM works in a similar asynchronous fashion to email where the message is sent immediately and there is no distinc-

tion between the signalling and the content. Session-mode IM requires both communicating parties to establish a session, thereafter the content is send over a separate media channel. While it is possible for the network elements to apply email filtering methods to pager-mode IM this is not the case for session-mode IM as the media may flow over a distinct and unknown path from the signalling.

The buddy-list system requires that some kind of consent be given to add a new buddy to the list. Presence spam refers to the practise of generating bulk buddy-list requests with some kind of additional information attached. In most IM clients buddy-list requests pop up immediately with the details of the request. The spammer may supply an advertisement and link to their site in place of their name and URL. Therefore while the buddy request will no doubt be denied, the spammer has still succeeded in displaying a message to the recipient. The problem of how to meet someone for the first time and decide whether or not to place the person on the whitelist is known as the "Introduction Problem" [65].

1.6.2 Spam Threats in the IMS / NGN

There are several categories of spam that have been identified, the most innocuous of which is called UCE (Unsolicited Commercial Email). UCE refers to the spam that promotes a commercial service or product usually with a link to a website where the product may be purchased. A common form of this spam is the sale of men's health drugs, such as Cialis and Viagra. If the customer receives anything then it will most likely be an unapproved generic drug. UCE often contains images in order to avoid text-based spam filters, and clogs email servers even further. Alternatively, spam filters can also be by-passed if the words in the email are obfuscated, for example the word Viagra can be written as V1agra. Modern spam filters specifically look for word obfuscation as it is a clear sign of junk email. UCE has also been used extensively on the Internet to advertise adult websites. These messages may contain explicit images and have largely been minimised of late thanks to tough legislation [60]. IMS SIP messages can contain embedded MIME (Multipurpose Internet Mail Extensions) that can contain text, images or short video clips. It is highly likely that this is the manner spammers will target NGN users for UCE.

Some spammers have no intention of selling a product or service but rather

LIVE FROM THE STREET!
Sym: (PRTH)
Price: .085 (UP 15%)
Announces the Opening of Two New Stores by
(PINKSHEETS: PRTH) is pleased to announce that Puerto Rico 7, Inc. has opened two new stores. The stores are recorded as Pinero II and Borinquen Towers. Both locations were researched demographically to deliver above average sales due to high traffic streets and communities directly surrounding the stores. The Management team believes that the stores will each quickly reach an annualized run rate of 1.2 Million dollars of sales.
IMAGINE IF YOU HAD THE CHANCE TO BUY A WAL-MART FRANCHISE IN MEXICO RIGHT WHEN IT FIRST OPENED ITS DOORS THERE AND ALL YOU NEEDED WAS A SMALL STAKE TO GET IN.
Hurry, we see this stock starting to make the turn NOW.
Big watch in effect for August 8, 2007!!!!!

Figure 1.1: An example of a stock touting email. This message was flooded over the Internet as a PDF attachment on the 8th August 2007. The stock in question trades for fractions of a cent.

just want to express a message to the recipient. These include religious, political and other propagandist messages. Twelve percent of registered voters in the 2006 US mid-term election received unsolicited email from political parties - probably due to the fact that political parties are exempted from spamming laws in the US [59]. Pre-recorded voice calls, also known as robo-calls, are ideal for spreading this type of message. The advent of IP-based telephony allows robo-callers to make many more simultaneous calls than is possible with today's PSTN (Public Switched Telephone Network), thus reaching a far larger audience.

A problem currently faced by micro-cap companies is the touting of shares via spam commonly known as the *Pump and Dump* scam. These micro-cap companies generally have a market capitalisation of less than $250m and do not meet the requirements to be traded on a stock exchange. These companies may trade via OTC Bulletin boards or the Pink Sheets Electronic Quotation service where real-time share prices and detailed company information are not readily available. The premise of stock-touting messages are that some particular stock is about to increase in value and that the wise investor should buy the stock immediately and sell once the stock has reached it projected value. The originator of the share-touting messages will have bought a significant number of these shares in advance. Once the touting messages are distributed these shares usually do increase in value very quickly, not because the respective company has done anything to affect its share price, but because so many people take this unsolicited advice. Once the shares reach a certain level the spammer and many others will sell, reducing the stock price back to its previous level or even lower. Research has shown that significant profits can be made this way for the spammer - profits that are funded from hundreds of gullible

investors [28]. Incredibly it has been found that many investors realise that they are buying into a pump and dump scam yet they feel that they might still make profit if they enter early enough [28]. There is a recent trend to send stock-touting messages as PDF, Excel or ZIP email attachments in order to bypass anti-spam software [29]. In the NGN context pump and dump scams can be perpetuated in several different ways using text, images or even short video clips.

Spam is also used as a tool to defraud Internet users. The most prevalent type of spam fraud is known as advance-fee fraud or the 419 scam. The number 419 is derived from the article of the Nigerian Criminal Code dealing with fraud as this type of fraud was made popular by Nigerians in the 1980s in order to defraud western investors seeking to tap into Nigeria's oil wealth. There are many variations on this type of fraud but in essence the scammer tries to illicit a small payment from a target with the promise of a large sum of money later on. For example the target may be informed that they have won a lottery and that a processing fee must be paid in order to receive the winnings of several million dollars. A variation on this theme is when a target is informed that a wealthy African leader has died and help is required to transfer their money out of the country. The target must provide some capital to facilitate the transfer, perhaps as a bribe to the local authorities, and in exchange the victim is offered a percentage of the considerable wealth. Research has shown that victims of advance fee fraud will lose on average 5000 US dollars, and in excess of one billion dollars has been lost by individuals and businesses lured by this scam in the last decade [35]. This 30-year old scam will no doubt proliferate throughout the NGN environment in several different forms and claim many more victims for years to come.

Fraudsters have also found inventive ways to discover sensitive information about their victims by posing as a trusted company. The act of criminally acquiring user names, passwords or credit card details is known as *phishing*, a play on the word *fishing*. Phishing emails typically request that the recipient visit a spoofed website and enter the details of their financial accounts. Upon acquiring these details the phisher will withdraw funds from the victim's account or use their credit card for online purchases. Financial institutions with online banking facilities and online payment systems such as PayPal are commonly impersonated in phishing attacks. Current estimates show that on a given day as many as 250,000 phishing attempts will be made against a given institu-

tion [7]. Phishing results in monetary loss for consumers, and soft monetary loss for the impersonated company in terms of brand erosion and undermined consumer trust. Phishing attacks over the years have evolved from telephone calls to mass emailing campaigns and will no doubt make the transition to packet-switched telephony networks.

The above threats are only a small subset of the huge problem that unsolicited email causes in the Internet. Inventive marketers and scammers will continue to abuse communications systems as long as it is profitable to do so. There are several benefits brought about by the rapidly decreasing telecommunications costs and the introduction of rich multimedia services. Friends, family and colleagues can experience a level of communication never before imagined, and business can operate more effectively. Unfortunately this also brings about an increase in unsolicited marketing, propaganda, scams and fraud. These attacks threaten to undermine the utility of new telecommunication services and cripple the uptake by consumers who will not tolerate the associated security risks.

1.7 Thesis Objectives

There are many issues that must be resolved before IMS can be deployed on a significant scale. Possibly the most pressing of these issues is the inherent security risks involved when adopting an architecture that mimics the Internet, as the Internet is rife with marketers, viruses, trojans and fraudsters. Not only will IMS terminals be at risk from attacks but the privacy of IMS users may also be jeopardised. If customers find that they are constantly bombarded with unsolicited voice calls and multimedia messages then they will no doubt lose faith in the technology and find alternative communication mechanisms.

This thesis has several objectives. Firstly it is imperative to perform a thorough analysis of the potential risks that face IMS networks and end users. This can be achieved by performing a review of the attacks that have been conducted throughout the last few decades via telecommunications networks.

Secondly it is important to perform a temperature check on consumer perceptions towards spam. If consumers are not worried about spam or if they feel that spam is a necessary evil in order to have access to the rich communications services that they desire then perhaps telecommunications companies can spend less time and resources in solving the problem. If, however, consumers

are outraged by spam and the potential breach of their privacy then clearly network providers and government will have to react quickly and preemptively to solve the problem, or risk alienating customers and losing revenue.

Thirdly it is important to identify how NGN networks such as the IMS can use technological solutions in order to combat the spam problem. Spammers are notoriously inventive and have thus far been able to circumvent practically all spam prevention methods to date. The question remains whether the NGN architecture can protect itself from attackers or whether this architecture is doomed to failure because of its security shortcomings.

The above three objectives can be achieved using a several scientific techniques. In order to analyse potential security risks to the NGN several case studies relating to multimedia spam must be discussed. These case studies can be applied to the NGN model to show whether or not these risks can be realised in the future particularly in IMS networks. When it comes to network security it is imperative to analyse past attacks as most new attacks are simply variations on an existing theme.

In order to analyse current consumer perceptions toward spam a survey is conducted amongst users of current generation mobile devices. This survey aims to ascertain how tolerant users are at the moment towards unsolicited voice calls and text messages and what level of intrusion they will accept in the future. The objective is to gather meaningful user data that can be used both in this work as well as in future work in this field of study.

A standards-compliant spam prevention architecture is proposed based on the above data and an analysis of the current state the art in anti-spam security measures. The primary objective of this thesis is to implement such a solution in an emulated IMS network in order to show proof-of-concept, and to perform a thorough evaluation of its reliability in a practical setting.

1.8 Thesis Scope and Limitations

There are currently several emerging packet-switched multimedia technologies trying to gain market share. Internet telephony companies, such as Skype and Vonage, use proprietary technologies in their networks. It is important that these companies examine the issue of spam prevention and privacy in their networks, however, no single solution will fit all network architectures. In this work only the 3GPP IMS architecture will be examined although it may well

be possible to extend the solutions proposed in the work to other network architectures. At present no network operator has yet adopted a fully IMS-compliant architecture, thus while the threat of IMS spam is probable there is no data to indicate the scale of the problem. The risks that are being examined in this thesis are based on an extensive review of literature and previous case studies but are still largely hypothetical.

The problem of spam and privacy is inherently a security issue. A major setback with security solutions is that it is impossible to predict future vulnerabilities, and the same is true for spam. History has shown that new technologies are often best exploited by criminals, as shown by the current high levels of email abuse, banking fraud and identity theft. Spammers and fraudsters will also always find inventive solutions for bypassing security measures; a prime example is that of Bayesian filtering. Bayesian filters use sophisticated statistical techniques to classify messages based on the words in the email. Soon after these filters became popular spammers began inserting random, legitimate words into the body of their messages in a process known as *Bayesian Poisoning*. The words inserted are harmless and have no relevance at all to the advertised product or the fraud being perpetuated, leading the filter to either accept the message as legitimate or, even worse, identify the email as spam but then train the filter with these harmless words, thereby increasing the likelihood of future false positives [51]. In scenarios like these, security experts and attackers are constantly playing a cat and mouse game. The thesis does not aim to predict future spamming behaviour but rather protect against the popular and effective attacks that currently exist.

At present there is a distinct lack of data as to how multimedia spam affects telecommunications networks and customers and consequently it is hard to know exactly what level of response to the threat is required. The thesis attempts to address this lack of data by performing two data collection exercises. Firstly by analysing the reaction of end-users to receiving an unsolicited multimedia, and secondly by performing a survey of current perceptions on multimedia spam. However, both these data collection exercises are limited both by the amount of resources available to tackle the problem and the time frame required to do a thorough analysis. Thus any results obtained cannot be considered as incontrovertible but rather they act as a guide to solving this previously unexplored area of research.

Another limitation exists in the relatively young age of the IMS specifi-

cations. The 3GPP have listed *Protection against SMS and MMS spam* as a future study item, possibly for first inclusion in Release 8 of their specifications. At present, however, no solutions have been proposed and the scope of this study item does not include other types of multimedia spam such as voice and video. The thesis aims to include in its scope several types of unsolicited multimedia, including real-time voice and video.

The thesis does not provide detailed descriptions of SIP or the 3GPP IMS. Throughout the text it is assumed that the reader has a good knowledge of SIP terminology, syntax and transactions. It is also assumed that the reader is familiar with the architecture and signalling of the 3GPP IMS. Readers that are not familiar with either of these specifications are referred to the SIP RFC [68] or an appropriate IMS text [17, 53]. A working knowledge of HTTP and XML is also recommended but not essential.

1.9 Thesis Outline

The document is divided into three parts. Part I introduces the history behind and the relationship between spam and the IMS. It then discusses the current state of the art in spam prevention mechanisms and examines several intriguing case studies. Part II describes the collection and analysis of spam-related data through experimentation and consumer surveys. Part III proposes solutions to the previously-identified problems using a multi-layered spam prevention architecture and analytical models. These analytical models are then realised in a practical setting and subjected to feasibility and performance evaluations. The chapter layout is detailed below.

In Chapter 2 a thorough literature review is conducted as to the current state of the art in both text-based and real-time spam prevention. Special focus is given to literature dealing with SIP spam that has obvious implications for the SIP-based 3GPP IMS. As such much of the chapter is dedicated to recent literature emanating from the IETF.

Chapter 3 explores pertinent case studies of past and current spam attacks in order to evaluate the danger of these attacks being perpetrated against future next generation networks.

Part II starts with Chapter 4. This chapter details a survey that was performed that gauges the current levels of multimedia spam in South Africa and evaluates consumer perceptions towards spam.

Chapter 5 presents The Robocall Experiment in which a typical spamming scenario was recreated in order to evaluate the behaviour of individuals when subjected to automated voice calls. The experiment also measures the participant response ratio to automated calling.

Chapter 6 begins Part III of the thesis and proposes a novel multi-layered architecture for prevention of unsolicited sessions and for user-defined privacy control. The architecture introduces two network elements: the Call Pattern Analyser and the Authorisation Engine that work together to reduce spam and enforce user privacy rules.

Chapter 7 describes the implementation of the proposed architecture in a practical standards-compliant network test-bed. The design and implementation of the test-bed is performed in such a manner that future researchers will be able to accurately reconstruct it for their own experiments.

In Chapter 8 the proposed architecture is subjected to several evaluations both through analytical assessment as well as practical validation in the test-bed framework. The evaluations demonstrate both proof-of-concept as well as highlight the effectiveness of several spam prevention techniques in an IMS environment.

Chapter 9 concludes the thesis and provides recommendations for further research in this field.

Chapter 2

Literature Review

The current state of multimedia spam research is relatively immature. Still, there have been advances in this field and several proposed solutions for preventing unsolicited sessions over IP telephony. The chapter investigates the state of the art in detection and prevention of voice, video and text-based spam in IP networks. There has been little discussion or research specifically dealing with the NGN, or indeed the IMS architecture, but many of the solutions proposed for general Internet telephony can be applied to these environments. This is in part due to the fact that the IMS is based entirely on Internet protocols, specifically SIP.

Internet spam solutions can be applied to the IMS landscape because the IMS provides intrinsically better security than the Internet due to its closed-wall design. The opposite is not true however. For example an IMS spam solution may rely on the strong authentication and authorisation mechanisms provided for by the IMS's authentication and key agreement protocols and physical sim-cards, whereas these mechanisms are not available for the Internet. Address spoofing and message source obfuscation have long been security problems in the Internet whereas the IMS has for the most part mitigated these risks [92]. Therefore the IMS spam problem may well be solved with the combination of the IMS security measures and the work of those who have tried to solve spam in the Internet.

The chapter begins with a review of literature dealing with the relationship between spam and SIP. It then discusses pertinent work on the detection of robo-callers through the methods of Turing tests and detecting irregular calling patterns. Relevant literature surrounding text-based spam filters is also

reviewed and the chapter concludes with an overview of the 3GPP current work schedule regarding spam in the IMS.

2.1 IETF Spam Prevention Literature

In July 2005 Rosenberg *et al.* first identified the problem of bulk unsolicited communications in SIP [67]. This document has led to several other IETF Internet drafts that deal with the problem of spam and SIP. Internet drafts are temporary documents that expire six months after they are issued. They can be changed at any time and do not take into account backwards compatibility with existing implementations. Only once these drafts reach an acceptable level of maturity are they published as an IETF RFC (Request for Comments). Due to the immaturity of research in this area, and SIP itself, Internet drafts are the only evidence thus far of IETF's progress in solving SIP spam. It must be noted however, that Internet drafts are considered to be works in progress, and may be modified or revoked at any time. Therefore the interested reader is encouraged to examine the latest version of these documents, which may have changed since the time of this research.

2.1.1 The Session Initiation Protocol and Spam

Rosenberg and Jennings's RFC *The Session Initiation Protocol (SIP) and Spam* [66] was released in January 2008. The document performs a thorough analysis of the similarities and differences between email and SIP spam, and the applicability of email spam solutions to SIP.

The introduction states categorically that the email spam plague would be much less significant had suitable solutions been deployed ubiquitously before the problem became widespread. Therefore, the problem of SIP spam must be dealt with as soon as possible. The work seeks to aid this process by analysing the applicability of email-space solutions, identifying applicable SIP mechanisms that can help the problem, and making recommendations to SIP implementers.

Three types of spam are identified as potential targets for the SIP spammer: call spam, instant messaging spam, and presence spam. It is noted that the costs of call spam are significantly lower than regular telemarketing because a SIP call costs only fractions of a cent in terms of network bandwidth

utilised. It is also expected that a bulk caller will be able to sustain as many as ten concurrent calls even on modest hardware. The value per spam message is expected to be significantly higher for a voice call than for email as the content is more likely to be examined by the average user, but the value of instant messaging spam is questioned due to the extensive use of whitelists in modern instant messaging implementations. Still instant message spam is seen as a problem due to flaws in the whitelisting solution. Presence spam is highlighted as a particularly dangerous threat due to the nature of the SIP consent framework. Presence requests are conveyed to the user often in the form of a pop-up box, thus presenting an opportunity to relay information to the user in the requester identity. For example the requester URI could be *sip:buy_viagra@drugstore.com*. A short message has been displayed to the user without consent and hence the marketer has been successful.

Several options are identified as possible solutions to the above problems. The first of which is content filtering; currently one of the most popular forms of email spam protection. Such filters inspect the content of the message for possible clues that it is spam. Content filters are a viable and probable protection mechanism against instant message spam due to the similarity with email spam. However, content filters are effectively useless against call spam for two compelling reasons. First, a voice call must be answered before any media will flow, making it impossible to do any kind of pre-inspection to determine if the content is legitimate. Second, if the call is recorded for later playback, say for example on a voice-mail server, the message would most likely be obfuscated sufficiently such that current voice and video recognition software would be unable to pick out any identifying words. This obfuscation would most likely be in the form of background noise, varied accents and incorrect grammar. The recognition of video is an even more challenging problem and hence at this time content filtering is not suitable for real-time or pre-recorded voice or video calls.

Blacklisting is proposed as an option to completely block those users identified as spammers. Blacklists have had limited success in blocking email spam because email address spoofing is very easy. Email addresses are also easy to come by due to the large number of free email providers, allowing blocked users to simply discard the account and open a new one. Similarly, there are currently several SIP providers that provide virtually unlimited URIs free of charge. It is generally unacceptable for network administrators to block an

entire domain for fear of blocking thousands of legitimate users. When SIP was first introduced it was also susceptible to address spoofing but recently new security measures introduced to the protocol allow for the sender of the message to be authenticated [52]. Therefore, so long as the spammer does not have an unlimited supply of addresses, blacklisting may be an effective tool in solving this problem. Blacklists may also be used in conjunction with a whitelist.

Whitelists are a list of URIs that have been previously authenticated and from whom the user is willing to receive messages. Throw-away usernames are of no benefit to the spammer when whitelists are in use but address spoofing remains a concern. However, whitelists are a very effective tool to prevent spam when sender authentication is employed. A problem that greatly limits the whitelist solution is that messages cannot be received from an otherwise legitimate user that has not explicitly been added to the contact list. The problem of how to meet someone for the first time and add them to the whitelist is an example of the Introduction Problem. While some techniques, discussed later in the chapter, exist to help solve the introduction problem, significant work is still required in this area to find an effective solution.

A consent-based solution is also proposed as a mechanism to stop unknown users from initiating SIP sessions. When an unknown caller first tries to setup a session it is rejected with the message that consent is being sought. The intended recipient of the session is informed that a call was attempted and asked whether or not future communications should be allowed. This solution again suffers from the lack of an effective consent framework whereby the spammer may still be able to communicate a short marketing message in their consent request. A content filter may be applied to these consent requests to try and avoid abuse of the system. The problem with consent-based solutions is that users may become frustrated responding to consent requests before communication can take place whereas a reputation-based solution allows trusted users to make calls without hindrance.

Reputation systems require that all users be assigned a reputation score, and users with negative reputations are not allowed to initiate sessions. Usually these reputations are gathered by a centralised system based on the feedback received from fellow users. This feedback can either be in the form of tattling, where bad behaviour is reported, or praising, where legitimate users are assigned higher reputation by their peers. The drawback of this system is that

a user with a poor reputation can simply discard their account. New users must be given the benefit of the doubt as to their reputation or they will never have an opportunity to prove themselves. Moreover, a collective of spammers can cooperate in order to increase each others reputation scores. An interesting variation of the reputation system is to consider the existence of users on a buddy list as a sign that they are trustworthy. Reputation can be inferred through a social network of reputation where the closeness of the relationship is taken into account. For example, users that are one buddy removed are given higher reputation and five buddies removed a lower reputation. Spammers cannot influence their reputation by collusion unless they somehow become part of the recipient's social network.

An example of the reputation solution is given in the paper *SIP Spam Detection* [61] by Rebahi *et al.* The paper proposes a framework for a reputation-based spam avoidance technique that determines the trustworthiness of callers. In this framework every user must rate how much they trust the users on their contact lists. These trust values are used to create a virtual weighted directional graph of users with the trust values determining the weight of the edges. An average reputation score is calculated from the source's neighbours. Requests are rejected from sources whose average reputation scores are below the predefined threshold.

Rosenberg and Jennings also propose the practise of address obfuscation in order to prevent address harvesting from websites. This entails manipulating the address into a form readable by humans but difficult for automata to identify as an address. For example, the user may list their address as *joe at domain dot com,* which would not be identified by an automata as a valid email address but easily decoded by a human. The address may also be shown as an image since current address harvesters only mine the text of websites for addresses. Unfortunately this solution does not solve the problem of spammers traversing the ENUM system for addresses. However, this problem may be mitigated if only number prefixes appear in the DNS, as opposed to the actual numbers.

The adoption of limited-use addresses will further protect the user so that if the address does fall into the wrong hands then it can be terminated. Limited-use addresses can be set to expire once spam is received at that address or after a fixed time period. The user must have a large number of addresses at their disposal in order to take full advantage of this technique. The drawback

of this technique is that all contacts that have been given the limited-use address will need to be informed of the new address once the previous address expires. The use of a presence service has been suggested as a solution to this problem. When a buddy requests the users presence state they will also be notified of the current limited-use contact address. Furthermore, every contact can be given a different single-use address [63]. This is possible due to the structure of the SIP protocol that allows tags to be appended to URIs that are ignored by intermediate proxies. For example, according the the SIP RFC, *sip:family@domain.com;member=john* is a valid SIP address.

Turing tests provide a mechanism to prevent automated callers from initiating sessions. The Turing test is a challenge or puzzle that can only be answered by a human. Turing tests, otherwise known as CAPTCHAS (Completely Automated Public Turing test to tell Computers and Humans Apart), are frequently used in the Internet to stop bots from signing up email accounts and posting spam on message boards. The most popular tests consist of a series of alphanumeric characters that are distorted and added to background noise. The user is required to identify the characters in order to gain access to the resource. This solution can be extended to voice systems where the user is presented with a voice recording distorted with background noise and asked to type in characters on the keypad. The SIP *Accept-Language* header can be used to determine the callers preferred language choice for the test. CAPTCHAs must be suitably complex due to advances in image recognition software and artificial intelligence algorithms, thus they can present problems to handicapped users. Furthermore, it is possible to employ cheap workers to take the tests in what are known as click sweatshops, but this does reduce the spammers profit margins significantly.

Another solution to prevent automated spamming is to request the caller perform some type of computational puzzle and return the result. The premise is that the puzzle will require an amount of resources that makes it prohibitively expensive to initiate bulk calls or messages, but not so as to burden legitimate users. A problem lies in the varying processing power available to different communications devices. A personal computer will be able to solve a CPU-intensive puzzle an order of magnitude faster than a mobile device. However, it has been found that the CPU memory bandwidth on most electronic communications devices is more or less similar, so a good puzzle should take this into account. This technique may be a viable solution for solving the

introduction problem as the caller has proved that they are willing to expend significant resources in order to complete the call.

Alternatively, the caller may be requested to make a micro-payment before the session is accepted. If the call is legitimate then the payment is refunded, otherwise the call recipient may keep the payment. It has been noted that this technique has several flaws. There may be significant costs involved for both sender and receiver to use a worldwide micro-payment framework since such payments usually attract fees of up to 15% of the transaction. Moreover, poor, selfish or vindictive users may choose to refuse the call and keep the payment of legitimate callers. The system may also be prohibitively expensive for users in countries with weak currencies.

Some countries may choose to pass laws that prohibit unsolicited communications. An example is the US "Do not call" list that prohibits companies from contacting users that have opted-out from receiving telemarketing. This registry has had remarkable success as a survey has shown that of those that signed up 92% reported receiving fewer telemarketing calls and 25% say they received none at all [18]. Unfortunately the source of spam is often from other countries that fall outside of the recipients legal jurisdiction so this solution may be limited somewhat.

The last solution that is proposed is to form a network of trusted providers. In this circle of trust every domain is responsible for preventing its users from spamming and is responsible for punishing transgressors. It is suggested that the providers agree to charge their customers a pre-defined amount and hence provide disincentives for spamming. This technique is not favoured as it relies on much of the rigidity and monopolistic traits that are currently evident in the PSTN.

Rosenberg and Jennings then go on to discuss methods by which the senders of SIP messages can be authenticated. SIP provides two reliable mechanisms for verifying the sender of a message. The first mechanism proposed by Peterson and Jennings [52] relies on two new header fields to be inserted into the message. The *Identity* header holds a signature that validates the sender's identity, and *Identity-Info* provides a reference to the signers certificate. The client must connect to their domain over a TLS (Transport Layer Security) connection in order to ensure that the domain is authenticated. The client itself is authenticated using some kind of digest exchange over the link; most likely using a shared secret such as a password. When a message exits the

client's domain, the identity of the sender is asserted and a signature is added to the Identity field to validate the assertion. The companion header field Identity-Info tells both the receiver's domain and the receiver's user agent where to find the certificate of the originating domain. Thus the recipient can verify that the originating domain has authenticated the user and permitted the user to use the address in the *From* header.

The second method of authenticating users is by using the *P-Asserted-Identity* SIP header [40]. This extension to SIP allows a network of trusted SIP domains to assert the identity of their end-users. There must be mutual trust between the domains because there is no mechanism to verify the authenticity of the asserted identity once it leaves the domain. The nodes within a domain are implicitly trusted by the users and end-systems within the domain to publicly assert the identity of party. Moreover, the nodes are also trusted to withhold the identity from other domains when privacy is requested. The P-Asserted-Identity is a far weaker solution than the Identity header but unfortunately it is the mechanism adopted by the 3GPP in the IMS.

The drawback of the P-Asserted-Identity is that it becomes exponentially harder to ensure mutual trust between providers as the number of interconnected domains grows. If a domain decides not to enforce the P-Asserted-Identity correctly, either because of a configuration error or a disinterest in protecting users from spam, then every network connected to them is under threat. The effectiveness of the P-Asserted-Identity is only as good as the security policies of the weakest provider. This is because every domain in the message path trusts that all previous domains followed the specifications correctly. When using the Identity header the recipient domain always has access to the certificate of the original signing authority regardless of how many intermediate networks were traversed. The use of the P-Asserted-Header is a security risk for large IMS providers interconnected throughout the world. It will be very hard to trace the origin of any unsolicited IMS sessions that originate from providers who are sympathetic to spammers.

2.1.2 Computational Puzzles for Spam Reduction in SIP

A hashcash is a proof of work supplied by the originator of a SIP session. The hashcash is used to prove that the sender of a SIP request has performed a modest amount of computing time calculating a textual stamp that is then

```
Puzzle: work=10
; pre="XPokF1n0+NG6iwRcYzeXuETrtDo="
; image="XPokF1n0+NG6iwRcYzeXuETrtDo="
; value=160
```

Figure 2.1: SIP Puzzle header proposed by Jennings - the image and pre-image are base64 encoded.

added to a header in the SIP request. Jennings discusses how the hashcash can be incorporated into the SIP architecture in his draft *Computational Puzzles for SPAM Reduction in SIP* [39]. The aim is to find a method that can allow previously unknown legitimate callers to be able to send messages while making it prohibitively expensive for spammers to initiate bulk SIP requests.

The hashcash mechanism dictates that the receiver may reply to a SIP request with a 419 response[1] including details of a puzzle that must be solved before the request can be resubmitted. The challenge is included in a new SIP header called *Puzzle*, an example of which is shown in Figure 2.1. The challenge requires that the sender find a *pre-image* that when hashed with SHA1 results in a value called the *image*. The only known way to find the pre-image is by brute force with the sender having to try many values before finding the solution. The recipient provides the sender with a partial pre-image with some of the lower order bits set to zero and also reports how many bits in the pre-image have been set to zero. In order to make the puzzle easier or harder to solve the number of zero-bits, known as the *work*, can be reduced or increased, and thus the puzzle complexity can be adapted for different situations. The maximum number of iterations required to solve the puzzle is 2^{work}, and hence the average iterations required to solve the puzzle is $2^{work}/2$. The recipient only needs to compute a single SHA1 hash in order to generate the puzzle, which requires very little computing power.

The author notes that proxies may do the work on both the sending and receiving sides. For example the receiver's proxy might challenge the sender with a hashcash and the sender's proxy might compute the hashcash on the sender's behalf, assuming of course that the proxy would be willing to do this only for authenticated users with a pre-existing service relationship. Still the mechanism will greatly reduce the number of requests that a single user agent will be able to create in a short period of time.

[1] 4xx responses in SIP indicate some type of client error. The 419 response is most likely a humorous reference to the slang term for advance fee fraud.

```
INVITE sip:bob@biloxi.com SIP/2.0
Via: SIP/2.0/UDP biloxi.com;branch=z9hG4bKnashds8
;received=192.0.2.1
;spam=-5
;spam-detail="Hormel-1.0;whitelist=-10,call_volume=5"
Via: SIP/2.0/UDP sip.atlanta.com;branch=z9hG4bKfjzc
;received=192.0.3.2
;spam=-100
;spam-detail="Jaeger-3.3;not-a-spammer=-100"
Max-Forwards: 70
To: Bob <sip:bob@biloxi.com>
From: Alice <sip:alice@atlanta.com>;tag=1928301774
Call-ID: a84b4c76e66710@pc33.atlanta.com
CSeq: 314159 INVITE
Contact: <sip:alice@pc33.atlanta.com>
Content-Type: application/sdp
Content-Length: 142
```

Figure 2.2: SIP message with spam scores appended to the via headers.

2.1.3 Spam Score for SIP

In the Internet draft *Spam Score for SIP* [93], Wing *et al.* propose a mechanism for SIP proxies to communicate a spam score to downstream proxies or user agents so that alternative call handling can be performed. This allows intermediate proxies equipped with spam detection software to warn downstream proxies of potential spam. The spam score is most beneficial when inserted by a proxy in the user's home domain, which should be trusted more than any previously traversed domains. It is recommended that the scores generated only by trusted SIP proxies be used, thus preventing untrustworthy proxies from purposefully reducing spam scores.

Positive spam scores indicate that the request is considered spam whereas negative scores indicate that it is considered legitimate, and the higher the value the more likely a request is spam. The draft does not specifically give a range for these scores but the ABNF (Augmented Backus–Naur Form) syntax dictates a one to four digit number giving a maximum range of [-9999, 9999], assuming no decimal places. The score is placed in the *Via* header adjacent the proxy address, and is indicated by the *spam* tag. Thus no new SIP headers are required. Several proxies along the path can insert their own identifiable spam scores, however, it is possible for subsequent malicious proxies to alter these scores. Therefore, it is recommended to cease processing these spam scores at the first Via header from an untrusted proxy. The draft also specifies a second *spam-detail* tag that enables the proxy to transmit further information regarding the nature of the spam. The example in Figure 2.2 illustrates a message that has passed through two proxies that have both added spam

scores and additional spam information.

Proxies and user agents may choose to handle a call differently based on the spam score. For example the call may be routed to voice-mail, the phone may ring inaudibly with a flashing light, or the caller may be challenged with a Turing test. The recipient may also choose to accept calls with high spam probability during the day but reject these calls at night. The draft gives no indication how these spam scores should be calculated nor does it dictate any specific behaviour that should be followed when receiving a call with a high spam call.

2.1.4 A Framework to Tackle Spam

In the draft *A Framework to tackle Spam and Unwanted Communication for Internet Telephony* [85], Tschofenig *et al.* describe a framework to piece together the numerous spam tackling mechanisms, specifically the solutions proposed by Rosenberg and Jennings [65]. The authors claim that there is no single solution that provides 100% protection against spam and consequently new building blocks must be continuously added to the solution space puzzle. The work aims to produce a model that defines internal device processing, protocol interfaces and terminology, that allows new protocols to be added seamlessly to existing solutions.

The proposed framework dictates that the user should upload an authorisation policy document to their network provider. The provider is then responsible for the enforcement of these policies thus reducing the load and bandwidth usage of the SIP device. Figure 2.3 shows the proposed framework in which the end user uploads the authorisation policies and these policies are enforced by the authorisation engine and the message routing engine, both of which reside within the same SIP proxy. The authorisation engine may use any part of the SIP message to make policy decisions but it is noted that in reality only a few headers are used in the majority of cases. It is argued that the most important header to consider is the authenticated identity header. The document recommends that the SIP Identity mechanism [52] that was discussed earlier be utilised for this purpose.

The *protocol interaction for authorising the message sender* refers to either the recipient or the proxy challenging the sender via a computational puzzle, a required micro-payment or a Turing test. This is a pull model because the

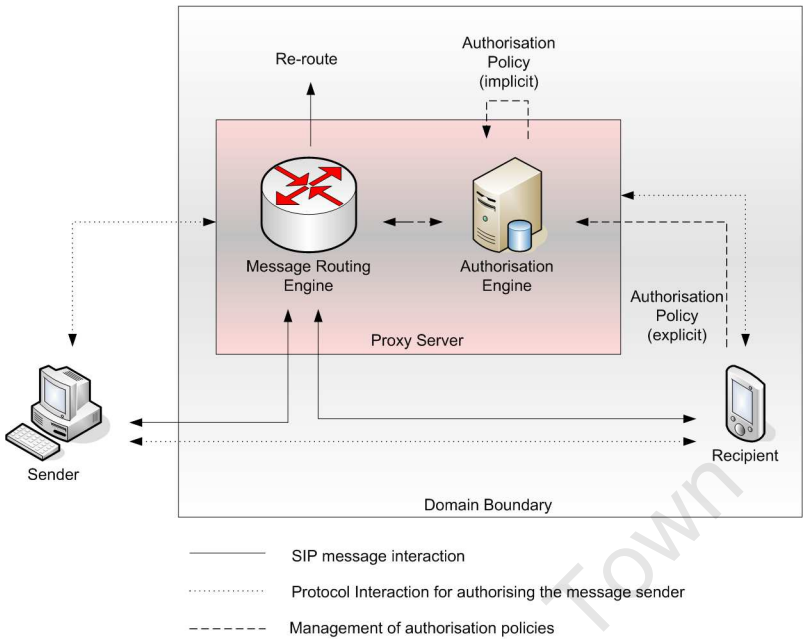

Figure 2.3: The framework proposed by Tschofenig *et al.* for the interaction between the end user and the authorisation proxy.

challenge is initiated by the recipient, whereas the authorised identity is a push model because the sender by default includes an authenticated identity. The framework enables the message to be rerouted (to an answering machine for example), passed on to the recipient or rejected. The message may also trigger further actions such as execution of a consent framework requiring permission from the recipient. If the SIP provider implements some kind of statistical spam blocking tools the framework provides a mechanism for the end-user to choose whether to enable these tools or not. The user may also choose whether or not to use information provided by other proxies, for example the Spam Score header [93].

The authors define three groups of communications users: closed, semi-open and open. Closed groups consist of people that will accept only communication from known and trusted sources. For example, parents that do not wish their children to be contacted by strangers will mandate that they are part of a closed group, consisting only of close friends and family. The authorisation policies of these groups is easy to manage with a straightforward whitelist.

Semi-open groups allow members unhindered contact with anyone else in the same group, such as in a company environment. An individual from outside

the group is classified as either a known contact, with whom contact has been made before, an interesting stranger, an uninteresting stranger or a transaction related stranger. Interesting strangers may include new business contacts or likewise, whereas uninteresting strangers are calls from people that the recipient would rather not respond to, possibly a spammer. The transaction-related stranger is a call that is based on a previous transaction, perhaps from a online shopping site or an airline. Transaction-related strangers present a difficult problem because their communications are often marked as spam by content filters.

Open groups include people that cannot afford to limit communication attempts such as help desks, government agencies or emergency services. Most of the communication attempts to the open group are from previously unknown callers and hence there is no way to provide any whitelisting or blacklisting services. The members of this group are especially hard to protect because they cannot risk rejecting a legitimate call. This may rule out the use of Turing tests and other spam prevention mechanisms.

Individuals may be part of different groups depending on the context. For example at work the individual may be part of a semi-open group whereas at home or during a meeting they may be part of a closed group. This concept is described as switching between different spheres and hence switching to a different authorisation rule set. These rule sets may be enforced by end host or by a trusted intermediary. It is not desirable for the end host to enforce the authorisation rules as this increases the complexity of the terminal and wastes resources. It is preferable if rule enforcement is performed by a trusted intermediary who is specifically designed for this task. The trusted intermediary is responsible for authenticating the sender, digesting the authorisation policies based on a standard format, and providing an interface for accepting the authorisation documents from the end user. XCAP (XML Configuration Access Protocol) [64] is recommended as the interface between the end user and the intermediary; it is used to upload, modify and delete XML (Extensible Markup Language) authorisation documents on the SIP proxy over an HTTP connection.

Due to the immaturity of many multimedia spam solutions the authors recommend that clients be able to query the SIP proxy as to its current spam blocking capabilities. Operators may choose to deploy solutions incrementally, therefore the proxy may support only basic whitelist and blacklist functionality

at first, but later be equipped with more sophisticated statistical filtering. The user is able to query the capabilities of the server using XCAP and the server will reply with an XML document listing its currently available spam blocking mechanisms. In this way, the authors' framework gives users access to the latest technologies available.

2.1.5 A Document Format for Expressing Authorisation Policies

The Internet draft *A Document Format for Expressing Authorisation Policies to Tackle Spam and Unwanted Communication for Internet Telephony* [86] by Tschofenig *et al.* defines an XML authorisation language for anti-spam policy documents. The language format allows users to upload their authorisation policy preferences to a trusted intermediary who will then enforce these policies, thus fulfilling the requirement identified in the Tschofenig *et al.* draft on tackling spam [85]. The proposed document format is an extension to the Common Policy authorisation framework for expressing privacy preferences [72], adding new conditions and actions.

The draft defines the *rule maker* as an entity that creates the authorisation policies. The rule maker could be the end-user, the network provider or even the concerned parent of a child using a mobile phone. The authorisation policies consist of a list of unordered rules and each rule has a condition, an action and a transformation component. The action and transformation components define the permission level of the rule, for example block or allow. Thus, depending on the rule, the permissions can be used to block any type of unsolicited multimedia attempts. The XML authorisation documents are uploaded to a spam blocking SIP proxy over the XCAP interface.

Rules are usually matched according to the sender of the request, therefore only authenticated identities are matched. All identities that have not been authenticated, either by the SIP Identity mechanism or the P-Preferred-Identity mechanism, will match rules only with empty identifiers, and the permissions of these rules will most likely specify that they should be blocked. The draft defines several unique extensions to the Common Policy framework specifically for blocking spam including the *sphere*, *spit-handling*, *media-list*, *method-list*, *mime-list*, *presence-status*, *rule-deactivated* and *time-period* elements. All these elements are used to fine-tune exactly which permissions are

```
<?xml version="1.0" encoding="UTF-8"?>
<ruleset xmlns="urn:ietf:params:xml:ns:common-policy"
xmlns:spit="urn:ietf:params:xml:ns:spit-policy"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
<rule id="r1">
  <conditions>
    <spit:spit-handling>
      <challenge result="FAILURE">captcha</challenge>
    </spit:spit-handling>
  </conditions>
  <actions>
    <spit:forward-to>
      <target>sip:answering-machine@home.foo-bar.com</target>
    </spit:forward-to>
  </actions>
  <transformations/>
</rule>
</ruleset>
```

Figure 2.4: An authorisation policy document that forwards all callers who fail a Turing test to voice-mail.

to be granted to the specified users. For example, using these elements it is possible to specify that an unknown caller trying to start a video call between the 9pm and 10pm on a Sunday night will be blocked, or that friends may start any type of multimedia session so long as it is not during office hours. The spit-handling element evaluates the outcome of spam blocking mechanism, such as a Turing test, and applies an action based on this result. Thus a rule can be specified that anyone who fails a Turing test is forwarded to voicemail using the *forward-to* element, as illustrated in Figure 2.4.

2.2 Detection of Voice Spam

Robo-calling refers to the practise of making automated calls with a computer and playing a pre-recorded message. These calls are made in bulk with companies such as PoliticalCalling.com claiming to have the capacity to make over 700,000 calls per hour; significantly more than can be made by telemarketers. Robo-calls are particularly intrusive because, unlike email, the phone will ring and the recipient must immediately choose whether to accept or reject the call. Moreover, it takes longer to identify a call as spam than a text message, because the recipient must listen for at least a few seconds to realise the other party is not human. Robo-calls have become a serious nuisance in some western countries such as the USA where political parties use them to garner votes and spread information about political activities [11]. The section covers some of the recent research in counteracting automated callers.

2.2.1 Progressive Multi Gray-Levelling

In the article *Progressive Multi Gray-Levelling: A Voice Spam Protection Algorithm* [75], Shin *et al.* discuss a new method of detecting voice spam based on the volume of calls made by a particular caller over a certain time period. The proposed mechanism determines whether the caller is a likely spammer or not based on a calculated gray level, a term used to highlight that the system does not categorically accept or reject calls in the way that whitelists and blacklists do. Rather, the system gives the caller a gray value that changes depending on their call patterns. Callers whose score exceeds the threshold, are blocked temporarily until their gray level returns to an acceptable level. Thus the system is seen to be superior to a blacklist that blocks senders permanently even if their behaviour improves.

Voice spam is considered to be more malicious than text-based spam, as it is possible to delete several junk messages at the same time. If spam is received in a voice mailbox the user must listen through each message individually before deciding whether or not it is legitimate. The real-time nature of voice calls is also seen as a problem for voice spam protection as there is no time to check and quarantine suspect calls. Thus the authors have proposed a mechanism to detect spammers before the call can even be initiated based on previous call pattern analysis.

Progressive gray levelling assigns two scores to the caller: a short term and a long term gray level. The short term gray level is calculated from the very recent past behaviour of the caller, perhaps in the last minute. This score is expected to rise very quickly when a robo-caller first starts making bulk calls but curtails within a few hours back to zero if the caller's behaviour improves. The short term gray level protects against surges of spamming but is ineffective against long term bursts of bulk calling. The long-term gray level, however, tracks the behaviour of caller over a much longer period of time, for example over a number of hours or days. The long-term gray level also takes into account the number of times that the caller was identified as a spammer, and multiplies the long term-term level by this number. In this way the system allows a spammer to make only a fraction of the calls that were made when they first started initiating bulk calls. The long-term level also takes a significantly longer period of time to reduce back to zero thus providing protection against only part time offenders.

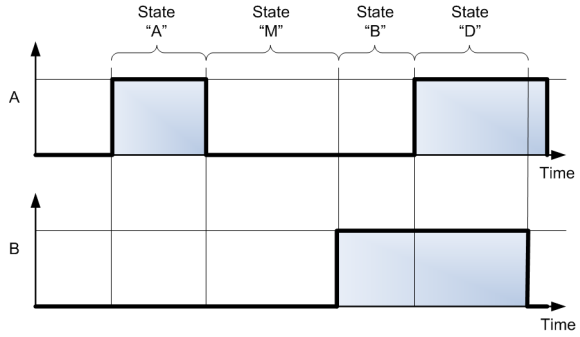

Figure 2.5: A four-state two-way telephone conversation model.

The authors implemented the progressive multi gray levelling scheme in a physical testbed environment using Iptel's SIP Express Router as the SIP proxy that houses their algorithm. Through their experiments it was found that the algorithm can successfully block short term spamming attempts and over a long period is able to prevent persistent spamming behaviour. The configuration of the algorithm parameters, such as the short term and long term gray level measuring period, is left up to the implementer. However, it was noted that obtaining the correct parameters for the algorithm is important, since there is a risk that measuring over too short or too long a time period can reduce the effectiveness of the system, and possibly blacklist a caller forever.

2.2.2 Detecting SPIT Calls by Checking Human Communication Patterns

Rosenberg and Jennings [65] proposed the Turing test for identifying non-human callers by subjecting them to a challenge that only a human can answer. Turing tests usually ask the caller to identify characters in a grainy picture or words spoken over background music. The problem with these tests is that they are intrusive, add to call setup delay, and are not suitable for disabled people with poor eyesight or hearing. Quittek *et al.* propose a solution to this problem in their paper *Detecting SPIT Calls by Checking Human Communication Patterns* [58].

The paper describes a mechanism for identifying automated callers based on the premise that human calling patterns are predictable and measurable. The authors use the Hammer *et al.* [33] four-state model, depicted in Figure 2.5, that describes the behaviour of two parties involved in a voice call. In state

A and B only one of the parties are talking, whereas in state M there is mutual silence and in state D, known as double talk, both parties talk at the same time. For obvious reasons conversations do not spend much time in the D state and it is noted that conversations typically exit the D state within 0.23 seconds before entering one of the other states. This pattern of human communication is exploited by what is referred to as a *hidden Turing test* that identifies when the caller is exhibiting atypical calling etiquette at the beginning of the conversation.

A normal conversation begins with the call recipient picking up the handset and announcing some type of greeting. Once the recipient's greeting is complete the caller then responds, although sometimes there is a brief period of double talk or mutual silence at the end of the recipient's greeting. The hidden Turing test checks for two types of anomalous conversational behaviour, silence checking and answer length checking, in order to verify that the caller is human.

Silence checking determines if the caller has entered into a period of double talk during the initial greeting by the recipient. The algorithm allows for a brief period of double talk that is normal in several cultures, but if the voice energy of the caller exceeds a pre-defined threshold during this period then the caller is classified as a machine. The algorithm also makes allowances for some background noise from the caller and peaks in the caller's noise energy that may be introduced by codec problems. The authors suggest one possible manner for ensuring a good environment for performing the test is to send a fake ringing tone at the start of the call and see whether the caller starts talking during this time. If this test is inconclusive then the answer length checking test is performed.

Answer length checking measures the length of the response to the recipient's initial greeting. Usually one would expect that the caller's first period of speech is fairly short. However, if the caller starts a very long period of voice activity, probably because a pre-recorded message is being played, then the call is classified as junk. Of course it is imperative that the recipient's first greeting does not illicit a long response or the system could incorrectly classify the caller as a machine.

The hidden Turing test is seen to have several advantages over the traditional challenge-response style of test. The test cannot offend the caller; it is performed quickly; it works with callers of different languages and pronuncia-

tions; it can be implemented on modest hardware; and it is a resource-intensive task for spammers to circumvent the test. Unfortunately the authors are unable to verify the effectiveness of their system in a practical environment due to the unavailability of spam call records.

The method described by Quittek *et al.* fulfills many of the objectives of this thesis, as presents a manner by which an automated caller can be identified in an unobtrusive manner. However, without evaluation in live environment it is unclear whether such a system will provide suitable protection. It is most likely that this technique will be most effective when used in conjunction with other techniques described in this chapter.

2.3 Detection of Instant Message Spam

In the previous sections the problem of SIP spam was introduced and several proposals were reviewed that aim to filter real-time multimedia sessions. Thankfully it is significantly easier to filter text messages than audio or video sessions, as content filters are able to parse and classify text messages using keyword detection before they are delivered to their final destination. Unfortunately there is a recent trend for spam to be sent in MIME bodies with both text and image parts, and even the most advanced image detection techniques struggle to identify such messages as spam. This section first discusses Naive Bayesian filtering, the most popular form of spam keyword detection, and then reviews methods for detecting image spam.

2.3.1 A Bayesian Approach to Filtering Junk E-Mail

In the paper *A Bayesian Approach to Filtering Junk E-Mail* [70], Sahami *et al.* propose that junk email can be detected by using a Bayesian word classifier. Bayesian networks are simple acyclic, directed graphs, with each node representing a random variable and the edge representing a probabilistic influence from the parent node to the child. The Naive Bayesian classifier assumes that each random variable is conditionally independent of every other feature. Modern popular spam filtering software applications, such as SpamAssassin, use Naive Bayesian filters to identify words that are most likely found in spam.

The Bayesian filter does not classify an email as spam or legitimate based simply on a single word or phrase. Rather, the existence of suspicious words

simply provides evidence to a probabilistic classifier to increase its confidence in a message being classified as junk. Moreover, the classifier does not base its judgement on the whole message but selects the most interesting features. These features may be words, exclamation points, alphanumeric characters or entire domains. The paper suggests choosing the 500 most interesting features of the email in order to classify the message.

The primary benefit of the Bayesian filter is that it is self-learning. The filter can be trained initially on a data set of known junk and legitimate emails in order to apply correct probabilistic values to the features. Thus the filter can be taught to recognise the different types of email that individuals receive. Moreover, training data can be fed to the algorithm at a later stage in order to adapt to changing spamming behaviour and new threats. Suspicious words, such as *Viagra*, increase the chance of an email being marked as spam, whereas innocent words used in everyday conversation increase the likelihood that the email is marked legitimate.

Unfortunately, due to the popularity of naive Bayesian filters, spammers have developed mechanisms to thwart the algorithm [96]. Most modern spam comes appended with a large number of random innocuous words or entire sentences that make the Bayesian filter believe the message is legitimate. In extreme cases the spammer may even attempt to change the data set of the Bayesian filter through a process called Bayesian poisoning. The spammer tries to trick the filter into believing that several different made-up pseudo-tokens are legitimate. This is achieved by sending a large number of seemingly legitimate or blank email just including the pseudo-token strings. These email are not classified as spam - the pseudo-tokens are given a neutral probability as they have never been seen before, and the end user does not mark them as spam as there is no spam content. Thus over time the Bayesian filter is trained that emails containing these tokens are legitimate. In this way the spammer improves the probability of future spam delivery by simply inserting the pseudo-tokens into the body of the message.

2.3.2 Image Spam

In the white paper *Image Spam - The New Face of Email Threats* [37], the authors discuss the solution adopted by the PUREmail system that detects various types of image spam. The system uses several existing image processing

Figure 2.6: An image spam advertising generic drugs. The text geometry is skewed and random noise is added making optical character recognition problematic.

techniques in order to identify image spam with a claimed accuracy of 98%. These techniques include optical character recognition, noise reduction, edge detection and skin detection.

Optical character recognition is used to identify alpha-numeric characters from the image that can then be passed through a text-based filter. Traditionally spammers have changed the geometry of the letters in image spam, or use different colours, in order to bypass recognition software. Additionally, the image may be sprinkled with random noisy pixels so that the filter cannot identify the patterns in the image; an example is illustrated in Figure 2.6. The noise also makes the image unique so there are not high-volumes of identical pictures that would quickly be identified by ISP spam filters. PUREmail removes noise from the image spam using noise reduction techniques, smoothing the image by taking an average value of nearby pixels. Unfortunately, this process reduces the contrast and sharpness of the image making character recognition harder. Thus the images are passed through an edge detection filter that identifies shapes in the image, using the statistical variance of the colour gradient in the image, at which point the character recognition software can operate effectively.

The PUREmail system also performs pornography detection by detecting the hue of human skin. It was found that in most pornographic content the skin hue makes up 80% of the colour in the image and varies only by shade and intensity. By simply inspecting the colours in the image it is claimed that the

filter can identify pornographic content with 99% accuracy, but no information is given as to the amount of expected false positives.

2.4 Spam Legislation

Several legal measures have been adopted to prevent different types of unsolicited communications. This section briefly discusses the South African legislation dealing with unsolicited electronic communications, and in the United States, the CAN-SPAM Act (Controlling the Assault of Non-Solicited Pornography And Marketing) and the Do Not Call registry.

2.4.1 Electronic Communications and Transactions Act of 2002

The South African *Electronic Communications and Transactions Act* [50] of 2002 deals with the transmission of unsolicited commercial communications. Section 45 of the act states that in order to send unsolicited electronic communications to South African consumers two provisions must be met. The first requirement is that the sender must give the recipient the option to cancel their subscription to the mailing list, but the act does not specify how this must be done. The second requirement is that the source from whom the recipients contact details where obtained must be made available on request. Persons found guilty of contravening this act are liable to a fine or imprisonment for a period not exceeding 12 months.

Therefore, in South Africa it is completely legal to send any kind of unsolicited electronic communication so long as unsubscribe or opt-out information is provided. If the sender is queried they must disclose where the recipients address was obtained, but there is no law against harvesting addresses off the Internet or corporate databases.

2.4.2 CAN Spam Act of 2003

In 2003 the USA signed a CAN-SPAM Act [6] into law thus establishing national standards for sending commercial email. The act establishes requirements for those who send commercial email, spells out penalties for spammers and companies whose products are advertised by spam, and gives end users the

right to opt-out from future communications. The act is primarily concerned with email promoting or advertising a product, service or website content, and stipulates four main provisions for such email.

First, the email's header information must be accurate and identify the source of the message. This includes the *From* and *To* headers as well as the originating domain and email address. Second, the message may not contain a deceptive subject line that could mislead the recipient about the contents of the message. Third, there must be a mechanism for the recipient to opt-out of future messages and the sender must honour this request within ten days. It is illegal to sell or transfer the addresses of persons that have opted-out unless this transfer is to allow another party to comply with the law. Lastly, all commercial email must be identified as an advertisement and include the sender's physical postal address.

The legislation dictates that email sent as part of an existing business relationship, otherwise known as a transactional message, may not contain false routing information but is otherwise exempt from the provisions. Otherwise, any violation of this law may be subject to a fine of up to $11,000. Fines may also be imposed for other spamming behaviour such as harvesting email addresses from websites, generating email addresses using a dictionary attack, using scripts to register for multiple email accounts, taking advantage of open mail relays, using a computer to send spam without permission, or spoofing IP addresses.

Critics of the act have claimed that it actually gives spammers more legal protection since it allows at least one shot at every email address. Some studies have also shown that since the adoption of the act the amount of spam in the Internet has actually increased [97], and that only an insignificant fraction of spam email actually complies with the act. Still the act does provide legal protection against spamming and users have reported significantly less pornographic spam since the act's adoption [60].

2.4.3 Do Not Call Registry

The US *Do Not Call Registry*, established in 2003, gives American consumers the opportunity to limit the amount of telemarketing calls they receive. Telemarketers may not call a number once it has appeared on the registry for 31 days. Telemarketers are also prohibited from calling any cellular telephone or

fax numbers. Inclusion in the registry is free and a listing is valid for five years after which time it may be re-registered. However, being listed in the registry does not mean an end to junk calls since there are several exceptions to the do not call regulations. For example, businesses telephone numbers may not be listed on the registry. Political organisations, charities and persons conducting surveys are also exempt from the regulations, as are companies that have had a business relationship with the call recipient within 18 months.

Thus far compliance with the regulations in the US has been encouraging with very relatively few complaints being received by the Federal Trade Commission [18]. Other countries, including Australia, Great Britain, Canada and New Zealand have implemented similar regulations to prevent the increasing amount of telemarketing. Unfortunately, these laws are hard to enforce on callers from outside the national boundaries and there have been reports of companies outsourcing their telemarketing to countries such as India in order to avoid prosecution.

2.5 3GPP Working Group

The 3GPP has proposed a new work item to be carried out with the OMA called *Protection against SMS and MMS spam* [74] in order to deal with mobile spam. The work item only covers text and multimedia messaging in the GSM system, but notes that in the future IMS messaging will also be taken into account. The objective of the study is to develop several techniques for controlling mobile spam by allowing the customer to decide which messages should be considered spam. Furthermore, it is noted that any devised solution should be effective regardless of whether the customer is on the home network or roaming. The goal of the work is to allow mobile operators to provide anti-spam services hence giving their customers increased confidence in both the messaging technologies and content provisioning services.

The work item suggests several solutions to the spam problem including whitelisting, blacklisting and keyword filtering. The document also suggests a junk folder of the mobile device for storing spam temporarily, thus avoiding deleting messages that have been wrongly identified as spam but also conserving storage space for legitimate messages. The mobile may also query the user for validation every time spam is detected so that the user can gain confidence in the system. A rate limiter is also proposed that allows a user to specify the

maximum frequency that they will accept messages from one source. As yet, apart from these initial recommendations, there are no outputs from this work item and no time scale has been set for completion.

2.6 Discussion

The chapter has detailed the current state of the art in SIP spam avoidance techniques, automated caller detection, spam detection and current legislation. It has also discussed the current work by the 3GPP dealing with spam in the IMS architecture. It is clear by the amount of literature available that the IETF has researched the problem of multimedia spam far more thoroughly than the 3GPP. However this should be of very little concern, since it is the policy of the 3GPP to adopt IETF standards wherever possible into the IMS standards [17]. Therefore, it is not unreasonable to assume that the work currently being performed by the IETF in combating SIP spam will soon be incorporated into the IMS specifications. For this reason the thesis places a strong emphasis on the solutions and mechanisms proposed by the IETF in developing an architecture for the prevention of spam and enhancement of privacy in the IMS.

Chapter 3

Case Studies

There have been several recent cases that highlight the growing problem of all types of spam. This chapter discusses some of the most concerning spam and electronic fraud attacks and the measures taken at the time in order to curb these attacks. The spam attacks detailed in this chapter are noteworthy because each was implemented in a novel manner and proved to be very successful for the attackers. A review of past attacks prevents future similar offences from being perpetrated in new telecommunications environments such as the IMS. Any proposed solutions to the IMS spam problem must take into account these case studies and attempt to provide suitable defence mechanisms.

3.1 Russian Business Network Harbours Online Crime

The Russian Business Network is an Internet business that specialises in hosting illegal websites devoted to spamming and phishing practises [43]. It is believed that half of all phishing attempts in 2006 were perpetrated from the company's servers, netting the phishers approximately US$150 million. The company also provides shelter for several other kinds of illegal activities including child pornography, online scams and piracy. However, the problem facing law enforcement is that the Russian Business Network itself simply provides hosting services and does not violate any laws. The criminals using the network also tend to target foreign companies and consumers so there are few complaints made to the local Russian authorities.

Thus far there has been little success in preventing attacks originating from

the Russian Business Network. Government agencies from other countries such as the US have not been able to elicit cooperation from Russian officials in arresting the individuals behind the company or to shut it down completely. This is most likely because the company has ties with the Russian criminal underground and bribed government officials. Such companies are able to survive in countries where laws, infrastructure and investigative support are poorly established. This presents a large problem for Internet Service Providers and telecommunications companies in other countries that are bombarded with spamming and phishing attempts with little legal recourse.

Out of desperation some smaller ISPs have adopted a strategy that involves blocking all traffic from the Russian Business Network regardless of its legitimacy. This strategy is far from ideal and may result in serious repercussions from consumer rights groups that advocate network neutrality. Large ISPs cannot afford to block off what is essentially a sizable portion of the Internet without provoking complaints from their customers. However, those ISPs that have blocked the company have found that the instances of spam and infected machines dropped significantly. They also spent less money on cleaning machines infected with viruses and spam zombies.

The Russian Business Network remains a problem for ISPs and telecommunications providers worldwide and no reasonable resolution has yet been proposed. This case demonstrates that despite the best efforts of network providers there is no way to stop companies with questionable business ethics from spamming and phishing. Since government legislation is essentially powerless to prevent these activities the responsibility lies with network providers to protect their customers from such attacks.

3.2 Pump and Dump Scams Cause Havoc on Stock Market

Studies have shown that 15% of the approximately 730-million unsolicited emails sent every week are stock touts. The messages suggest that the recipient should invest in a particular penny stock because the price is about to rise. This type of fraud is known as the *pump and dump* scam, or micro-cap stock fraud, referring to the small market capitalisation of the companies being traded.

The spammer will have already invested in the stock and will therefore gain

from any rise in the stock price and increased trading activity. The practice of buying penny-stocks, touting them and then selling immediately is sometimes referred to as *buy low and spam high*. Evidence has shown that stock touts are often very successful; for example one study has found that a stock that only had a 6% chance of being the most actively traded in a day has an 81% chance of being the most actively traded on the day of a mass email stock tout [28]. Unfortunately for the unwitting investors the study also found that the price of the stock usually drops by an average of 5.25% in the two days following the touting, most likely because the perpetrator of the scheme has offloaded their stocks during this period.

Pump and dump campaigns have serious repercussions for the companies whose stocks are being touted. In May 2007 the US Securities and Exchange commission was forced to suspend trading on 35 small companies due to spam email campaigns urging investors to buy their shares [94]. The suspension not only served to prevent investors from losing money but also to protect the companies themselves. Shares have been known to increase in value by 800% following stock touting, causing chaos and confusion for executives with stock options, and sending the market capitalisation of the company on a roller-coaster. These examples highlight the fact that spam touting campaigns are highly effective because trading is taking place not on the basis of underlying fundamentals but rather the misconceptions of those who received spam touts.

There have been other examples of the powerful effects of spam on the stock market. In mid-2007 stocks in technology giant Apple tumbled when a hoax email was distributed claiming that the release of the iPhone and the Leopard Operating system were being delayed [10]. Within six minutes, US$4-billion of Apple's market value had been wiped out. The company later managed to calm investors by labelling the email as a hoax and the share price recovered accordingly. The perpetrators of the scam no doubt made a healthy profit from buying the undervalued shares as soon as the price crashed.

Stock touting spam is increasingly being sent as image spam or by PDF attachment, thereby avoiding content filters. Due to the prevalence of stock fraud it is highly likely that it will quickly manifest itself in various guises throughout the IMS. Indeed, before spamming technologies became so advanced most micro-cap stock fraud was perpetrated via telemarketing. Prerecorded voice and video messages delivered by robo-callers are likely to be even more effective than the current text-based touts and may become a serious headache for

IMS network operators.

3.3 Users Defrauded with the Unknown Missed Call

In the missed call scam fraudsters call random cellular phone numbers, let the call ring for a fraction of a second and then hang up leaving a missed call indication on the recipient's phone. When the individual calls the number to find out who was trying to contact them they are unknowingly directed to a premium rate number. This scam can be highly profitable for the perpetrators since it costs nothing to make an unanswered call yet they are able to charge huge amounts when the call is returned.

In 2004 the British Independent Committee for the Supervision of Telephone Information Services was forced to shut down two firms found guilty of this practice [49]. When users called the premium rate number they were informed that they had won a prize but in reality the were simply accruing a hefty phone bill, sometimes as much as US$20. In Britain the practise is considered illegal since the perpetrators are committing fraud, but some countries do not have legislation governing such activities.

This fraud can easily be conducted in an IMS environment due to the ease of which missed calls can be made by robo-callers. Moreover, like presence spam, the missed call notification could be used as a marketing tool, for example the missed call could be from *sip:buy_a_goldfish@pets.com*. Using rudimentary robo-calling software it is possible to cancel the call immediately after receiving a ringing notification and probably before the recipient's phone has had a chance to make an audible notification to the user. This novel type of spamming has the potential to become a huge inconvenience for users of IMS services who might find several missed call notifications on their phones throughout the day, thus greatly reducing the utility of this feature.

3.4 Robo-call Flood in 2006 Mid-Term Elections

Robo-calling is a highly intrusive form of spam that has seen wide adoption in recent political events such as the 2006 US mid-term elections. A study has shown that 64% of all registered voters received a pre-recorded voice call in

the final two months leading up to the election [59]. Robo-calling has become the second most popular mechanism to reach voters behind direct mailing campaigns that targeted 71% of voters. Only 14% of voters reported receiving email solicitations and 24% received a call from a human.

The Republican party was mostly blamed for the deluge of robo-calls that triggered a huge backlash from the public, outraged by the constant invasion of privacy [11]. Some voters would receive an average of three to five calls per day from Republicans and Democrats. There were also reports of people receiving the same pre-recorded call over and over again for weeks at a time. Unfortunately for the public, political parties and charities are not governed by the US do-not-call registry, therefore no laws were broken. However, due to massive public outrage and allegations that the calls were misleading the public, the Republican Party agreed to end calls to people on the do-not-call list.

Records have shown that Republicans spent approximately US$600,000 on robo-calls during the campaign [11], yet this remains one of the cheapest mechanisms to reach voters. The calls are often scathing attacks on fellow running mates accusing them of trying to raise taxes or other misgivings. The originator of the call is often disclosed only at the end of the message, if the person decides to listen for that long. This can result in confusion for the recipient as to who was responsible for the call.

Celebrities, comedians and well-known political figures are used to make the recordings in order to retain the interest of the recipient. The campaigns are also strategically directed at those who are more likely to vote, especially affluent and older voters. This type of electioneering has thus far proved to be very successful for the parties involved and will therefore continue to be utilised as a tool for garnering votes. Currently, consumers have little protection against this invasion of privacy apart from hanging up the phone once they realise they are listening to a recording.

It has been speculated that spamming campaigns could seriously influence the forthcoming US presidential elections [46]. In 2006, for example, 14000 Latinos received letters incorrectly informing them that it was illegal for immigrants to vote. In another incident during 2004 fraudsters opened a fake website claiming to be collecting money for the John Kerry campaign thereby stealing campaign funds and credit card numbers. It is very hard for consumers to spot phishing attempts like these and it is expected that similar dirty tricks

campaigns will be launched in the lead up to the 2008 elections.

3.5 Vodacom MMS Campaign

Over the 2006 holiday period South Africa's largest cellular network provider bombarded their own customers with a flood of unsolicited MMS (Multimedia Messaging Service) messages [83]. The company claimed the campaign was aimed at informing and educating customers about the MMS technology but critics labelled the exercise as spamming. To make matters worse, consumers could opt-out from receiving further messages only by replying to a premium rated number.

It is unclear whether the Vodacom MMS campaign violated section 45 of the Electronic Communications Act as the act fails to clearly define what constitutes spam. Still, the campaign sparked outrage amongst some customers especially due to the costly opt-out mechanism. Despite the fact that the company stood to profit from increased use of MMS messaging, Vodacom claimed that the campaign did not seek to promote a chargeable product or service and thus could not be classified as spam. Furthermore, the incident highlighted that the company provided no mechanisms for their customers to opt out of unsolicited communications.

3.6 Discussion

This chapter has highlighted a number of case studies in which individuals and companies have attempted to extract profit from unsolicited communications. The problem of spam is no longer restricted to email marketing and now encompasses several types of fraud and propaganda. Furthermore, it is increasingly being distributed over different electronic communication mediums including voice and mobile messaging systems. Spam is not only an inconvenience for network providers and users alike, but also has the power to alter stock markets, defraud unsuspecting consumers and perhaps even alter the outcome of elections.

Unsolicited mobile communications are on the rise in many countries, including South Africa due to reducing costs and poor legislation dealing with the issue. This was highlighted in the Vodacom MMS incident in which a

company sent thousands of unsolicited messages with complete impunity. The traditional communications model dictates that anyone can be contacted electronically as long as their address or phone number is available. This model needs to be completely retooled for emerging next generation telecommunications networks in order to guarantee the privacy of users and protect them from annoying, and possibly harmful, unsolicited multimedia.

Part II

Data Collection

Chapter 4

Consumer Perceptions on Spam - A Survey

This chapter details a survey conducted in order to identify current consumer perceptions on multimedia spam. As there are currently no full IMS network implementations, the survey is primarily interested in spam trends in current mobile telephony, as it is the closest reference point thus far to next generation communications.

Cellular telephones have an extremely high penetration amongst the South African population compared with fixed lines [8], and provide additional communications services apart from basic voice, such as text and multimedia messaging. The 3GPP plans to incoporate these advanced services, amongst others, into the IMS as it is a evolution of the current cellular network architecture. Hence the survey focused on GSM telephones thus providing a platform for perceptions towards spam in the IMS environment.

4.1 Motivation

Several surveys have been performed in order to determine public reaction towards email spam [23, 60, 82]. However, a review of literature suggests that there is currently a critical shortage of data available regarding consumer perceptions of multimedia spam. Furthermore, South African corporations and the government have tended to overlook consumer privacy and have thus not researched the growing problem of multimedia spam [50]. Therefore, it was deemed useful that the thesis conduct its own spam survey in order to perform

a temperature check on current consumer perceptions towards spam of various types, and how spam affects the utility gained from electronic communication devices.

Given the tentative nature of the spam survey, scope and time-frame, the development and execution of this sub-project was limited accordingly. Although clearly limited in terms of generalisability of the results, the survey, in the absence of any other available data, provides a relevant and useful basis of departure for any discussion around multimedia spam in the South African context.

4.2 Item Construction

Construction of the initial survey questionnaire was informed and guided by the available literature pertinent to the thesis. Several exchanges with experts in the field helped to refine the question content. A psychology graduate student from UCT (University of Cape Town) was employed to further develop and execute the actual survey. With the assistance of this student a preliminary set of questions were constructed that yielded a conceptually valid survey for attitude research. The resulting questionnaire consisted of three subsections.

4.2.1 Section A

The section consisted of items one to six and assessed participants exposure to various forms of multimedia spam (i.e. calls, SMS, MMS) in terms of frequencies. The first four items employed a multiple-choice format, asking participants to rate the frequency of their exposure to spam from daily to never. The last two items assess the frequency of spam exposure that would compel a person to discontinue use of their cellphone. The questions offer multiple choices ranging from *10 per day* to *10 per month*, and also included an option indicating that *I would not stop using my cellphone*.

4.2.2 Section B

The 15 items of Section B were designed to form a coherent scale that measures the levels of each participant's negative attitude towards cellphone spam. It is most feasible to design direct questions in order to assess attitudes among

populations [41]. It has been stated that an attitude is comprised of three components: a favourable or unfavourable evaluation; a belief; and a behavioural disposition, all of which were integrated to form the items of the scale [16]. A multiple-choice design in accordance with Likert-Type scaling has been proven to be the most effective tool for attitude assessment [77]. This format offers an answer range of *Strongly Agree*, *Agree*, *Disagree*, *Strongly Disagree*. The content of some of the items was reversed in order to control for acquiescence, i.e. a participant's tendency to agree with item regardless of their content [69].

4.2.3 Section C

The last section, consisting of six items, aimed to assess where participants locate responsibility of managing the multimedia spam to which they are exposed (i.e. self, government, network providers). Again, Likert-Type scaling, ranging from *Strongly Agree* to *Strongly Disagree* was used to assess this dimension.

4.3 Validity and Reliability of the Questionnaire

The initial version of the questionnaire was subjected to a pilot survey in order to test the administration of the study, and to produce data for the necessary item analysis for possible subsequent revision of Section B of the questionnaire. Sections A and C did not require such an analysis, since the individual items were designed to produce data that were conceptually relevant and valuable in their own right. Section B required pilot data testing, because the set of items was designed to form a coherent scale that produces a single score for interpretation from each participant. Therefore, these questions were scrutinised in terms of validity ensuring inferences made from the tests/scale scores were considered meaningful, appropriate and useful, and for reliability in ensuring all items of the test measured the same construct [31].

4.4 Data Collection

4.4.1 Pilot Survey

Participant selection for the pilot study was convenience based. Although survey research design ideally uses more sophisticated sampling approaches, such

as randomisation or stratification, in order to achieve accurate representation, the scope of the project has to be taken into account when deciding on a sampling approach [15]. The spam survey aimed to make merely tentative attempts at filling the void with relevant and useful data in relation to the topic, therefore employing a convenience sampling strategy was apposite under the constrained circumstances surrounding the brief sub-study of the present research. In the case of the pilot, the sample consisted of 82 first-year psychology students at UCT (62 female; 20 male). The mean age of the participants was 20 years. The students were asked to complete the questionnaire in one of their lectures. The actual administration of the survey was unproblematic and no questions arose from the participants, indicating clarity and face validity of the questionnaire and data collection procedures. Only one questionnaire script had to be excluded from analysis because it was returned incomplete.

4.4.2 Pilot Analysis and Revisions

Administration of the pilot survey indicated that Section A of the questionnaire could be maintained in its original form, since no issues arose during the course of data collection. Given the design of Section B as a coherent scale, the data was tested for reliability issues using the Cronbach's Alpha test-statistic. Cronbach's Alpha is essentially a set of correlations that can single out items within a scale that are problematic, in that it measures something besides the construct under observation [27].

The initial analysis for the pilot data for Section B yielded a score of 0.62, which indicates sufficient levels of reliability in the moderate range of the acceptable spectrum. Individual item analysis suggested that five items were problematic. The data from these questions were, for Alpha testing purposes, excluded from the analysis, which immediately resulted in a strong Alpha score of 0.69. Accordingly, these problematic items were eliminated from the final version of the survey questionnaire.

In terms of issues of validity, given the scope and time-frame of the survey, it was neither practical nor possible to employ the spectrum of the more extensive means for testing in this respect. For example, it was not viable to correlate the pilot data from the scale developed for this project with data from existing scales, because research in the context under scrutiny here has been, at best, scarce. Therefore, the survey had to rely on the idea that proper

state of validity can be established through extensive and multidimensional consultation of experts and sources in relation to the construct under observation [15]. Thus a detailed review of the literature provided the validity-basis for this survey.

A review of the data for Section C suggested over-complication of this part of the survey, since the data appeared somewhat incoherent, even at face value. For the sake of simplification, brevity, and improved face- and overall validity, the items of Section C were collapsed into a single question. The item was constructed to give participants a direct multiple choice as to whom (i.e. network provider, government, self) they assign responsibility to for managing cellphone spam. These data could then be displayed in form of a simple frequency distribution.

4.4.3 Final Survey

The sample for the final survey consisted of 152 respondents (95 female; 57 male). Again, in congruence with the pilot study, participants were selected on a convenience-basis. They were approached either in lectures or in the various departments at UCT. The mean age was 22 years, although the ages ranged from 18-36 years because participants were draw from the student body as well as the academic staff, thus generating a slightly more generalisable sample by comparison to the pilot. Administration of the survey proceeded without any evidence of problems, indicating high face validity and easy comprehension among participants [15]. Only two incomplete survey scripts had to be eliminated from the data analysis. The pilot and final survey documents are included in Appendix A of this thesis.

4.4.4 Data Analysis

The data for Sections A and C were captured by numerically coding the applicable answer categories (i.e. 1-4 or 1-5, depending on the number of multiple-choice options) in order to enable simple computation of frequency distributions and corresponding histograms. In accordance with 'Likert Type' scaling, the items of Section B were scored using values of one to four for each multiple-choice item, depending on the direction of the question [69]. A value of one indicated a low negative attitude towards spam, a value of four a high one. The values were subsequently added to form a single score for each participant,

which could then be computed and displayed in terms of an overall distribution of scores. The results were analysed and plotted using Matlab.

4.5 Current State of Spam (Items 1-4)

The first part of the survey attempted to tentatively gauge the current state of electronic spam in South Africa. In congruence with the focused of this thesis, the questions specifically focus on cellular telephony. However, a question assessing email spam was included to provide a comparison dimension.

The following four questions were posed to the respondents:

- How often do you receive unsolicited email ("Junk Mail")?
- How often do you receive unsolicited phone calls (e.g. Advertising/Sales Calls) on your cellphone?
- How often do you receive unsolicited SMSs (e.g. Advertisements)?
- How often do you receive unsolicited MMSs?

The responses to the above questions are detailed in Figure 4.1, Figure 4.2, Figure 4.3 and Figure 4.4, respectively.

4.5.1 Email Spam

Unsurprisingly over 50% of respondents indicated receiving email spam on a daily basis. 26% of respondents reported receiving spam email at least once per week and 20% reported receiving spam either monthly or less frequently. Only four respondents reported receiving no email spam at all. This is an expected result considering the current prevalence of email spam worldwide. However, it can be assumed that without the current ubiquity of spam filters and firewalls that these results would be quite different.

4.5.2 Unsolicited Voice Calls

The respondents reported receiving far fewer unsolicited voice calls than spam emails, with 24% claiming to receive such calls monthly and 48% claiming to receive unsolicited calls less frequently than monthly; 9% reported receiving

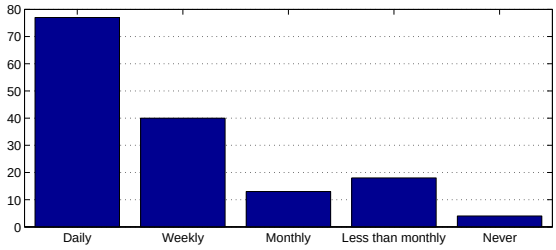

Figure 4.1: Item 1 - How often do you receive unsolicited email?

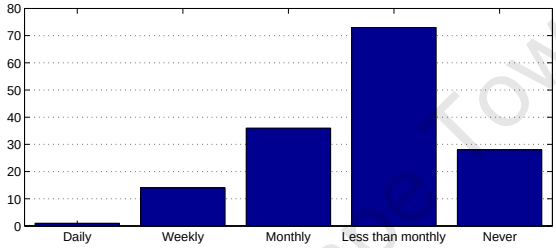

Figure 4.2: Item 2 - How often do you receive unsolicited phone calls on your cellphone?

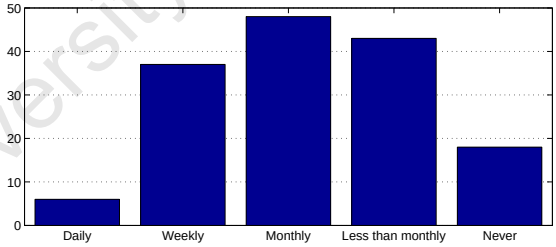

Figure 4.3: Item 3 - How often do you receive unsolicited SMSs?

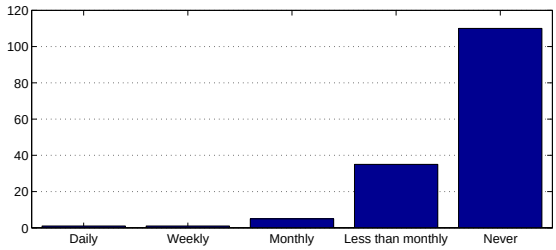

Figure 4.4: Item 4 - How often do you receive unsolicited MMSs?

weekly unsolicited calls and only one respondent claimed to receive such calls on a daily basis; 18% reported never receiving unsolicited calls.

The numbers reflected above are most likely all telemarketing attempts which has only recently gained popularity in South Africa, and at present the literature shows that there are few reports of robo-calling in South Africa. The low penetration of voice spam is most likely due to the current high costs of telephony in South Africa due to many years of a government-protected telecommunications industry.

4.5.3 Cellular Telephone Text Messaging Spam

Cellular telephone text messaging, or SMS (Short Message Service), is a popular mechanism for sending text messages of up to 160 characters over a GSM network. Recently, however, this service has been exploited by spammers looking to deliver advertising directly to users mobile handsets. This is evident by the relatively large number of recipients reporting receiving spam text messages: 4% of respondents reported receiving unsolicited messages on a daily basis, 24% reported receiving these messages at least weekly, 32% reported receiving monthly and 28% less than monthly; only 12% claimed that they never receive unsolicited text messages.

The results indicate that unsolicited text messaging is a popular marketing mechanism in South Africa. This is most likely due to the relatively low cost of sending bulk text messages compared with voice calls. In South Africa bulk messaging companies charge approximately US$0.03 per message [78] - a price that is apparently highly attractive for marketers looking to deliver a message instantly to the user's handset.

4.5.4 Multimedia Messaging Spam

MMS (Multimedia Messaging Service) allows users to send images, video, audio and rich text to one another over the GSM network. Unlike SMS, users are not restricted to only 160 characters per message. MMS has never seen the same popularity as SMS and this is reflected in the comparatively low amount of MMS spam reported in the survey. Only one respondent each reported receiving MMS spam daily and weekly; 3% reported receiving monthly MMS spam and 23% reported receiving less often than monthly; 72% reported never having received an unsolicited MMS.

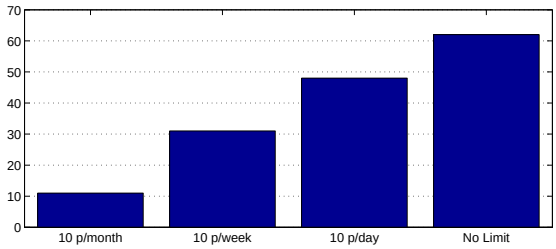

Figure 4.5: Item 5 - How many unsolicited phone calls would make you stop using your cellphone altogether?

These results indicate that MMS spam has not received the same attention from marketers as SMS. This may be surprising for some considering the rich advertisements that can be conveyed via a multimedia message. However, the lack of MMS spam can most likely be attributed to there being fewer bulk MMS providers and the much higher costs involved; an MMS message may cost as much as three times the price of a bulk SMS.

4.6 Spam Tolerance (Items 5 and 6)

In the second section of the survey the respondents were asked how much spam they would tolerate before abandoning their cellular telephones altogether. The aim of this section is to measure the loss of utility that the user experiences because of unsolicited communications. Spam tolerance was examined with these two questions:

- How many unsolicited phone calls would make you stop using your cellphone altogether?
- How many unsolicited SMSs or MMSs would make you stop using your cellphone altogether?

The respondents were asked to answer with one of the following four responses: *10 per month*, *10 per week*, *10 per day*, or *I would not stop using my cellphone*. The feedback from the above questions are shown in Figure 4.5 and Figure 4.6.

4.6.1 Tolerance For Unsolicited Voice Calls

The survey respondents are either surprisingly tolerant of receiving unsolicited calls or they value the utility of their cellular telephones very highly. This is

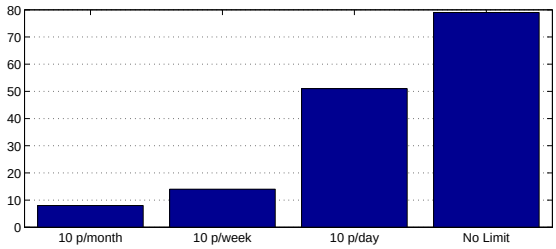

Figure 4.6: Item 6 - How many unsolicited SMSs or MMSs would make you stop using your cellphone altogether?

evidenced by the fact that 41% of respondents stated that they would not stop using their mobile handsets regardless of the amount of unsolicited calls they received; 32% stated that they would not use their phones if they received ten junk calls per day; 20% said that ten per week was too much and 7% said they would not even tolerate ten per month.

It is clear that some people value their privacy very highly while others will hardly tolerate any intrusion whatsoever. This result should be of concern to network providers who stand to lose customers should there be voice spam allowed on their networks.

4.6.2 Tolerance For Unsolicited Text and Multimedia Messages

The survey respondents clearly show a higher tolerance for spam text and multimedia messages than they do for unsolicited voice calls. 52% of respondents said that they would not stop using their mobile telephones irrespective of the number of junk message received; 32% indicated that they would stop using their handsets if they received ten junk messages per day; 9% said that ten per week was their limit and only 5% stated that they would not tolerate ten junk messages per month.

This result is not surprising as text messages are far less intrusive than voice calls. For a start they do not need to be tended to immediately as with the case of voice spam. It is also possible that the anonymity of text messaging makes the intrusion appear less severe.

4.7 Attitudes Towards Spam (Items 7-16)

Items 7 to 16 made up section B of the survey that aims to gauge negative attitudes towards spam. The following questions were posed and the respondents were asked to answer either *strongly disagree*, *disagree*, *agree* or *strongly agree*.

- The amount of unsolicited cellphone communication (calls / SMS) is getting worse.
- I do not mind receiving unsolicited (i.e. "Junk"-) email.
- Receiving unsolicited cellphone calls bothers me.
- I feel bothered by unsolicited SMSs.
- I do not mind receiving unsolicited MMSs.
- Unsolicited calls and SMSs reduce the enjoyment I receive from using my cellphone.
- I find unsolicited cellphone advertising informative.
- I would tolerate cellphone spam if I receive call credits or airtime in return.
- I am concerned with the ease at which access to my cellphone details is available to spammers.
- Cellphone spam can trick the recipient into accepting hidden costs for such calls / SMSs.

For brevity, only the responses for items 9, 10 and 12 are shown in Figure 4.7, Figure 4.8 and Figure 4.9 respectively. The results show that the majority of respondents either agree or strongly agree that unsolicited calls and SMSs bother them. Furthermore, the majority of respondents agreed that they enjoyed their cellular telephones less because of spam.

In order to gauge the respondents overall attitude toward spam the scores from each of the answers are summed. Each answer correlates to a score between one and four and there are ten questions in this sections in total. Therefore, the minimum score possible is 10, indicating no negative attitude

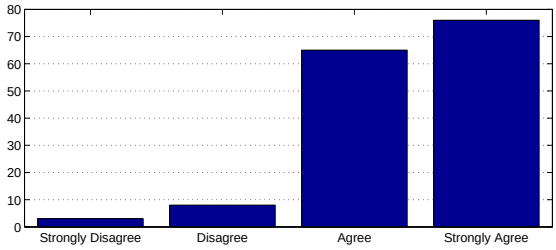

Figure 4.7: Item 9 - Receiving unsolicited cellphone calls bothers me.

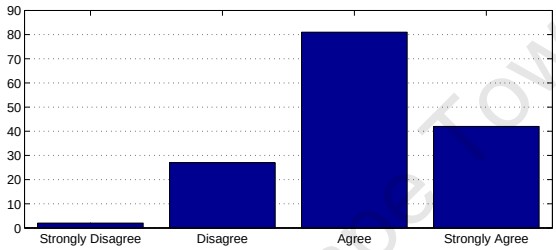

Figure 4.8: Item 10 - I feel bothered by unsolicited SMSs.

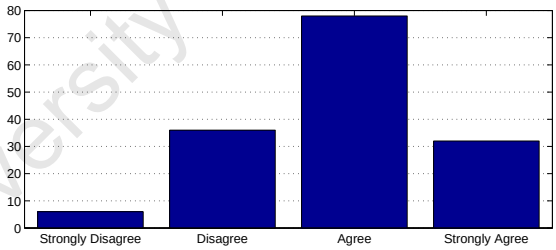

Figure 4.9: Item 12 - Unsolicited calls and SMSs reduce the enjoyment I receive from using my cellphone.

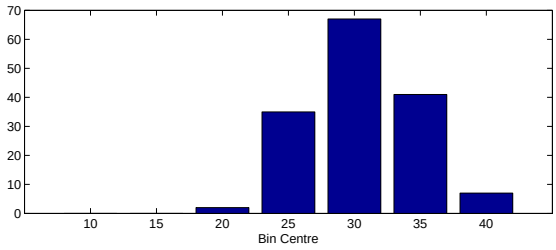

Figure 4.10: Negative spam attitude score frequencies.

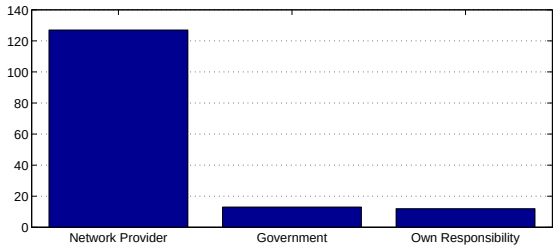

Figure 4.11: Item 17 - Who, in your opinion, is responsible for managing cellphone spam?

toward spam, and the maximum score is 40, indicating high negative attitude. A histogram representing the frequency distribution of all respondents spam attitudes is shown in Figure 4.10. The mean attitude score was 30.68 with a standard deviation of 3.97. The results indicate that in general there is a very high negative attitude towards multimedia spam, with several participants scoring the maximum negative attitude score of 40.

4.8 Accountability (Item 17)

Respondents were asked about spam accountability in the final section the survey with the following question:

- Who, in your opinion, is responsible for managing cellphone spam?

In many countries there is a distinction between the network operator and the service provider, and generally the customer deals only with the service provider. However, in South Africa this distinction is blurred somewhat as the major network operators tend to also act as the service providers and as such the general public do not usually distinguish between the two. Therefore the survey refers to the term network provider to describe this amalgamated entity.

Three possible answers were offered: *My network provider*, *government (legislation)*, or *it's my own responsibility*. The responses are represented graphically in Figure 4.11.

The overwhelming feedback from the survey is that network providers are responsible for protecting users against spam with 84% of respondents in agreement with this. 9% state that government should be held responsible, and remarkably 8% of respondents feel that preventing spam is their own responsibility.

4.9 Discussion

The survey outlined in this chapter provides insight into the current spam problem and the general attitude of consumers towards spam. The survey finds that currently there is a high prevalence of spam in South African mobile networks, with 28% of participant claiming to receive unsolicited SMSs at least weekly. The large amount of voice call advertising is highlighted by the fact that over a third of participants receive unsolicited telemarketing at least monthly.

It also finds that a large proportion of consumers will abandon their cellular phones if the problem escalates further out of control. The participants showed a greater tolerance for text-based messaging than voice calls. Still, almost half of the participants would cease to use the service if they received ten unsolicited messages per day. This is not an unlikely scenario considering the current state of email spam.

In general the participants showed very high negative attitude towards spam indicating that this is not an issue to which consumers will simply become accustomed. The invasive nature of multimedia spam most likely accounts for the high negative attitudes reported. While consumers are increasingly becoming accustomed to email spam it is unlikely, considering the high negative attitudes observed, that they will become accustomed to multimedia spam. The survey finds that this is an issue that frustrates mobile users, therefore it is unfortunate that thus far the multimedia spam problem has been overlooked. Thus the need for this work, and indeed future work in this field, is fully justified.

The survey further finds that the majority of consumers hold their service providers accountable for preventing spam and protecting their privacy. Thus one can infer that service providers risk losing customers if they do not provide suitable security measures in their networks. While the evolution to IMS might lead to greater opportunities for spamming it also presents an opportunity to put into place preventative measures against it. Service providers must utilise the transition to next generation architectures as an opportunity to provide safeguards against spam.

Chapter 5

Robo-call Experiment

Recently there have been significant advancements to the state of the art in detection and prevention of SIP-based robo-calling. Chapter 2 reviewed some of the most widely accepted robo-calling prevention mechanisms, some of which are based on variations of the Turing test. A problem with the Turing test is that it requires the call to be answered before the test can be performed. This can consume resources in the terminating network and cause frustration for legitimate callers. One solution to this problem is the grey-levelling mechanism proposed by Shin *et al.* [75] that blocks callers based on the volume of calls made - regardless of whether the caller is human or not. This is not an unreasonable proposition since it would be difficult to make a significant number of unsolicited calls without the use of a machine and a pre-recorded message.

However, the grey-levelling solution assumes that the call-blocking proxy has perfect knowledge of the caller's history. One drawback to this assumption is that the call-blocking proxy must be located on the spammer's home network in order to obtain a full picture of the caller's history. In reality it is unlikely that this will be the case, since unscrupulous network operators that harbour spammers will not block outgoing spam. A more likely scenario is that the spammer will operate from different network to the recipient, possibly in a different country and legal jurisdiction. Thus, the spam-blocking proxy will need to be located on the terminating network of the call. Knowing this, the logical spammer will target several different networks at a time in order to avoid sending an excessive number of requests to any one network. Unfortunately this leads to the spam-blocking proxy on each terminating network compiling

an incomplete call history of the spammer.

5.1 Hypothesis

The thesis proposes that it is possible to identify suspicious calling patterns derived on an analysis of the previous call durations of each particular caller. This hypothesis is founded on the assumption that humans will react in a similar fashion when subjected to a pre-recorded message and hence some form of recognisable pattern will emerge from the call history. Therefore, by simply measuring the length of the calls over a period of time it is possible to identify certain characteristics that imply abnormal calling patterns.

The benefit to this method over the gray-levelling solution is that irregular calling patterns can be detected at the terminating network regardless of how many other networks have been targeted by the spammer. The spammer's call duration history will look similar irrespective of the numbers of calls made to the particular terminating network per hour.

The hypothesis also states that the spammer's call duration history is sufficiently dissimilar to that of a legitimate caller's such that an automaton might be able to distinguish between them. It should be possible for the automaton to assign a spam likelihood score to the current call based purely on the past calling behaviour of the caller.

5.2 Motivation for Experiment

Research into IP spam is still very much in its infancy and therefore there is little data available by which to test hypotheses. Simulations are not an effective tool for this type of data collection as this problem deals with humans who are notoriously unpredictable, thus the only solution is a practical data collection exercise. Therefore, this thesis introduces the robo-call experiment that aims to collect data about common human behaviour when subjected to an automated caller. The robo-call experiment has two distinct goals. First, it aims to measure the typical amount of time that a human will listen to a pre-recorded message before terminating the call. Second, it aims to evaluate the effectiveness of voice spam by measuring the user response ratio.

5.3 Limitations

The experiment is not designed to provide conclusive data but rather to obtain an insight into human behaviour when confronted with robo-calling. While the experiment suffers from several limitations the data collected is invaluable for this research. The first limitation is the unavailability of an IMS network for testing, and as such the experiment was conducted over a fixed PSTN network. The second limitation is the limited available sample size, mostly due to the ethical considerations of this type of research. While the aim of the experiment is to further knowledge in this field in order to better protect users from spam, the experimental process itself can essentially be considered spamming and is highly intrusive. Hence, at the request of the ethics committee governing this research, the sample was limited to a relatively small number of UCT (University of Cape Town) staff and post-graduate students. The third limitation is the lack of previous voice spam attacks from which to derive the pre-recorded message.

In order to obtain statistically significant results this experiment would need to be conducted on a scale several orders of magnitude larger. The evaluation does not consider the different behaviours of persons from various social backgrounds, and how the type of pre-recorded message affects the results. Furthermore, it is not known to what level users have experienced voice spam in the past.

5.4 Methodology

The experimental methodology of the robo-call experiment is detailed below. All parts of the methodology were approved by the relevant ethics bodies.

5.4.1 Pre-recorded Message

The message stated that the recipient's telephone number was drawn randomly and that they may or may not have won some sort of prize. The message then congratulated the callee and listed the terms and conditions of the contest. The call recipient was directed to visit a website in order to verify if they had indeed won the prize and to find out more about the competition details. The recorded message was a total of 98 seconds in length and was stored as a loss-

less wave file. The recording purposefully included a short silence period at the start in order to allow the recipient some time to perform their greeting. However, the message did not try to obfuscate the fact that it was a recording in any other way. A full transcript of the pre-recorded message is included in Appendix B.

5.4.2 Website

A website was created in order to determine the response ratio to the pre-recorded message. The domain *www.callrewards.co.za* was registered and the DNS entry directed to a machine on the UCT network. UCT assigns all machines on the campus network to a specific IP address range, thus it was possible to block all non-UCT visits to the site by filtering requests from non-conforming source IP addresses. Furthermore, all visitor IP addresses were logged and only new visitors were permitted, in order to prevent repeat visits to the website. A cookie was placed on the visitor's machine for added protection assuring that even if the user's IP address changed they were still identified as a repeat visitor. The website informed the visitor that the robo-call was in fact part of experiment, and asked the visitor to please not divulge the address to any colleagues in order to avoid potentially skewing the results.

5.4.3 Robo-call

The experiment was performed with the windows dialler program over the local campus telephone exchange. Telephone numbers were selected completely randomly from the the UCT internal telephone directory. The dialler played the pre-recorded message as soon as the call was connected. The call duration is defined as the elapsed time between the call answer and call termination. If the call was not terminated by the remote user before the end of the pre-recorded message then it was terminated by the automated dialler.

A total of one hundred successful calls were made. Successful calls are those that are answered within 20 seconds and are not redirected to a voice-mail system or fax machine. Unsuccessful calls, discontinued numbers and calls to busy numbers are not relevant to the study and are therefore omitted from the results.

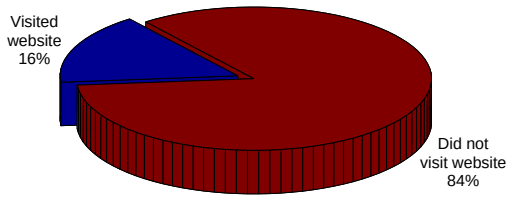

Figure 5.1: Robo-call experiment - response ratio

5.5 Results

The two useful results that can be obtained from the experiment are the response ratio and the typical call duration of the robo-calls.

5.5.1 Response Ratio

Of the one hundred successful calls there were sixteen unique hits on the website. All of the hits were first time visitors from within the allowed IP address range. The hits on the website always followed very shortly after a batch of robo-calls, usually within minutes. It can be said with a fair degree of certainty that there were no errors or extraneous factors that could have affected this response value significantly.

The 16% response ratio is interesting when compared to the tiny response ratio that email spammers currently achieve. The good response ratio is likely due to several factors. For example, the experiment targeted University employees and students while at work and it can be assumed that most had Internet access and could visit the site almost immediately. One would expect that had the calls been made to mobile subscribers or persons without Internet access that this ratio would be significantly less.

However, one would also expect that a true spammer would use a far more convincing pre-recorded message than simply mentioning that the recipient may or may not have won a competition. In any case it appears that the response ratio is highly attractive to a potential spammer and possibly worth the extra expense over email spam considering the significantly better returns.

Table 5.1: Robo-call experiment - interesting figures.

Total calls	100
Mean call duration	54.54s
Standard deviation call duration	34.06s
Calls ended in first 30 seconds	36%
Calls reaching conclusion	26%

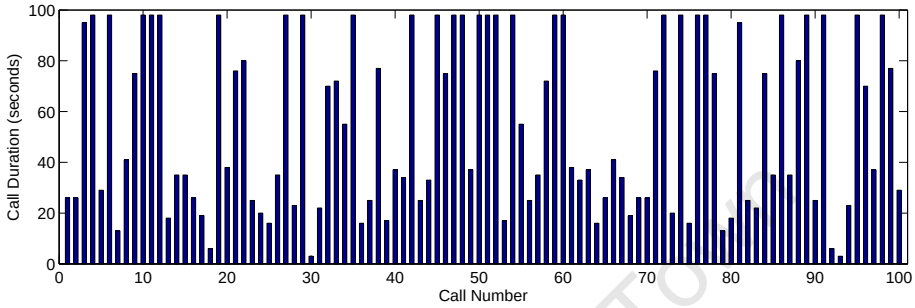

Figure 5.2: Robo-call experiment - call duration results.

5.5.2 Call Durations

Table 5.1 and Figure 5.2 show the call duration results obtained from the experiment. It is interesting to note that over a quarter of all recipients listened to the entire 98 second message resulting in the call being terminated by the automated dialler, and on average recipients would listen to the message for almost a minute. There may be several reasons for this surprisingly patient response to the intrusiveness of an automated call. It is possible that the recipients were not accustomed to receiving automated calls and were intrigued to listen further. Perhaps boredom at work or a genuine belief that they had won a prize accounts for the results. However, it may be the case that traditional phone etiquette does not condone hanging up the call while the other party is still speaking, hence the large proportion of recipients that listened all the way to the end of the message.

As with the response ratio these results are also encouraging for the potential multimedia spammer. It is clear that a voice call is a far more powerful form of communication than email. This is no doubt the reason that robo-calling has become so popular with political parties in the USA [59].

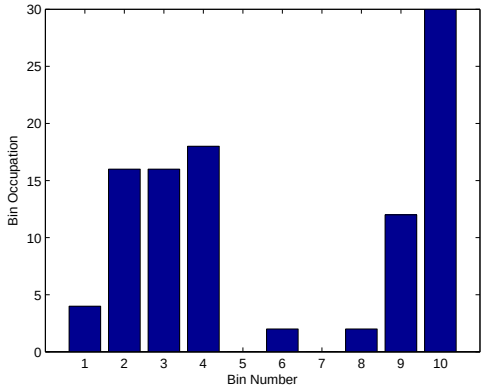

Figure 5.3: Robo-call experiment - histogram of call durations.

5.6 Analysis

The hypothesis earlier in the chapter stated that it may be possible to distinguish between automated callers and legitimate callers by examining the call duration history. The best manner by which to test this hypothesis is to construct a histogram of the call duration results and try to find any distinguishing characteristics; this histogram is shown in Figure 5.3. The histogram is divided into ten equally spaced bins in order to give a smooth representation of the data.

5.6.1 Histogram Analysis

From the histogram it appears that the most heavily populated bins lie in two distinct areas: in the first half of the call and right at the end of the call. This implies that it is possible to group people into one of two categories. The first category is the individual who does not want to be bothered by the automated caller and terminates the call as soon as they realise that it is unimportant. It takes some individuals a little longer than others to deduce the call is spam, but most have decided within the first 30 to 50 seconds that they are not interested in the call. Very few people hang up the call right at the beginning implying that it does take a fair amount of time in order to establish the legitimacy of a call.

The second category is the individual that listens to the end of the call either due to interest or some type of misplaced courtesy for the automated caller. It is unknown whether the individuals would be willing to listen to

a message longer than 98 seconds but such a long message would surely not be employed by a spammer looking to minimise costs and maximise the total number of calls made.

5.6.2 Bimodal Probability Density Function

The histogram represented above can be modelled as a Probability Density Function (PDF). Single-mode PDFs cannot take into account the two different types of individuals identified above therefore a bimodal distribution must be employed. The first category of individuals can be modelled with the Gaussian distribution shown in Equation 5.1. Values of $x < 0$ are not considered as a conversation cannot have a negative duration.

$$g(x, \sigma, \mu) = \frac{1}{\alpha\sqrt{2\Pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) \tag{5.1}$$

The second category of individuals can be modelled better with an exponential distribution shown in Equation 5.2. Note that $h = 0$ for all values of x greater than the length of the call x_0.

$$h(x, \lambda, x_0) = \begin{cases} \lambda \exp(\lambda(x - x_0) & x < x_0 \\ 0 & elsewhere \end{cases} \tag{5.2}$$

This yields the combined bimodal PDF, f, described in Equation 5.3. The mixing parameter p determines the fraction of calls that fall within the normal distribution, where $0 \leq p \leq 1$.

$$f(x, p, \lambda, x_0, \mu, \sigma) = p.g(x, \sigma, \mu) + (1 - p).h(x, \lambda, x_0) \tag{5.3}$$

Figure 5.4 shows the original histogram with bin occupancies divided by the total number of calls in order to give a probability distribution, with the bimodal function f overlaid. A least-squares fit is used to determine the parameters of f; details of this fitting process are discussed in subsequent chapters.

5.7 Comparison with Legitimate Callers

The results of the robo-caller experiment yield interesting insights into the behaviour of humans when subjected to voice spam. The question remains whether the call duration history of a spammer is sufficiently different from

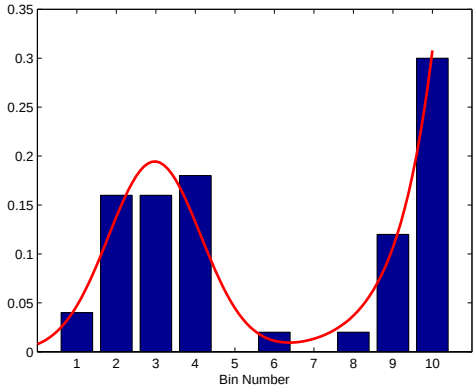

Figure 5.4: The original histogram normalised and overlaid with the bimodal probability density function f.

that of a legitimate caller. In order to solve this question the call histories of legitimate callers must also be examined. For this purpose the phone records of sixteen individuals are analysed in a similar fashion to the analysis performed on the robo-caller experimental results. The purpose of this comparison is to illustrate the difference between the observed call duration analysis of the robo-call experiment and typical call durations of legitimate callers. The goal is to demonstrate that the observed robo-call data is not typical and hence indicates a suspicious calling behaviour.

5.7.1 Data Source

The data is sourced from the call records of a European VoIP provider courtesy of NEC laboratories Germany. The call records were anonymised by the VoIP provider in order to protect the user identities by hashing the caller and callee information. There is an almost zero chance that two caller identities will hash to the same value so while it is not possible to know the identity of the callers, it is possible to recognise which call records belong to the the same callers.

The data was parsed to find callers with a history of over one hundred non-zero length calls. The aim is not to compare the robo-call data results with a very large data set of legitimate callers but rather to demonstrate a trend in the legitimate caller data. Sixteen callers were chosen at random from the qualifying data set. The robo-caller experiment shows the result of 100 calls. Therefore only the last 100 call records of the qualifying callers are examined for comparison.

5.7.2 Outliers

In the analysis of the robo-call durations above a histogram was created from the data with equally spaced bins. The problem with call data from legitimate callers is that periodically a caller will make a very long call. It has also been observed that sometimes in the call records there are abnormally long calls, probably as a result of a call timer malfunction. For most users these long calls do not fit well within the normal calling behaviour patterns, and they badly distort the produced histogram. The solution to this problem is to remove the data outliers and focus on the general calling behaviour of the caller.

There is no definitive method for the removal of outliers from data, indeed some argue that the best method is usually by human inspection of a plot [36]. However, an automated method was utilised that suffices for this situation - the values that fall outside two standard deviations from the median value are removed. This process typically only removes two or three call records out of a hundred but greatly improves the resultant histogram plot.

5.7.3 Call Duration Histograms

The histogram plots for the sixteen random callers are shown in Figure 5.5 and Figure 5.6. A quick observation of the plots reveals that the lower bin numbers have the greatest occupation. To further illustrate this phenomenon a standard exponential PDF $y = \lambda e^{-\lambda x}$ has been overlaid onto the plots. For simplicity this overlaid plot has not been fit to the data and the value of λ is kept at unity.

5.8 Discussion

The hypothesis at the beginning of the chapter stated that it may be possible to distinguish automated callers from legitimate callers by examining their respective call histories. In order to confirm this hypothesis a robo-call experiment was devised and implemented in a practical setting. The result of the experiment shows that humans react in different ways when receiving an automated caller - some terminating the call after a short while and others listening to the whole message. This behaviour can be best viewed as a normalised histogram of call durations.

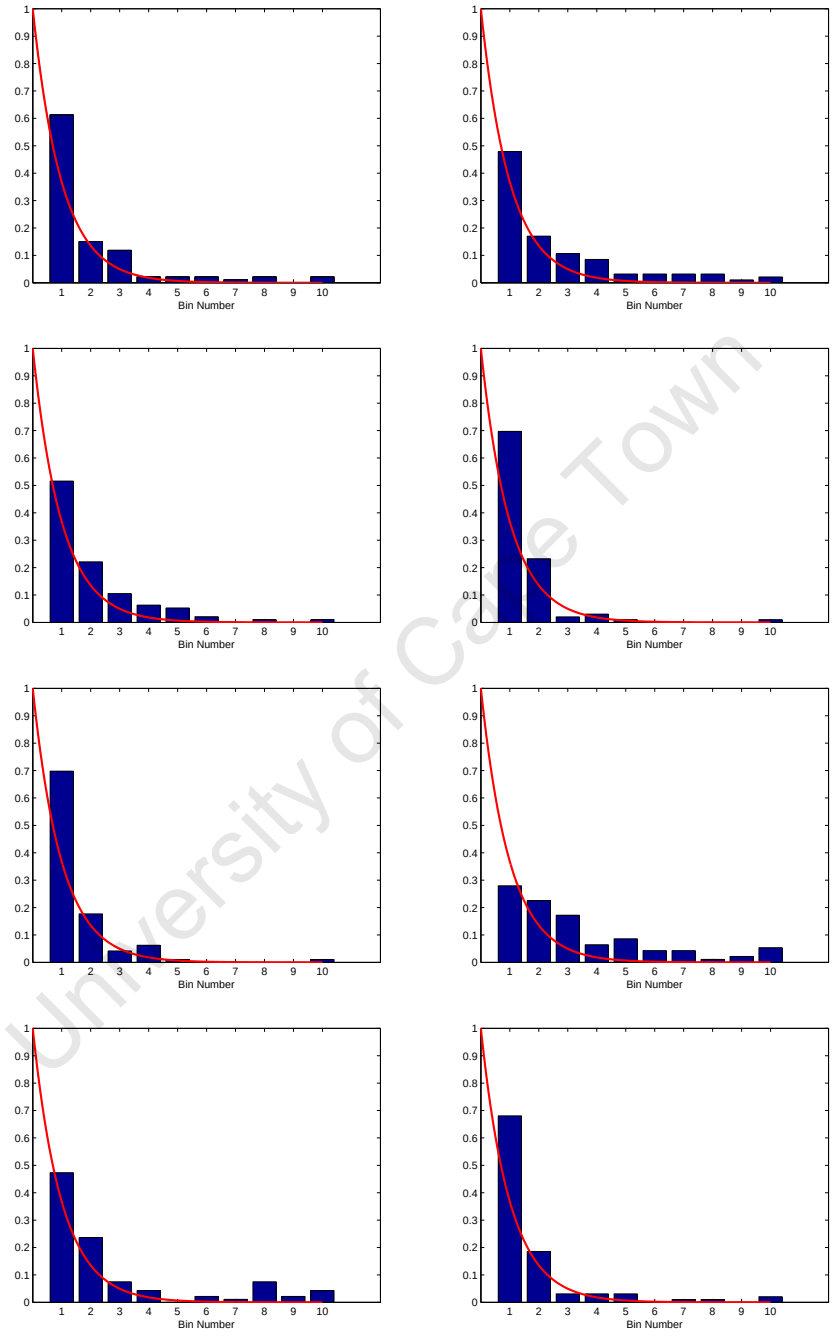

Figure 5.5: The call duration histograms of random legitimate callers (excluding outliers) overlaid with an exponential probability density function (Part I).

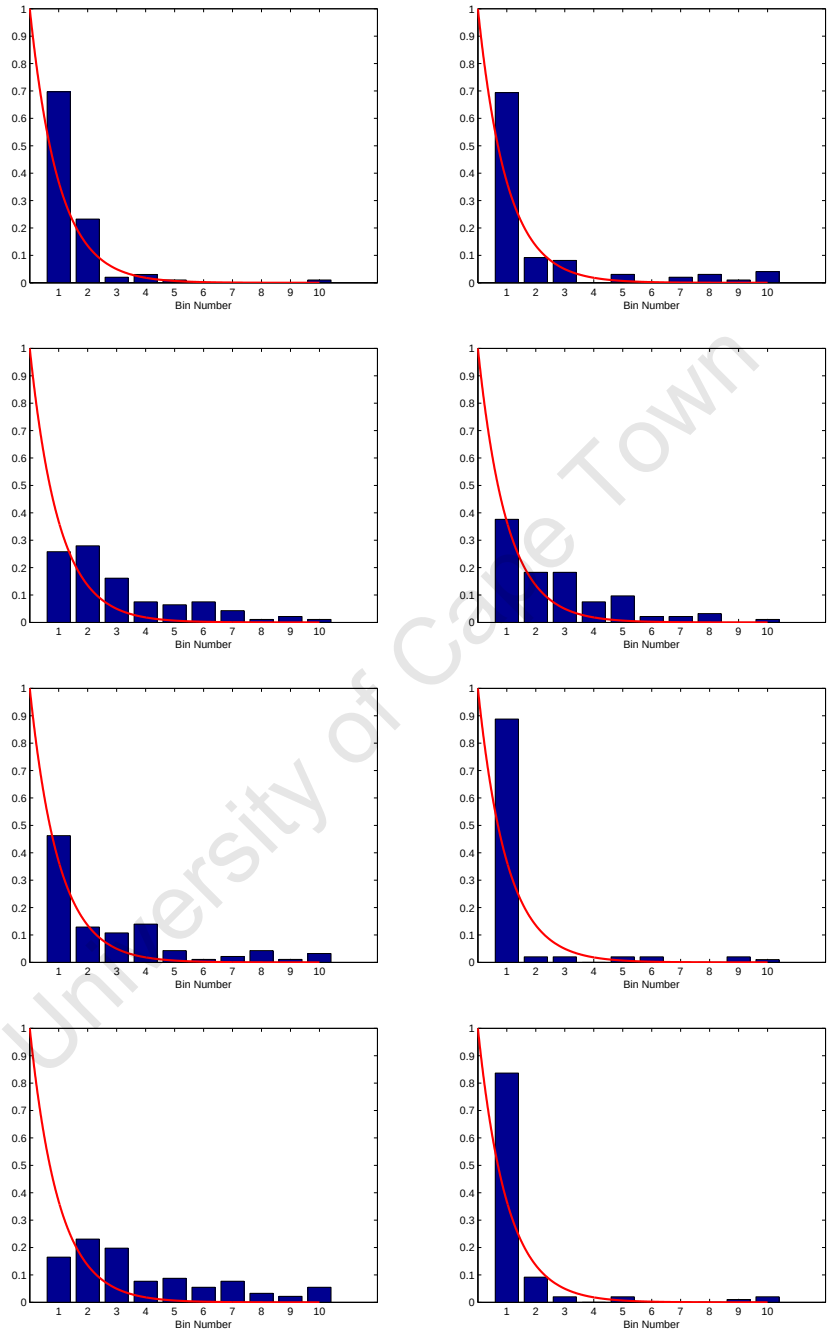

Figure 5.6: The call duration histograms of random legitimate callers (excluding outliers) overlaid with an exponential probability density function (Part II).

This result is only useful if the robo-caller's duration history is distinguishable from that of a legitimate caller. The call records from a large European VoIP provider were obtained and examined in order to compare the typical calling behaviour of a normal caller to the experimental results. Sixteen random caller were selected and their last 100 call records were extracted. After eliminating the outlier data a histogram was composed in the same manner as the robo-caller histogram.

The histograms of the legitimate callers show a clear trend. An exponential probability density function is a close fit to every call duration history examined. It is highly unlikely that the sixteen random callers chosen are all exceptional cases. Therefore it must be reasoned that their behaviour is not significantly different from the behaviour of the majority of the users of this VoIP service. The most likely explanation is that typical human behaviour is to make many more shorter calls than longer calls.

Thus it can be concluded that the call duration history observed by the automated caller is not typical of a legitimate caller. This result confirms the hypothesis that an examination of the call duration history can distinguish an automated caller from a legitimate one.

Part III

Proposed Solution

Chapter 6

Spam Prevention Architecture

The previous chapters highlighted the problem of spam in IMS networks and the current state of the art in SIP spam research. As yet, there is only limited literature dealing with the problem of spam in IMS networks and thus far the 3GPP has made few attempts to address the issue. This chapter proposes an architecture for the prevention of spam specific to IMS networks in order to fill this research void.

The chapter begins with a list of design considerations that are formulated from the literature review. It then describes the multi-layered solution proposed by this thesis and a generic IMS architecture that supports the solution. The proposed architecture details only the logical elements of the system. The goal of the architecture is that its eventual realisation should be implementation-specific.

6.1 Design Considerations

The literature review and findings from the previous chapters show that the design of an IMS-compliant spam protection solution should take several fundamental considerations into account. As with any design there are always trade-offs to be made, and hence some design considerations must be given precedence over others. This section details these considerations ordered from the highest priority to the lowest.

6.1.1 Discourage Spam and Ensure Privacy

An important design criterion is that the system discourages the sending of voice, video and text-based spam. The IMS potentially supports many different types of services but consideration of all types of spam presents too large a scope for the thesis. While the design is limited only to these basic services, the system should be extensible in order to cater for new services in the future.

It is unreasonable to expect any mechanism to block all spam completely, especially since it is hard to say exactly what constitutes a spam message. The most reasonable goal of a spam prevention architecture is to reduce the profitability of mass unsolicited sessions thus discouraging the practice [65]. A further goal is to ensure user privacy by providing the facility to automatically block unwanted callers and allow known callers.

6.1.2 Unobtrusive

The design should ensure that the spam prevention mechanism is unobtrusive to the network users. One of the primary reasons for blocking spam is because it frustrates users; thus the whole exercise is futile if the solution itself is overly obtrusive. Turing tests and authorisation frameworks are known to be highly obtrusive. One of the biggest drawbacks to the Turing test is that it requires human interaction that can become frustrating to users over time. For example, if a user is required to identify a distorted image or garbled voices every time they wish to make a call this might seriously affect their willingness to use the service. Furthermore, Turing tests can be difficult to solve for persons with poor vision or hearing, or for users that are not familiar with the language.

Authorisation frameworks that require users to be authorised before they are allowed to initiate sessions can also be highly obtrusive. Users may be required to respond to several authorisation requests each day, which may prove to be more frustrating than the spam itself.

An important consideration for a system used for general telephony is to ensure emergency calls are unimpeded [17]. Emergency calls are often government regulated and the requirements for support differ from country to country. In some countries the telephony device is required to make emergency calls without the presence of a smart card and in other countries networks are even required to handle emergency calls from non-subscribers. Countries

around the world usually use different numbers for their emergency services, causing problems for call filters. A call filter that is programmed to automatically allow emergency calls will not work if the user is roaming in a different country. Any spam prevention solution should ensure that it provides the least possible intrusion to legitimate callers and, most importantly, it cannot block an emergency call attempt.

6.1.3 Conformance to Standards

Several standards bodies have been integral in creating the IMS specifications. The first priority of the design is conformance to the 3GPP specifications, specifically regarding the architecture and signalling of the network elements. As the IMS reuses the specifications of the IETF, the design is also focused on these specifications.

The specification document describing the complete architecture of the IMS is 3GPP TS 23.228 [57]. The IMS core network's primary concern is to handle authentication, authorisation and call routing, therefore it does not provide any services to the user, apart from basic voice and video calling. Rather, the 3GPP has specified that service execution occur in application servers. The IP Multimedia Subsystem (IMS) specifications do not define how advanced services should be provisioned, but do provide an interface which connects to the application layer. This is known as the ISC (IMS Service Control) interface and is based on SIP. The ISC interface connects SIP application servers, OSA/Parlay gateways, and CAMEL gateways to the S-CSCF. CAMEL is generally a technology relating to legacy circuit switched intelligent networks so for brevity it is omitted from this discussion. The Sh interface is based on Diameter and connects the application servers to the HSS (Home Subscriber Server).

An API (Application Programming Interface) defines abstracted methods for interacting with a network. Often these APIs do not define on which programming language they should be implemented so that they can be ported across platforms. While 3GPP specifies what its requirements are it is up to the joint API groups to deliver suitable solutions for their service needs. APIs aims to allow developers to create applications that are independent of the underlying network technology. The Parlay APIs include call control, conferencing, interaction and charging. The API is specified in UML but

mappings are provided for CORBA IDL, Java (JAIN) and WSDL. Parlay is an open standard and is jointly defined by Parlay group, ETSI and 3GPP. However, in 3GPP it forms part of the Open Services Access (OSA) group and hence the common term OSA/Parlay. Essentially Parlay and OSA are the same thing even though the same standards may be assigned different version numbers by the different groups.

Parlay X, released in 2003, is a simplified version of the Parlay API. Parlay had mappings to CORBA and Java but Parlay X addresses the need of web developers. This is clearly an attempt to further increase the number of developers capable of providing services. While Parlay X provides less functions to the developer its reduced complexity is very attractive to network providers and it has already been adopted by large Telecoms such as BT and Sprint. Parlay X specifications are also specified by ETSI and 3GPP but unlike Parlay only one set of specifications is released.

The 3GPP also specifies native SIP application servers that do not require a gateway interface or specific API. Native SIP services do not enjoy the benefits brought about by APIs above such as rapid development, security or network abstraction. However, native SIP servers can be faster to deploy for smaller projects as no execution environment is required apart from a simple SIP stack. Moreover, native SIP servers can also be more flexible as the developer is not limited by restricted API calls. Thus, due to the limited scope of research projects, native SIP servers adhering to IETF SIP specifications [68] and the IMS call control protocol [55] are most likely to be the optimal architecture.

6.1.4 Resistance to False Positives

False positives are defined as legitimate sessions that are incorrectly identified as spam, whereas a false negative is defined as a spam message mistakenly marked as legitimate. Paul Graham in his essay titled *A Plan for Spam* [30] emphasises that missing a single legitimate communication is an order of magnitude worse than receiving spam. Graham also states that a user is less likely to notice a legitimate email in their spam folder if they receive a large amount of daily spam. Furthermore, the better the filter is, the more likely the user is to trust it and will ignore all messages marked as junk. Thus, it is imperative that any solution err on the side of caution by minimising false positives, even at the expense of increased false negatives.

6.1.5 Scalability

Scalability is of particular importance when dealing with network security due the risk of a DoS (Denial of Service) attack. The most simple solution for ensuring scalability is to distribute the network elements across several different servers and provide fail-over support. In a SIP environment scalability often comes at the expense of reduced functionality. For example, SIP servers can either act in a stateful or stateless manner. Stateful proxies must maintain transaction state (as apposed to call state), whereas stateless proxies do not. There are distinct benefits to stateful proxies as they can facilitate call forwarding and other features. However, stateful proxies do not scale as well as stateless proxies as they are required to store the state of all transactions. According to the specifications SIP proxies must be stateful if they use TCP, multicast or provide forking, otherwise they can be stateless.

It is possible for both stateful and stateless proxies to maintain call state through use of the *Record-Route* header. In SIP all responses for the same transaction return via the return path of the request. The *Record-Route* header ensures that all future transactions that are part of the same dialog also traverse the SIP proxy. This also negatively affects scalability as now each future transaction has more hops to traverse in order to reach its destination.

Scalability also relies on load balancing and redundancy, and in SIP these can be achieved using simple DNS mechanisms. DNS allows for a single domain to resolve to several servers, therefore a simple DNS solution can help the S-CSCF to choose the most appropriate application server. An efficient security solution must be distributed, maintain as little state as possible, and utilise load sharing and fail-over mechanisms.

6.1.6 User Configurability

Different users have different privacy requirements. For example, parents may wish that their children are part of what Tschofenig *et al.* [85] describe as a closed group whereas call centres or businesses would most likely be part of an open group. Only the users themselves will be able to choose exactly the level of spam protection and privacy that they require. Thus it should be possible for the end user to fine-tune their privacy rules and change these rules at a later date, preferably without the assistance of the network operator.

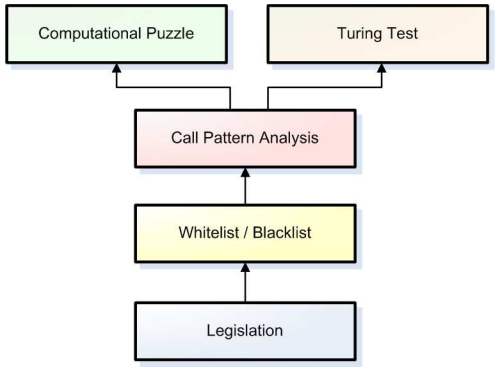

Figure 6.1: Proposed multilayered spam prevention architecture.

6.2 Proposed Multilayered Solution

The thesis proposes a multilayered spam prevention architecture founded on the review of current literature, the design considerations above, and input from fellow academics. Presently there is no silver bullet for solving the spam problem, hence the architecture offers several layers of protection in order to discourage spamming behaviour and to ensure user privacy in the IMS.

The layers of the architecture are depicted in Figure 6.1. The four tier protection model's first line of defence is the least intrusive to users, with the last layer of protection being the strongest but most intrusive.

6.2.1 Legislation

The first layer of the spam prevention model is government legislation. Legislation has been shown to greatly reduce the number of telemarketing calls in the US. A Harris Interactive survey released in 2007 shows that 72% of US adults have signed up the for the Do Not Call Registry [38], and of those that signed up 59% stated that they receive far less unsolicited calls, 14% stated they receive a little less, and a full 18% claimed that they now receive none at all. 6% stated that the amount of telemarketing had not changed for them and only 1% claimed to be receiving more. Political parties, charities and those conducting surveys are unfortunately exempt from the registry in the US and this accounts for 41% of people on the registry stating that they had received one or more calls from survey companies. Overall the Do Not Call Registry has shown remarkable success in a very short amount of time and highlights the impact that government legislation can have on spam.

Unfortunately the US CAN-SPAM act has not had as much success. A survey conducted a year after the option of the act showed that 28% of users reported receiving more email spam and only 22% reporting receiving less [23]. It is unclear whether the increase in spam was due to poorly written legislation, a lack of enforcement or simply the worldwide trend of increasing spam levels. However, there was a noticeable decrease in the amount of pornographic email spam after the adoption of the act so in one respect the legislation was successful.

While telecommunications operators cannot directly control legislation they are able to lobby government to outlaw spamming behaviour. The survey in Chapter 4 found that only 9% of the participants feel that the government is responsible for reducing spam. However, legislation can reduce the amount of spam with little or no intrusion to end users and is therefore a vital tool in dealing with the problem.

6.2.2 Whitelist / Blacklist

The second tier of protection comes from specific admission lists. Blacklists are used to bar unwanted callers and whitelists are used to ensure that known callers are guaranteed to be connected. Users are required to manually select the addresses of the persons on their whitelists and blacklists.

The blacklist may not be permanent however, as it is possible to add and remove users from the lists dynamically. For example, if a caller that does not appear on either the whitelist or blacklist (an unknown caller) fails the Turing test, that caller may then added to the blacklist for 15 minutes. Other transgressions may also result in a caller being added to the blacklist temporarily, such as a highly suspicious calling behaviour.

The whitelist is also not fixed. For example, if a user makes a call then the address of the recipient may be added to the caller's whitelist so that any return calls will be automatically accepted. The exact behaviour of the whitelist and blacklist is user configurable depending on the level of protection required.

6.2.3 Call Pattern Analysis

The third layer of protection comes from CPA (Call Pattern Analyser) modules that aim to detect suspicious calling patterns. The term *call* is used instead

of *session* for congruency with the literature, but it should be noted that the mechanism operates on all types of IMS sessions, including voice and video calls, text messaging and push-to-talk.

Three modules are used in the proposed architecture, namely a session volume analyser, session duration analyser and concurrent session analyser. The modular design of the CPA allows new technologies to be added when they become available. Call pattern analysis is performed on each caller's usage history whenever they initiate a session. For example, Bob attempts to initiate a voice call to Alice. However, Bob has already initiated ten calls in the last minute to other users in Alice's network. The CPA in Alice's network flags this behaviour as suspicious, and gives Bob's current session a spam score indicating a high probability that this particular call is spam.

Session volume analysis has already been proven to be an effective tool in identifying suspicious calling behaviour in SIP networks [75]. The analyser can also be used to detect other types of spam such as text and multimedia messages based purely on volume.

The previous chapter discussed the short-coming of a spam detection solution based purely on call volume analysis. Thus a session duration analyser module is incorporated in order to detect spammers that intentionally target several different networks successively, thereby reducing the volume of sessions detected by each of the networks. The session duration analyser forms a PDF (Probability Density Function) of the call duration history. If this PDF resembles a curve of known spamming behaviour then the spam score is increased according to the closeness of the match. This module is only applicable to sessions with a discernible length, such as voice and video calls, and session-based instant messaging. The session duration analysis module is a novel advancement to the current state of spam detection research and is implemented for the first time in this thesis.

The third module utilised in the proposed architecture is a concurrent session analyser. The module detects an abnormal number of concurrent sessions and increases the caller's spam score if the number of concurrent sessions exceeds the amount that would be expected of a legitimate caller. This module is also only applicable to sessions that can occur concurrently.

A weighted average of the three CPA modules is calculated in order to inform further network elements as to the probability of the message being spam. The weighted average is known as the caller's overall spam probability

score (or simply spam score) and is embedded in the session request. The modular architecture allows for additional call pattern analysis components to be plugged into the system if required.

6.2.4 Turing Test / Computational Puzzle

The last layer of protection is provided by either a Turing test or computational puzzle, depending on the nature and severity of the perceived threat. As this layer of protection is highly intrusive it is only used when there is a high probability that a session is spam.

A computational puzzle is used in order to slow down a spammer that is generating an unusually high amount of session setup requests (be they voice calls, text messages, or otherwise), and thus has a high spam score. The computational puzzle requires the calling terminal to solve a processor intensive puzzle before the call can continue. The difficulty of the puzzle is proportional to the spam score calculated by the call pattern analysis and the level of protection required by the individual user. The computational puzzle difficulty dynamically adjusts to the level of the current threat and therefore does not block callers indefinitely.

The Turing test, on the other hand, requires the caller to correctly identify a distorted image that is streamed to their terminal. If the terminal does not have video capabilities then an audio test is used instead. The user is required to input the correct sequence of alphanumeric characters in order to continue with the call, otherwise the caller is informed that the call cannot continue and the call is terminated. The Turing test is only used in rare instances when the session is given a very high spam probability score.

6.3 Solution Architecture

The reference solution architecture is depicted in Figure 6.2. The architecture is comprised of three reference elements, namely the Call Pattern Analyser, Rule Database and Authorisation Engine. As this is a reference architecture the actual physical architecture is implementation specific, therefore the elements may be combined or distributed over a several machines. The elements are shown adjoined to an IMS core network, a user terminal and a spam source.

The solution is always located in the user's home network and only operates

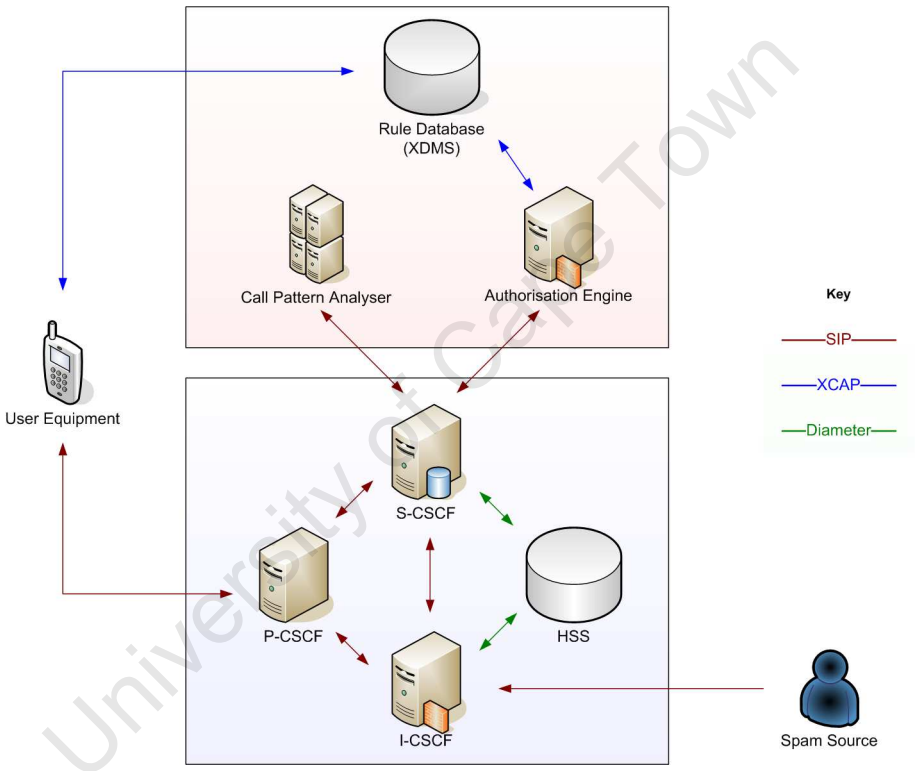

Figure 6.2: The solution architecture depicted with an IMS core network, user terminal and spam source.

on terminating calls for several reasons. Firstly, it is hard to trust that a foreign network will perform correct call pattern analysis and call filtering, or will execute the authorisation policies correctly since there is no pre-existing trust relationship. The foreign network may be harbouring spammers and therefore choose not to provision suitable spam avoidance solutions. Secondly, it is easier for users to upload authorisation policies to a machine located within their home network as they will already be authenticated and authorised. Thirdly, since spam blocking is generally required only when receiving calls and all incoming calls must traverse the user's home network, it is a natural solution to apply filtering in the terminating network. Lastly, since the solution is only applied to terminating calls emergency numbers can be set by the network operator to bypass any authorisation decisions.

6.3.1 IMS Core Control

The spam prevention architecture interacts closely with the core IMS network that provides the call routing functionality. When an IMS terminal initiates a call the signalling must traverse various SIP servers, known as CSCFs (Call Session Control Functions), in order to reach its destination. The call is first routed to a P-CSCF (Proxy CSCF) that may be located on either the user's home network, or on a visited network. The call must then be routed to a S-CSCF (Serving CSCF) that is always located on the user's home network. It is the S-CSCF that performs complex admission control and routing functions in the IMS. If there are several S-CSCFs in the network then an I-CSCF (Interrogating CSCF) first selects an appropriate one to handle the call. Each IMS URI is associated with a User Profile that is stored on the HSS (Home Subscriber Server) and is downloaded to the S-CSCF when a terminal registers on the network. This profile contains authentication and authorisation information, and allows users and the network to specify what services are available, for example the proposed spam blocking service.

Every IMS call is subject to initial filter criteria. These filter criteria are specified in the User Profile, and are also downloaded to the S-CSCF when a user registers on the network. In the proposed architecture all user profiles specify that terminating calls must first traverse the Call Pattern Analyser. This allows the network to build up a detailed history of every user that makes a call to the network.

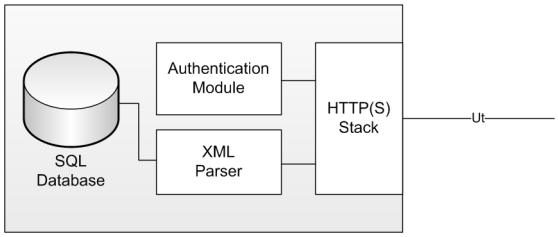

Figure 6.3: Structure of the Rule Database.

However, terminating calls do not necessarily need to traverse the authorisation engine. This is because users may choose for some reason not to use the spam blocking service perhaps because the service is charged separately by the network provider. In this way the proposed architecture can still analyse the calls made to users that have not enabled spam protection, hence improving the effectiveness of the call pattern analysis for other users in the network.

6.3.2 Rule Database

The Rule Database is an XDMS (XML Document Management Server) that provides an XCAP (XML Configuration Access Protocol) [64] interface for interaction with user equipment and other network elements. 3GPP defines the connection to the XDMS as the Ut interface and the protocol as HTTP, as shown in Figure 7.7.

An XDMS is able to store any type of data in XML format, thus it is commonly used for storing service-related information for address books, conferencing applications, instant messaging and push-to-talk contact lists. In the proposed architecture the XDMS is primarily concerned with storage and manipulation of user-specific authorisation rules in an SQL (Structured Query Language) database.

The XDMS is also responsible for other tasks, including authorising the users that attempt to upload or modify any documents, and checking that all uploaded documents are well formed and valid. If an unauthorised user attempts to upload or modify a document, or if the XML of an uploaded document does not parse correctly, then the XDMS is responsible for responding with an appropriate error code. The rule database only accepts XML documents in the form of the spit-policy schema proposed by Tschofenig *et al.* [86], which is based on the well-known common-policy schema [72].

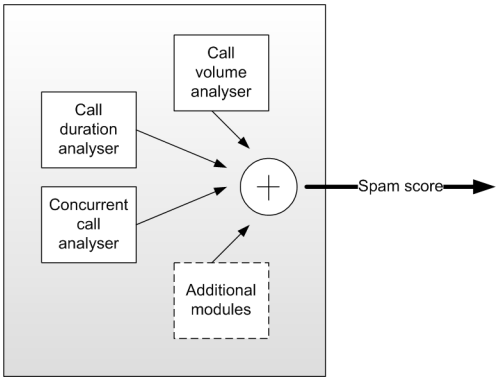

Figure 6.4: Modular architecture of the Call Pattern Analyser.

6.3.3 Call Pattern Analyser

The CPA (Call Pattern Analyser) is an application server that proxies incoming requests. The use of a SIP server is highly recommended as it must perform SIP message manipulation that is not easily supported by either CAMEL or Parlay gateways. Furthermore, the CPA is on the call setup path and must therefore provide rapid response times in order to avoid call setup delays.

The server appends an appropriate spam probability score to its own IP address and inserts this string to the SIP *Via* header, as described by Wing *et al.* [93]. The modular architecture of the Call Pattern Analyser is shown in Figure 6.4.

The server constructs a history for each caller, uniquely identified by their *P-Asserted-Identity* headers. The server not only stores the volume of sessions initiated by each unique caller but also pertinent information about the calls themselves. Several items of data are collected for each session; the type of session (multimedia, text, conference, etc.), the initial INVITE time, the call start time (identified by the user agent server's 200 OK response), the call termination time (identified by the BYE message sent by either the user agent client or user agent server), and a dialog identifier (constructed with the *To*, *From* and *Call-ID* tags). Some sessions, for example pager-mode text messages, may not have a call start or end time in which case these fields are left blank.

The time of the initial INVITE message is recorded for the purposes of the session volume analyser. The session duration analysis module is only concerned with the duration of the call from when it was answered not from when it was initiated, thus the requirement to store the call start time. The call

is stored regardless of whether it receives any provisional or final responses from the user agent server. Therefore, spammers trying many URIs will still increase their spam score regardless of whether the user's terminals are switched on or if they answer the call. Unfortunately, it is not possible to measure the number of calls made to non-existent URIs, as these URIs will have no user profile associated with them and consequently no filter criteria that forwards the call to the Call Pattern Analyser.

The server's analysis modules can calculate their appropriate scores for session volume, session duration and simultaneous sessions from the various items of stored information. The number of session records stored by the server is implementation specific but since each record uses only a few bytes of data storage it is expected that the server can store the history of millions of users without requiring purging of older records. Furthermore, only a few hundred recent call records for each user are required in order to make an accurate analysis of their calling behaviour. This also ensures that the system reacts quickly to spamming attacks by emphasising recent calling trends over past behaviour.

It should be noted that it is desirable to store as many call records as possible to prevent a smart spammer from using several pre-recorded messages of different lengths to fool the call pattern analyser. Of course having to analyse many records will negatively affect the speed of the analyser and hence there is a tradeoff between accuracy and speed.

6.3.4 Authorisation Engine

The Authorisation Engine, shown in Figure 6.5, takes input from several sources in order to make an authorisation decision regarding a new session attempt. The Authorisation Engine may be implemented using any service platform, however, for performance purposes it is recommended that a native SIP server be used.

The server requires three sources of information: the XML rules, downloaded from the Rule Database; the network's default rules, a set of static rules defined by the network operator; and the SIP request itself. The Authorisation Engine may also take auxiliary information deduced by the server itself into account, such as the time of day or day of the week.

There are several header fields that the Authorisation Engine examines

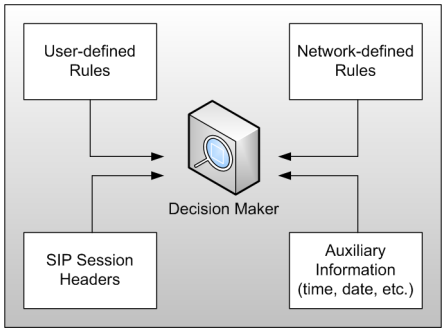

Figure 6.5: The Authorisation Engine makes decisions based on several sources of information.

in the SIP request, including the *Via*, *To*, *P-Asserted-Identity*, *Contact*, and *Content-Type* headers. Recall that the Spam Probability Score is co-located with the Call Pattern Analyser IP address in the *Via* header. The *To* header indicates which XML rules to download from the Rule Database and the *P-Asserted-Identity* header determines the identity of the user that initiated the session. The *Contact* header field shows the IP address of the user agent client that, like the *P-Asserted-Identity*, may also be subjected to rules such as blacklisting. The *Content-Type* header shows what type of session is being started; for example, multimedia, text-message, or otherwise.

The various sources of information are fed to a decision maker that chooses how to handle the session. The authorisation has several options: reply with an error message, allow the call to continue, redirect the call, seek authorisation from the callee, challenge the caller with a computational puzzle, or challenge the caller with a Turing test.

Take for example a call that is received at 10pm on a weeknight from a user that does not appear on the callee's whitelist or blacklist. The caller requests a multimedia session, but the callee's rules specify that only text messages may be received from unknown callers after hours. The Authorisation Engine responds with a 488 error response (Not Acceptable Here) stating that only text messages are allowed; this information is conveyed in the SIP response reason phrase as stipulated by the SIP specification [68]. The caller now sends a text message instead of a multimedia session that is now accepted by the Authorisation Engine and proxied to the callee.

6.4 Discussion

The chapter has detailed a proposed architecture for privacy and spam control in the IMS. The architecture employs a modular design and follows the design considerations as closely as possible. It is envisaged that future work in the field will result in new techniques to identify suspicious calling behaviour. Thus the proposed architecture is extensible, allowing new modules to be inserted and older ones removed if necessary. The architecture does not hinder emergency calls as it operates on the terminating network and hence can be easily omitted from the emergency call path by the network operator.

The distributed nature of the proposed architecture ensures that it is scalable and flexible, as users are given the option whether or not to use the spam prevention service, without jeopardising the effectiveness of the solution. The solution conforms exactly to the 3GPP IMS specifications, the relevant IETF RFCs and IETF Internet Drafts hence ensuring interoperability and extensibility.

The next chapter details a practical implementation of the proposed architecture in order to show proof-of-concept and provide a platform for evaluations.

Chapter 7

Test-bed Implementation

The previous chapter proposes an architecture for the prevention of spam and the assurance of privacy for IMS users. The architecture is designed such that it is fully standards compliant, extensible and scalable. In order to confirm that the architecture is realisable in a practical network environment the proposed design must be implemented in a network test-bed. A simulation of the proposed architecture would not provide meaningful validation as the research questions whether or not such an architecture can be proved in a practical setting - as opposed to theoretically. Furthermore, simulations tend to oversimplify the problem and require several assumptions that may or may not be correct, whereas in a test-bed environment a new technology can be proved beyond doubt.

This chapter discusses the implementation of a standards compliant IMS test-bed at the University of Cape Town, and the software tools constructed as part of the thesis. The test-bed serves two important purposes: it provides a proof-of-concept prototype of the spam prevention architecture proposed in the thesis, and it allows for several evaluations to be performed in order to measure the effectiveness and performance of the network elements in a practical setting. The proof-of-concept implementation and subsequent evaluations will provide invaluable data for evaluating the feasibility of such an architecture in future IMS networks.

The test-bed places a strong emphasis on being reproducible by others, so wherever possible the network components are comprised of free and open-source software. Open source software licensed under the GPL (GNU Public License) is always provided with the source code or with the provision that

the source code is readily available. This ensures that the test-bed components can be easily and legally altered if any changes are required in the provided software, whereas proprietary software usually does not offer this advantage. Furthermore, many of the tools required for the test-bed do not run on proprietary operating systems. Thus the test-bed utilises the Ubuntu Linux [4] operating system throughout.

7.1 IP Connectivity Access Networks

IMS terminals require a packet-switched CAN (Connectivity Access Network) in order to interface with the IMS core network and services. Four different CANs are implemented in order to test the proposed architecture in scenarios that are typical of practical IMS networks. This section discusses the choice of IP version, the CAN technologies implemented and acquisition of IP addresses. The P-CSCF FQDN (Fully Qualified Domain Name) and port number is manually programmed into the hosts, as specified by the OMA [9], as none of the CANs utilised are able to support the procedures required for automatic P-CSCF discovery.

7.1.1 IPv4 vs IPv6

Due to the limited IPv4 address space there is a critical shortage of public IPv4 addresses in the world, creating a large problem for new IP-based networks such as the IMS. The IETF has proposed a move to IPv6, which uses 128 bit addresses, effectively ending the shortage of IP addresses [20]. Therefore, the IMS was initially developed for use with IPv6 exclusively as it was thought that by the time IMS reached the market IPv6 would be the dominant protocol on the Internet. Unfortunately this prediction was incorrect and as of 2008 take up of IPv6 has been limited to only a few research institutions and government agencies. None of the major ISPs in South Africa, or the cellular network providers, provide IPv6 access at this time. Due to the slow uptake of IPv6 the 3GPP has introduced inter-working elements to allow for the temporary use of IPv4.

NAT (Network Address Translation) has been used extensively in the Internet in order to extend the life of IPv4. A NAT server translates the private IP addresses of the many hosts it serves into a single public IP address and does

the opposite in the reverse direction. Unfortunately, SIP, RTP and RTCP, protocols that are used extensively in the IMS, struggle to traverse NAT servers as these servers assume that all connections are initiated from within the network. Of course, in telecommunication applications, signalling and media can be initiated from both inside and outside the private network. There are several work-arounds for traversing NATs, mostly with the use of session border controllers or additional network elements.

Thus, due to the continued use of IPv4 and to avoid complicating the testing process, the test-bed hosts use the IPv4 protocol and are allocated public IP addresses.

7.1.2 Local Area Network

The fastest and most reliable form of CAN available to the test-bed is a 100 Mbps Fast Ethernet link from the user equipment to the machines hosting the IMS core network. The typical RTT (Round Trip Time) in the test-bed for an ICMP Echo Request packet is approximately 0.13 ms. Errors are so uncommon that they are considered negligible. The host obtains a dynamic IP address from the UCT DHCP (Dynamic Host Configuration Protocol) server, and uses the UCT DNS servers for domain name resolution.

7.1.3 IEEE 802.11

The test-bed terminals implement the IEEE 802.11 wireless LAN technology. The terminals utilise the 802.11g standard that offers a maximum theoretical throughput of 54 Mbps but a range of only a few hundred metres. The typical RTT is approximately 1.24 ms in the test-bed. As with the LAN, the terminal also receives a dynamic public IP from the UCT DHCP server and uses the UCT DNS servers.

7.1.4 EDGE

The test-bed terminals are equipped with Huawei E620 Mobile Connect PCMCIA cards. The card gives the terminal access to wireless cellular networks, in this case the South African Vodacom network. The cards supports EDGE (Enhanced Data Rates for GSM Evolution), a faster version of the vanilla GPRS access technology. When connected to the EDGE network the terminals have

a theoretical maximum down-link throughput of 236 kbps but in reality this tends to be significantly slower depending on the signal strength and current network usage. The range of the radio link is in the order of several kilometres from the cell towers and at present Vodacom coverage extends throughout the majority of South Africa. Through an agreement with Vodacom the card is allocated a public IP address, although this practice is not standard for security reasons.

7.1.5 HSDPA

The Huawei E620 cards also support the 3G HSDPA (High-Speed Down-link Packet Access) access technology via the Vodacom network. The particular version of HSDPA offers a theoretical maximum down-link transfer speed of 1.8 Mbps, significantly faster than the EDGE connection. The card connects to the network and obtains a public IP address in the same fashion as when connecting to the EDGE network.

The hosts utilise the card by means of the Mobile Connect Card Driver for Linux released by the Vodafone R&D labs in Spain [44]. The software is written in the Python language and is licensed under the GPL.

7.2 IMS Core Network, HSS and Registration

The test-bed requires a core network infrastructure for authentication, authorisation and SIP routing. The 3GPP specifications state that the core network is comprised of three CSCFs, for routing and registration, and an HSS (Home Subscriber Server), for storage of user data and service profiles.

7.2.1 Call Session Control Functions

The IMS core network is implemented using the Fraunhofer FOKUS OSIMS (Open Source IMS Core) [25]. The OSIMS project provides an open source implementation of the three IMS CSCFs: Proxy, Interrogating and Serving. The project was first released as open source in November 2006 under the GPLv2 (GNU Public Licence v2) and since then has seen several improvements in functionality and reliability. The OSIMS CSCFs were chosen for the test-bed as they are open source, provide an excellent standards compliant solution, and

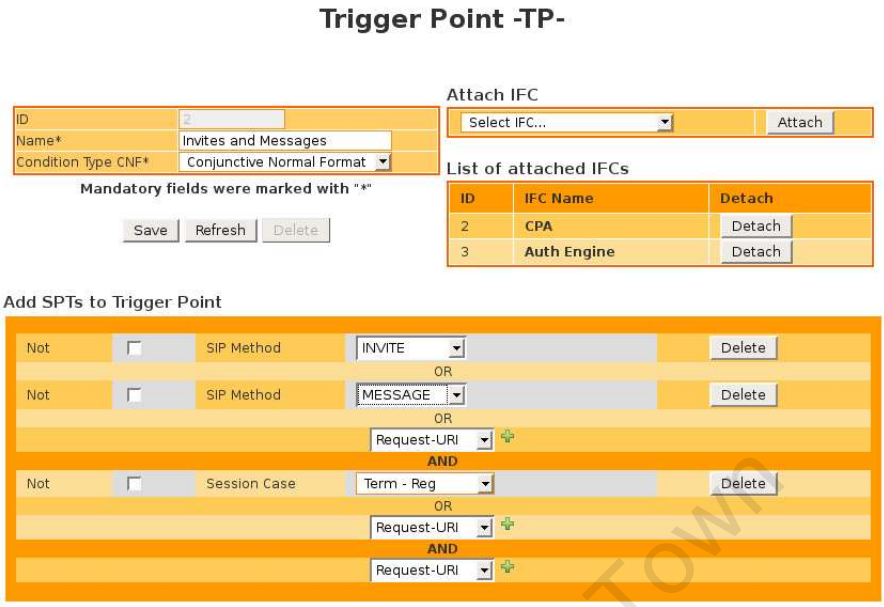

Figure 7.1: A trigger point is configured using the FHoSS web interface.

can support a huge volume of SIP signalling as the project extends the highly efficient SER (SIP Express Router). SER and consequently the OSIMS CSCFs are written in the C programming language that is well-known for excellent speed and performance in delay sensitive applications.

7.2.2 Home Subscriber Server

The HSS is implemented using the FHoSS (FOKUS Home Subscriber Server) that also forms part of the OSIMS project. Unlike the CSCFs the FHoSS is written in the JAVA programming language and uses a MySQL database in order to store data. The FHoSS provides a web interface that allows network operators to easily add and remove users and provision application servers. It is also possible to add initial filter criteria to sessions, which is a requirement for the proposed spam prevention architecture. Figure 7.1 demonstrates a trigger point that is configured to include the mandatory CPA (Call Pattern Analyser) and the optional Authorisation Engine for all terminating INVITE and MESSAGE requests.

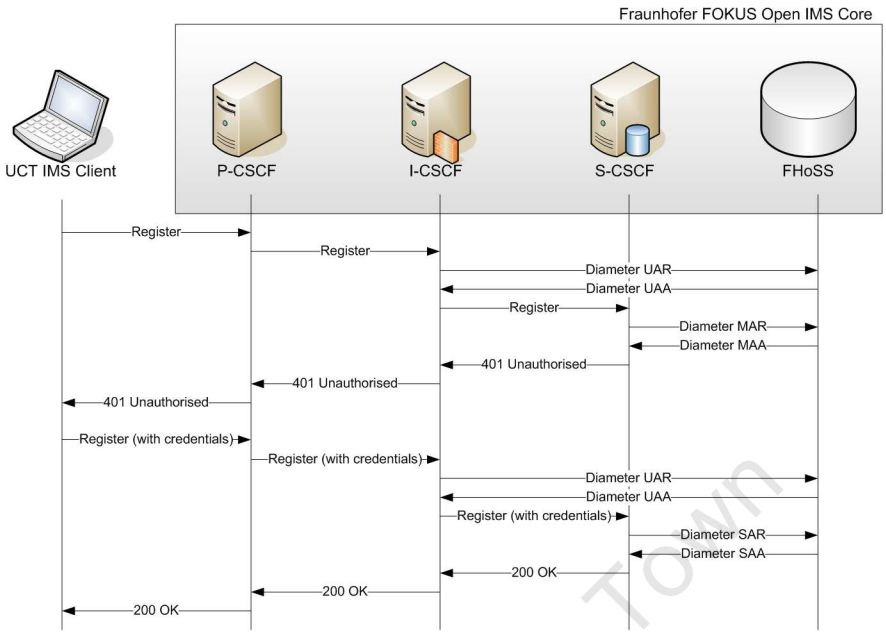

Figure 7.2: The UCT IMS Client registers with the FOKUS Open IMS Core.

7.2.3 IMS Registration

The first part of the implementation involves registering a user with the IMS core, as illustrated in Figure 7.2. The registration serves the purpose of binding the users SIP address to their contact IP address, authenticating the client, authenticating the network and downloading the user's service profile from the HSS to the S-CSCF. The UE sends a SIP REGISTER request to the P-CSCF that has been manually configured in the client software. The request contains the IMPU (Public User Identity) that the user wishes to register in the *To* header and the user's IMPI (Private User Identity) in the *Authorization* header. It is possible for a user to have several IMPUs and several IMPIs associated with their IMS subscription, however, for the purposes of the test-bed the users are allocated a single IMPU and IMPI per subscription.

The REGISTER request also contains the home domain name of the network on which to register. The P-CSCF resolves this domain name and forwards the request to an appropriate I-CSCF. The I-CSCF contacts the HSS with a UAR (User Authentication Request) in order to select an appropriate S-CSCF. The HSS uses the secure Diameter protocol exclusively for communication with the CSCFs and application servers. If an S-CSCF has not been previously assigned to the user, the HSS responds with a UAA (User Authentication Answer) including a set of S-CSCF capabilities. The request is then forwarded

```
<?xml version="1.0" encoding="UTF-8"?>
<IMSSubscription>
  <PrivateID>david@open-ims.test</PrivateID>
  <ServiceProfile>
    <PublicIdentity>
      <Identity>sip:david@open-ims.test</Identity>
    </PublicIdentity>
    <InitialFilterCriteria>
      <Priority>0</Priority>
      <TriggerPoint>
        <SPT>
          <Method>INVITE</Method>
        </SPT>
        <SPT>
          <Method>MESSAGE</Method>
        </SPT>
      </TriggerPoint>
      <ApplicationServer>
        <ServerName>sip:137.158.112.24:7080</ServerName>
      </ApplicationServer>
    </InitialFilterCriteria>
  </ServiceProfile>
</IMSSubscription>
```

Figure 7.3: User profile containing the initial filter criteria for the Call Pattern Analyser.

to the selected S-CSCF where-after the S-CSCF will see that the user has not yet been authorised and will retrieve the user's authentication data from the HSS using a MAR (Multimedia Auth Request) message. This message also instructs the HSS that future requests for the user should be handled by the current S-CSCF. The HSS stores the S-CSCF address in the user data and replies with a MAA (Multimedia Auth Answer) message containing the user's authentication data.

The S-CSCF challenges the UE with a 401 Unauthorised response to which the UE must generate a reply. The test-bed uses the AKAv2 protocol [84] exclusively as it is a supported authentication protocol of the 3GPP IMS. The UE now sends a SIP REGISTER request with the authentication information as requested by the S-CSCF. In a similar fashion to the first request, the message is routed via the I-CSCF to the S-CSCF, which then checks the authentication information is correct. If the UE has provided the correct credentials the S-CSCF send an SAR (Server Assignment Request) in order to inform the HSS that the user is now registered and request a copy of the user profile. The HSS responds with the SAA (Server Assignment Answer) that contains the user profile encoded in XML format. Figure 7.3 shows an example user profile for the IMPU *sip:david@open-ims.test* (some fields are omitted for simplicity).

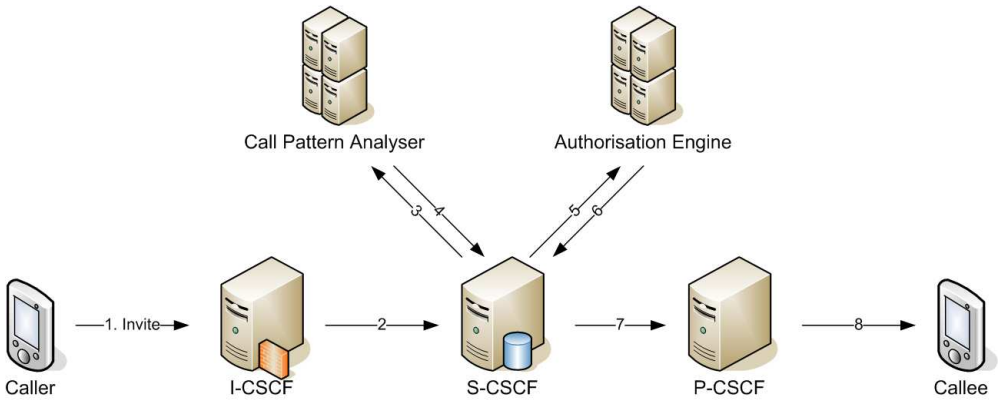

Figure 7.4: The Call Pattern Analyser and the Authorisation Engine are traversed in a serial fashion.

7.2.4 Application Server Invocation

For the purposes of testing the proposed spam prevention architecture it is required that incoming sessions traverse both the CPA and Authorisation Engine. This is achieved using the filter criteria, as defined by the 3GPP [56]. Two types of filter criteria exist: initial filter criteria, that are evaluated for requests that start a new dialog, and subsequent filter criteria, that are evaluated on receiving any other request which already forms part of an existing dialog. Due to an oversight in the specifications, the implementation of subsequent filter criteria would violate the SIP routing rules for proxies, hence the test-bed only implements initial filter criteria.

The two application servers of the proposed spam prevention architecture must be traversed in a serial fashion, as shown in Figure 7.4. Every IMPU in an IMS network is associated with a service profile that contains zero or more initial filter criteria that determine which services are invoked for the different session types. In the case of the spam prevention architecture the user requires two initial filter criteria. The first specifies that all INVITE and MESSAGE requests must traverse the CPA. The second initial filter criterion states that these same requests must traverse the Authorisation Engine.

The order in which these application servers are traversed is vitally important, as the Authorisation Engine relies on information inserted into the SIP message by the CPA. For this reason the initial filter criteria are given integer priority values, zero being the highest priority. Thus it is essential that the CPA initial filter criterion priority value is lower than that of the Authorisation Engine.

The initial filter criteria are assigned trigger points that match particular SIP messages. In the case the test-bed the trigger points match SIP INVITE and MESSAGE requests on the terminating leg of the session. The initial filter criteria also specify the URI of the application server to which the request should be forwarded.

The S-CSCF adds the address of the next application server to be traversed to a *Route* header, hence ensuring the request with be transmitted to the server using regular SIP loose routing rules. The S-CSCF also adds its own address to a second *Route* header so that, once the application server has been traversed, the request will be routed back to itself. State information is added to this second *Route* header in order to inform the S-CSCF which filter criteria have already been processed, thus avoiding a routing loop.

7.3 User Equipment

The UE (User Equipment) is responsible for several tasks including registration with the network, session initiation, session termination, media handling and user interaction. Many SIP applications have been developed since the introduction of the protocol, however, as the IMS is a relatively young technology there are currently few IMS clients available. IMS introduces several new SIP headers and requires a great deal more signalling than vanilla SIP, thus a SIP client cannot be used in place of an IMS client. Figure 7.5 illustrates the very different session setup procedures followed by vanilla SIP and IMS. Fortunately two clients, SIPp and the UCT IMS Client, fulfil the needs of the test-bed in that they offer true IMS signalling.

7.3.1 SIPp

SIPp [3] is an open source software tool that generates SIP traffic. The software is primarily a test tool for measuring call rates, round trip time and message statistics, hence it has only a rudimentary text-based user interface and very limited support for the media plane. SIPp is able to read a session scenario from an XML file and generate appropriate SIP requests and responses. Recently support has been added to enable the software to interact with the FOKUS Open IMS Core using AKAv1 and AKAv2 authentication.

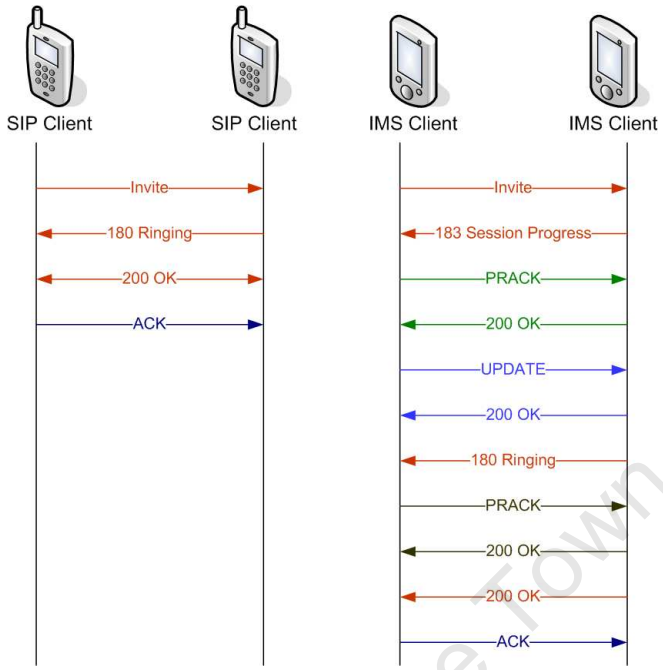

Figure 7.5: Session setup between two SIP clients (left) and two IMS clients (right).

7.3.2 UCT IMS Client

At the time that the Open IMS Core project was released in November 2006 there were no open source IMS clients available. The UCT IMS Client [89] was co-developed by the author of this thesis and a fellow researcher in order to fill this gap and to provide researchers an opportunity to experiment with true IMS signalling. Since the first release of the client in December 2006 it has seen several improvements in functionality, has recorded in excess of 3000 downloads worldwide, and has been used in several open source IMS projects [90].

At present the client supports full IMS session setup, instant messaging, audio, video and XCAP support. The client is able to register with the core network using MD5, AKAv1 or AKAv2 authentication protocols. UDP is used exclusively for SIP transport and TCP for all XCAP requests. The software is free, open source and released publicly under the GPLv3.

The UCT IMS Client is based on the oSIP and eXosip libraries [48] for SIP signalling that are both released under the LGPL (Lesser GNU Public Licence). The libraries allow developers to write code that is fully compliant to RFC 3261. The media plane is supported by the Gstreamer library [1] that

incorporates many different media codecs and supports network streaming via RTP (Real-time Protocol). Gstreamer is also released under the LGPL. The support for XCAP is provided by the Libcurl library. Libcurl is released under a MIT/X derivative license that, similar to the GPL, gives modification and redistribution rights to software developers.

The actual creation of the authorisation documents is beyond the scope of this work, however, this is expected to take place from a dedicated piece of software or possibly a web interface. The client is responsible for uploading these documents over the *Ut* interface to the Rule Database.

7.4 Call Pattern Analyser

The design of the CPA (Call Pattern Analyser) in the previous chapter calls for three modules to collectively assign a spam probability score to a particular session. The three modules are the session duration analyser, session volume analyser and concurrent session analyser. Each module is responsible for assigning an integer spam probability score to the call history in the range [0, 100]. A weighted average of all the module scores is then calculated and appended to the SIP message as it traverses the proxy.

The CPA is implemented in the C programming language using the oSIP library. The server is located in the user's home network and is invoked only on terminating calls.

7.4.1 SIP Proxy Behaviour

The CPA is an application server that acts as a SIP proxy in that it does not initiate or terminate sessions but rather amends SIP requests and proxies them back to the S-CSCF. Therefore the first requirement of the CPA is to follow the general rules for SIP proxies by adding itself to the list of *Via* headers. This ensures that all responses to the transaction will be able to traverse the same proxies that the initial request traversed, as specified by the SIP specifications. The spam probability score is also appended to this *Via* header as described in the design of the proposed architecture.

The CPA also adds itself to the *Record-Route* header containing the *lr* (loose routing) parameter, thus ensuring that all future transactions that form part of the same dialog will also traverse the proxy. This is necessary as

the session duration analysis module must measure the length of the session between the 200 OK message in response to the initial INVITE and the BYE request sent by either party. The 200 OK final response will be routed via the CPA regardless because it forms part of the same SIP transaction as the initial INVITE request, however, the final BYE request is a new SIP transaction (although still part of the same SIP dialog). Hence the need for the *Record-Route* header. Unfortunately this adds additional overhead to the session setup procedure as it means that all transactions in the dialog must traverse the proxy and, due to the verboseness of IMS session setup, this can add up to a significant amount of unnecessary signalling.

The third action when receiving a request is for the CPA to remove its own address from the top *Route* header, thus preventing the request from being re-routed back to itself. As described above, it is not necessary to add the address of the S-CSCF for the next hop as the S-CSCF does this before forwarding any requests to the application server.

On receiving a provisional (other than 100 trying) or 2xx final response for a request the proxy removes itself from the *Via* header and sends the request back to the S-CSCF. All other responses are handled on a hop-by-hop basis by the proxy. As the CPA implementation is for research purposes only it does not fully conform to the SIP specifications for proxy behaviour, however, it handles all call scenarios used in this project.

7.4.2 Data Collection

The architecture design dictates that the CPA store the session history of each unique caller. This information is stored in a MySQL database table. MySQL is an efficient open source database solution that offers quick look-ups even in very large tables. The caller's URI is hashed with the MD5 algorithm in order to alleviate any privacy concerns. Each session record contains the session type, dialog identifier (DID), time of initial INVITE, call start time and call termination time. In order to simplify the call duration calculations the time values are stored in seconds from the Unix epoch (1 January 1970). In a separate table each caller is also assigned a record that contains the current number of active calls. An examples of the session records table is shown in Table 7.1; some of the fields are shortened due to space constraints.

When a new SIP INVITE arrives the CPA checks to see whether the dialog

Table 7.1: An example of the session record table stored by the CPA.

Type	From	DID	Invite time	Call start time	End time
application/sdp	9452ec52	1	1203940996	1203940999	1203941020
text/plain	3c2caff6	2	1203940834	1203940838	1203940859
text/mime	b53f7a85	3	1203941205	1203941208	1203941267

identifier already exists in the table. If not, a new session record is created and the INVITE time is recorded. In this way retransmitted INVITE requests are ignored so as to avoid incorrectly increasing the spam score for users with slow or lossy links.

On receiving a 200 OK message, the CPA looks for the corresponding INVITE message and marks the start time of the call. The user's active call count is increased by one. Similarly, when the CPA receives a BYE message from either the UAC (User Agent Client) or UAS (User Agent Server) the call record is updated with the termination time of the call and the active call count is decremented by one. The call duration is not specifically recorded as this is easily calculated by subtracting the start time from the termination time.

7.4.3 Session Duration Analysis Module

The session duration analysis module monitors suspicious calling behaviour based on the previous session durations of the current caller. The task of this module is to quickly and efficiently analyse the user's call history and assign a spam probability score that is proportional to the similarity between the caller's history and known spamming behaviour. In Chapter 5 an experiment was conducted showing that the calls from an automated caller can be effectively distinguished from a legitimate caller by means of a PDF (Probability Density Function). In this implementation the CPA attempts to fit a known PDF to the measured call duration history. A thorough overview of function minimisation algorithms is outside the scope of this thesis. However, for completeness a brief summary of the call duration analysis module implementation is discussed.

Through experimentation and observation typical spammer PDFs can be obtained. The CPA must determine if any known spammer PDFs are an accurate model for the current callers history histogram. If so, it can be concluded that the caller's history is suspicious and a high spam probability

score should be apportioned. This thesis proposes the bimodal PDF presented in Chapter 5 as an appropriate departure point for this study.

A least squares fitting procedure is used to minimise the PDF function whereby a function minimiser takes in an initial guess of the function parameters and adjusts them accordingly. This minimisation is non-trivial, particularly as the proposed function has five parameters that must be adjusted in order to obtain the best fit. The parameters that must not be adjusted are the bin number and the bin occupation. The minimisation is performed by a C++ library co-written by the author and a doctoral researcher at Forschungszentrum Jülich, and is adapted from the code by Press *et al.* [54].

The minimiser consists of two methods. One minimises according to a single parameter and the other performs multi-parameter minimisation. Single parameter minimisation is achieved using the Golden section search proposed by Kiefer [42]. First, a region must be found in which the function minimum is contained. This is done by starting at the value of the single parameter and finding the direction in which the function decreases. The algorithm then moves in that direction, doubling the amount of the move at every iteration, until the function increases again. Once a region has been found it is progressively made smaller until the minimum is located.

The proposed bimodal function requires multi-parameter minimisation. This can be achieved by choosing a direction in the multi-parameter space according to the gradient of the function and moving in the direction in order to find a minimum. However, this method produces a naïve, and consequently poor, guess of direction [54]. Thus the conjugate gradient method is utilised that gives a better guess of direction and generally leads to faster convergence [34]. In order to minimise the function in the chosen direction a one-dimensional minimiser is utilised. The function takes in the point and the direction of change, and only the amount of change is allowed to vary, hence becoming a function of only a single parameter.

Having determined the function parameters of the comparative curve the observed session records can be correlated against it. Once the outlier data has been removed, as described in Chapter 5, the next step of the correlation procedure is to construct a histogram of ten equally spaced bins from a session length of zero to the maximum session length observed from the user. The histogram is normalised by dividing each bin occupation by the total number of calls in the sample size. In order to get a smooth curve the histogram should

use as much data as possible. A session history of fewer than 50 sessions of a particular type will result in a poor histogram, therefore this technique cannot be used until a reasonable session history has been established for the user. Naturally there is a preference to use as many session history records as are available, however using too many sessions records will reduce the sensitivity of the module to changing calling behaviour. Therefore only the last 100 sessions records are used in this implementation.

The CPA must find the Pearson product-moment correlation coefficient between the predicted data of the spammer curve and the observed data of the caller's history. The coefficient r, shown in Equation 7.1, is a measure of the correlation between the variables X and Y, where z is the standard score and n the sample size.

$$r = \frac{\sum z_x z_y}{n-1} \tag{7.1}$$

The value of r determines the degree of linear relationship between the two variables, and falls within the range [-1,1], where -1 signifies a perfect negative linear relationship, 0 no linear relationship and 1 a perfect positive linear relationship. It is unlikely, if not impossible, that any of these values are achievable in a practical setting but higher values do indicate greater correlation than lower values thus still allowing for meaningful conclusions to be drawn. The CPA is only interested in positive linear relationships thus negative r scores are simply treated as zero.

It is a complicated problem to determine how the r score be related to the final spam probability score. As the range of r scores effectively encompasses [0,1] it is tempting to simply multiple this value by 100 in order to calculate the spam probability score. Unfortunately, this would not be correct as only an r score of greater than 0.5 can really be considered meaningful, and even then this is a low correlation coefficient. Therefore, as an initial guess the SPS (Spam Probability Score) for the module is calculated as shown in Equation 7.2. Evaluations in the forthcoming chapter will attempt to provide more insight as to the validity of this calculation.

$$SPS_{Duration} = \begin{cases} 200(r-0.5) & r \geq 0.5 \\ 0 & r < 0.5 \end{cases} \tag{7.2}$$

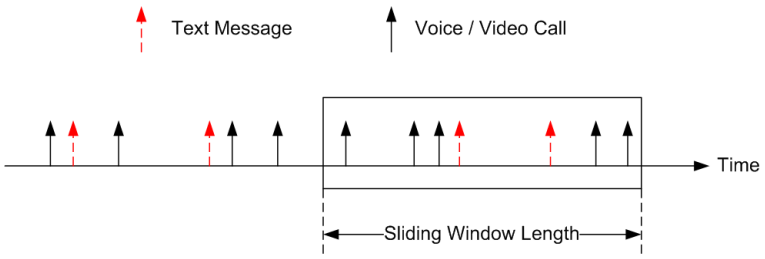

Figure 7.6: Users are allowed to initiate a maximum number of sessions during a sliding time window.

7.4.4 Session Volume Analysis Module

The call volume analysis module is significantly more simple to implement than the call duration analysis module. Shin *et al. [75]* proposes the concept of using a short-term and a long-term analysis of the call history in order to make an accurate assessment of the user's call volume. However, for the sake of simplicity it was decided to take a far more basic approach to the implementation of a session volume analysis module.

The module distinguishes between different media types because users might realistically be expected to initiate far fewer voice and video calls, than text or multi-part instant message sessions. The media type is extracted from the SIP message so that calls are treated separately to text messages. This is because a spammer is unlikely to launch an attack using several different types of media simultaneously.

The architecture assumes the use of session-based text message transfer. Session-based messages split the signalling from the media, similar to multimedia calls, except that instead of RTP in the media plane, the MSRP (Message Session Relay Protocol) is utilised. Therefore, an instant message dialog may be made up of many messages between the two participants. If a new dialog is started for every message, which can be the case for pager-mode messages transmitted over the SIP MESSAGE request, then the module would count each message in the conversation as a new session. Clearly this is not ideal as the user would quickly be identified as a spammer. A work-around can be introduced so that the module identifies messages between two participants as belonging to the same conversation but this feature has not been implemented at this time.

In this implementation the module is configured with two parameters for each media type: a sliding window time period, shown in Figure 7.6, and the

maximum number of sessions that are allowed during this window. The module subtracts the time window from current time and calculates how many sessions of that type where initiated during this time. Equation 7.3 shows how the spam probability score for the call volume analysis module is calculated. The numerator is reduced by one so that a user who makes exactly the allowable volume of calls is not given a spam score. The division is a modulus division so any non-integer remainder is discarded.

$$SPS_{Call-Volume} = 10 \left\lfloor \frac{InitiatedSessions - 1}{AllowedSessions} \right\rfloor \tag{7.3}$$

7.4.5 Concurrent Session Analysis Module

The implementation of the concurrent session analysis module is not difficult because the CPA stores the number of active calls for each user. The module has only one adjustable parameter - the allowable number of concurrent sessions. The spam probability score calculation shown in Equation 7.4 is similar to that for the session volume analysis module.

$$SPS_{Concurrent} = 10 \left\lfloor \frac{ConcurrentSessions - 1}{AllowedSessions} \right\rfloor \tag{7.4}$$

7.4.6 Overall Spam Probability Score Calculation

Once each of the modules present in the CPA have delivered their respective spam probability scores an overall score can be calculated. This is simply a weighted average of the modules scores also in the range [0,100], as described by Equation 7.5, where each w_n lies in the range [0,1] and $\sum w_n = 1$.

$$SPS = w_1.SPS_{Duration} + w_2.SPS_{Volume} + w_3.SPS_{Concurrent} \tag{7.5}$$

The values of w are mostly dependent on the type of session. For example an audio call might place more emphasis on the session duration and concurrent sessions modules, whereas an instant message might place all the weight on the session volume module. In this implementation each module is given an equal weighting of 0.33.

7.5 Rule Database XDMS

The Rule Database is a server that allows clients to store, modify and retrieve stored XML rule documents. A server of this type is known as an XDMS (XML Document Management Server) in IMS terminology. The rule documents are stored in an MySQL table that enables quick an efficient look-ups. The documents are transported to and from the server using the XCAP protocol. XCAP maps XML documents, sub-trees and elements to HTTP URIs thus allowing these elements to be directly accessed using HTTP. An XCAP server supports the HTTP commands GET, PUT and DELETE for downloading, uploading and deleting documents respectively [64].

7.5.1 XCAP URIs

The XCAP URI is comprised of the XCAP root and a document selector. The XCAP root is a valid HTTP URI that can be resolved by the client software, for example a typical XCAP root for a server running on port 8000 might be *http://rd.open-ims.test:8000/xcap-root/*. The document selector is a concatenation of the AUID (Application Unique ID) and the users sub-tree. This implementation utilises the spit policy document format specified by Tschofenig *et al.* [86], which specifies the AUID as *spit-rules*. The users sub-tree informs the XDMS which user's authorisation rules are being requested, or the word global if the default rules are required. Optionally the XML file name where the document is stored can be appended to the URI. An example of a full request URI is shown below.

http://example.com/spit-rules/users/sip:david@open-ims.test/spit-rules.xml

7.5.2 XCAP Signalling

The server is implemented using the OpenXCAP project [2] that is written using the Python language under the BSD licence. In order to secure connections to the server OpenXCAP supports TLS encryption and basic HTTP authentication. The server is able to validate the XML documents that are uploaded via the PUT command using stored XML schema files. Figure 7.7 shows a typical signalling flow between the client, Rule Database and the Authorisation Engine. The authentication mechanisms for HTTP are very similar to those of SIP due to the fact that SIP is based on HTTP. However, unlike the

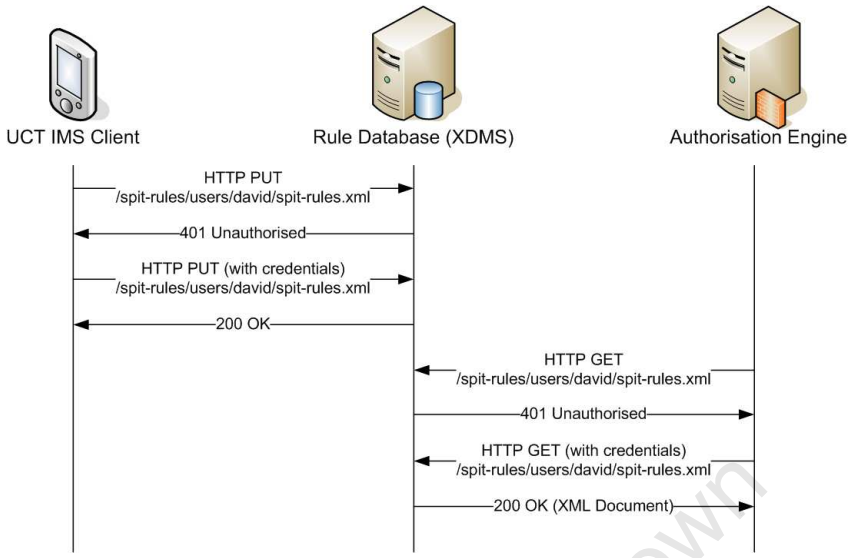

Figure 7.7: Storing and retrieving documents from the Rule Database (XDMS).

IMS authentication detailed earlier in the chapter, the XCAP packets do not flow through the IMS core elements but rather over the *Ut* interface directly between the client and the XDMS.

7.5.3 Default Behaviour

The user is required to upload their authorisation preferences to the rule database before any incoming calls are received. If the user has not uploaded any rule document, or the rule document has expired for some reason, then the rule database issues an error. The requester then has the option to download the default rules as defined by the network operator. In the case of this implementation the default rule document contains a rule to allow all sessions. An example of the XCAP URI for this request is shown below.

http://example.com/spit-rules/global/spit-rules.xml

7.6 Authorisation Engine

The Authorisation Engine is tasked with making authorisation decisions for incoming session requests on behalf of the user. These decisions are based on the authorisation document obtained from the Rule Database and the spam probability score appended to the *Via* header of the SIP request. In the

proposed architecture there are several possible outcomes. First, the call may be blocked outright with no chance of recourse. Second, the call may be allowed to traverse through to the recipient without further obstruction. The final option is that the call may be provisionally accepted depending on the outcome of either a computational puzzle or a Turing test.

7.6.1 Server Tasks

The first task of the Authorisation Engine is to extract the identity of the called party from the *To* header of the SIP request, and to generate a suitable XCAP request to the Rule Database XDMS. The authorisation document for the specified user is downloaded from the XDMS and stored it in temporary memory. The Spam Probability Score is then extracted from the CPA's *Via* header in the SIP request, following which the Authorisation Engine implements the rules defined in the authorisation document.

7.6.2 Proxy Behaviour vs. UAS Behaviour

An application server can act in several different modes. It can behave as a SIP proxy, a UAC (User Agent Client) or as a UAS (User Agent Server). In proxy-mode the server simply proxies the request back to the S-CSCF following the rules described for the CPA. As a UAC the server can initiate calls and as a UAS the server can terminate calls.

The proposed architecture calls for the Authorisation Engine to act in either proxy or UAS mode depending on the outcome of the authorisation decision. If, for example, a caller appears on the target user's whitelist then the call must be allowed to continue uninhibited. In this case the server will simply act as a proxy, similar to the CPA, with the exception that the Authorisation Engine does not need to add a *Record-Route* header as it is not interested in any future transactions in the SIP dialog. However, if the authorisation decision requires that either a computational puzzle or Turing test is required then the server is required to terminate the call itself, and thus act as a UAS.

As the SIP specifications call for different behaviour from SIP clients and proxies this presents an interesting implementation problem. The chosen solution is to always proxy the request, but if the server must terminate the call then the request is simply proxied back to the Authorisation Engine either to a different port or a different server altogether. This behaviour is demonstrated

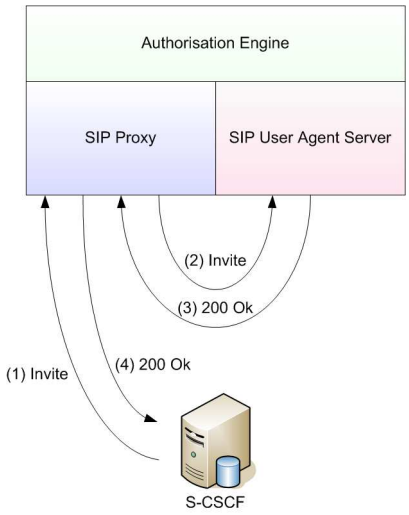

Figure 7.8: The Authorisation Engine proxies a request back to itself so that it can be answered by the user agent server.

in Figure 7.8. Thus, while the Authorisation Engine is a single logical entity, it is actually implemented as several differently behaving servers. This intermediate proxying is often not shown explicitly in future signalling diagrams.

The SIP proxy is implemented using the oSIP library, whereas the SIP UAS is implemented using the eXosip library which extends oSIP to provide UA capabilities.

7.6.3 Decision Engine

The Authorisation Engine must make decisions based on the user's authorisation document. The document contains an unordered list of rules which are made up of conditions, actions and transformations. Conditions and transformations inform the authorisation engine to enforce the respective action. This implementation does not fully conform to the spit policy and common policy specifications [86, 72], as it only provides support for a subset of the conditions and actions elements.

A rule is executed when the conditions of that rule have been met. Typical conditions include the identity of the sender and the validity of the rule. The identity of the sender can be specified in several different ways but the most common is to either specify a specific sender's URI, using the *one* attribute, or to specify an entire domain, using the *many* attribute. The validity of the rule specifies during which time period the rule should take effect and must

contain two attributes, *from* and *until*, that define the start and end of the rule validity respectively.

The spit policy framework introduces a *spit-handling* element as a possible condition part of a rule. This element is useful for cases that rely on the outcome of some other mechanism, for example when the caller is subjected to a Turing test or computational puzzle. The *spit-handling* element is activated when the reply to the test has been received. The element specifies which rules should be enforced depending on whether or not the caller passed the test. The spit policy framework also extends the common policy validity element to allow more detailed time periods for rules. For example, the spit policy framework allows a rule to be enacted only on specific days of the week and during designated times. Therefore, different rules can be specified for periods when the recipient is likely at work, on vacation or sleeping. This fine-grained control allows users the ability to construct highly personalised rules for when and by whom they can be contacted.

The design of the proposed architecture requires that rules should also take into account the Spam Probability Score. Unfortunately, there is no condition defined in either the common policy or spit policy frameworks that handle this condition. Therefore, a new namespace is introduced containing a conditional element named *spam-score*. This element has two attributes, *from* and *until*, and must contain at least one of these attributes. If only the *from* attribute is used then any score equal to or higher than the specified score matches the rule. Similarly, if only the *until* attribute is used then any score equal to or less than the specified rule will match. Figure 7.9 shows an example of the new element in an authorisation document. The document species in *rule1* that if a session arrives with a spam score between 10 and 90 then a computational puzzle should be performed. A second rule specifies that if the computational puzzle is solved correctly then the session should be allowed to continue.

The actions part of the rule specifies what should happen when one the conditions have been met. Specific actions are defined by the application-specific usages of the common policy framework. In this case these actions are defined by the spit policy framework, which defines three actions - *allow*, *block* and *forward-to*. The spit policy framework also allows other actions such as *hashcash* [39] or *captcha* [85] - the implementation of these actions is discussed in further detail below. There is an inconsistency in the spit-policy Internet draft as to whether the command for actions should be *execute* or

```
<?xml version="1.0" encoding="UTF-8"?>
<ruleset xmlns="urn:ietf:params:xml:ns:common-policy" xmlns:spit="urn:ietf:params:xml:ns:spit-policy"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:sc=”http://example.com/spam-score”>
<rule id="rule1">
  <conditions>
    <identity>
      <many/>
    </identity>
    <sc:spam-score>
      <sc:from>10</sc:from>
      <sc:until>90</sc:until>
    </sc:spam-score>
  </conditions>
  <actions>
    <spit:execute>hashcash</spit:execute>
  </actions>
<transformations/>
</rule>
<rule id="rule2">
  <conditions>
    <spit:spit-handling>
      <challenge result=”SUCCESS”>hashcash</challenge>
    </spit:spit-handling>
  </conditions>
  <actions>
    <spit:execute>allow</spit:execute>
  </actions>
  <transformations/>
</rule>
</ruleset>
```

Figure 7.9: An authorisation document illustrating the use of the spam-score element.

spit-handling. In this implementation the *execute* command is used as this is the element specified in the supplied XML schema.

The Authorisation Engine must incorporate a mechanism by which rule conflicts can be resolved. This is achieved by requiring each rule to assign a permission setting to the particular session. It is possible that a particular session matches the permissions of two separate rules. In this case the common policy framework specifies that the highest permission be allocated to the session. For example, if a caller appears on the recipient's whitelist and blacklist, then the highest permission will be applied and the session will be allowed.

7.6.4 Computational Puzzle

The proposed spam prevention architecture requires that there be a method by which to challenge the caller with a computational puzzle. Such puzzles are also commonly referred to as a hashcash puzzle as they are often based on hashing algorithms. The implementation of the computational puzzle is based primarily on the Jennings' Internet draft [39]. The draft discusses a method by which a server can calculate a computational puzzle of variable difficulty using a SHA1 hash [21].

The server constructs a string by concatenating a random number, the current time and various header values from the SIP request. The string is then hashed with SHA1 to create what is known as the pre-image. The pre-image is concatenated with the string *z9hG4bK* and then hashed again with SHA1 to create the image. The SHA1 hash is used since it is very quick to create a hash but almost impossible to find the input value of the hash without trying every possible combination of alphanumeric characters. The puzzle difficulty is determined by a parameter known as *work*. The value of work determines how many bits of the pre-image are set to zero. The image, pre-image and work values are then sent to the user agent client that must attempt to recreate the pre-image using a brute-force method. The difficulty of the puzzle increases exponentially as the number of work bits increases.

In order to simplify the implementation a very similar hashcash algorithm by Black [12] is implemented in place of the Jennings' algorithm. The Black algorithm involves discovering a pre-image that when concatenated with a string and the date produces an all-zero string. The difficulty of the puzzle

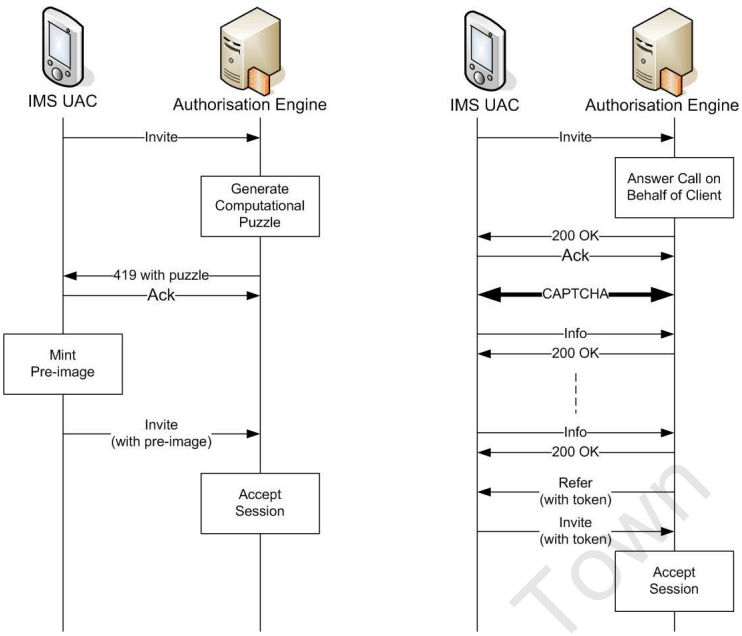

Figure 7.10: Signalling flows for the computational puzzle (left) and the Turing test (right) between the UAC and the Authorisation Engine UAS .

is determined by how many leading zeros are required in the image and thus requires the same amount of work as the Jennings' algorithm. The Jennings' hashcash technique is based on Black's work, hence the similarity, however, the Black algorithm is favoured due to the fact that source code for the algorithm is readily available [14].

If a rule states that a computational puzzle is required the Authorisation Engine proxies the request to the UAS, which then generates a random string. The string and the amount of work required are sent back to the UAC in a *419 with puzzle* SIP response. In the proposed architecture the difficulty of the puzzle is related to the Spam Probability Score extracted from the session request. Thus a session with a high spam probability will be required to solve a harder puzzle than one with a lower score.

The UAC is then responsible for minting a pre-image that satisfies the work requirement of the puzzle and can then re-send the original request. This time the request includes the Puzzle header with the pre-image, as proposed by Jennings [39].

Refer-To: <sip:david@open-ims.test; token="76a9K1nS">

Figure 7.11: An example of the *Refer-To* header field.

7.6.5 Turing Test

The Turing test server is implemented as a SIP UAS using the eXosip library for signalling and the Gstreamer framework for media delivery. The Gstreamer framework encodes the video in the H.263 codec and the audio in the MPEG layer 1 codec and transmits the stream over RTP to the UAC. The codecs were chosen for their high quality so as to reduce the chance of the test being illegible or inaudible for legitimate callers.

If a rule dictates that a Turing test must be completed then the Authorisation Engine UAS answers the session on behalf of the intended recipient. If the session contains a video component then a visual CAPTCHA is streamed to the UAC, or if no video component is defined then an audio CAPTCHA is used instead. Alternatively, a combined audio/video stream can be used. In all cases the caller is required to identify the visual or audio characters and type them into their IMS client. These characters are transferred to the Authorisation Engine using the SIP NOTIFY request, in a similar fashion that DTMF tones are transmitted over SIP signalling.

If the caller correctly solves the CAPTCHA then the UAS sends a SIP REFER message to the UAC. The URI in the REFER method includes in the *Refer-To* header [79] the original SIP URI of the intended recipient plus an additional token made up of pseudo-random characters included as a URI parameter, as shown in Figure 7.11. This token indicates to the Authorisation Engine that the caller has previously passed the Turing test. The supplied token can be used only once and expires after 3 minutes, thus the UAC is required to re-initiate the call immediately or otherwise risks having to pass another Turing test if the token expires.

Figure 7.10 shows the signalling between the UAC and the Authorisation Engine UAS for both the computational puzzle and the Turing test scenarios - non-essential signalling messages are omitted from the figure.

7.7 Summary

This chapter has detailed the implementation of the proposed spam prevention architecture. Thanks to software projects such as the Fraunhofer FOKUS Open Source IMS Core, UCT IMS Client, oSIP libraries, and SIPp traffic generator, the implementation was performed exclusively with open source components. The benefits of open source are numerous, but the most important is that the test-bed implementation is fully reproducible by future researchers in the field. Furthermore, each of the components can be easily modified to incorporate experimental technologies such as the ones proposed in this thesis.

In the following chapter the proposed spam prevention architecture is subjected to several evaluations, many of which would not be possible without a practical network test-bed. Thus while the test-bed has effectively demonstrated proof-of-concept it also serves as a platform for experimentation and validation.

Chapter 8

Evaluations and Results

The previous two chapters have discussed the design and implementation of a proposed IMS spam prevention architecture. In order to confirm the suitability of the proposed architecture it is subjected in this chapter to an array of evaluations regarding the effectiveness of the solution in detecting spam, as well as the associated resource overheads that accompany such security solutions. Previous work by the author [91] has described evaluations performed on this architecture using simulation tools. However, these simulations did not take into account all the variables of a practical network and are overly simplified. As such they do not provide sufficiently accurate results. For this reason it was decided in this thesis to perform the majority of evaluations using the practical IMS test-bed implementation described in the previous chapter. While it is acknowledged that there are always differences in practical networks and outcomes are not always fully repeatable, the results do provide a relevant and accurate departure point for the expected performance of the proposed architecture in a practical IMS network setting.

The chapter begins with evaluations of the traffic and delay overheads experienced when implementing the proposed architecture in a practical setting. It then proceeds to evaluate the session volume, session duration and concurrent session analysis modules of the Call Pattern Analyser under various scenarios. Finally the chapter describes evaluations performed on the computational puzzle mechanism and the Turing test mechanism. For brevity, the chapter does not seek to evaluate the most basic functionality of architecture, including whitelisting, blacklisting and time-based filtering.

The results of the performed evaluations aim to demonstrate the effective-

ness of the proposed solution as well as limitations to guide future studies.

8.1 Overheads

The first evaluations performed are dedicated to determining the overheads introduced by the implementation of the spam prevention architecture. There are two types of overhead that must be considered: the increased traffic load in the network and the additional session setup delays.

The proposed spam prevention solution functions during session setup and therefore has no significant effect on registration delay. However, as the user must specify to use the spam prevention service in their user profile there is some increased traffic overheads during registration. An increase in traffic overhead during registration is not critical to the network performance as registration events occur relatively infrequently.

On the other hand an evaluation of traffic and delay overheads during session setup is highly invaluable as these events occur frequently in typical telecommunication networks. An increase in traffic volumes in the core network could overwhelm the capacity of the network nodes, and an increase in session setup time reduces user utility in the network. The evaluations aim to determine the severity of the overheads introduced by the proposed solution.

8.1.1 Method

Throughout the evaluations in this chapter there are three recurring cases that are examined; these are illustrated in Figure 8.1. The first is the reference case, where the network does not implement any part of the spam prevention solution. The second case, Case II, is where the network implements the solution but the user does not utilise the service. This type of user perhaps does not trust the service or does not wish to pay the extra fee that may be associated with it. Emergency service numbers and open groups also fall into this category. The service provider includes the Call Pattern Analyser (CPA) in these users service profile but not the Authorisation Engine. The third case, Case III, is when the network implements the proposed solution and the user chooses to use it. In this case the user profile includes both the CPA and Authorisation Engine.

The evaluation of traffic overheads requires only a single measurement as

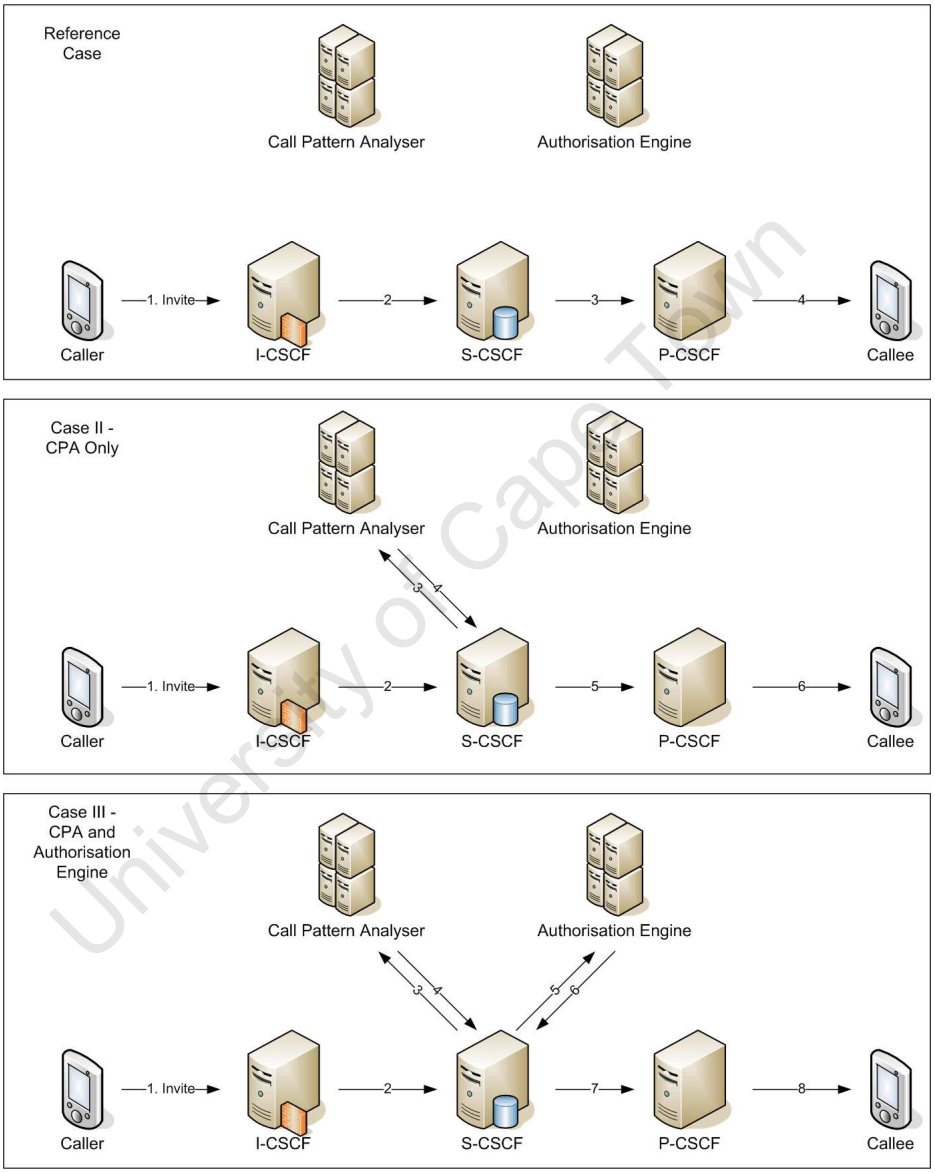

Figure 8.1: The three evaluation cases.

Table 8.1: IMS-level registration traffic overheads.

	Reference Case	Case II	Case III
Total core traffic (bytes)	22326	22926	23522
Traffic overhead (%)	-	2.69%	5.36%

the amount of traffic flowing from one event to another is constant. However, the evaluation of delay overheads requires several measurements as there are many nodes involved in the session setup process and each has variable response times. Therefore fifty samples are averaged for all experiments where delays are measured in order to achieve accurate measurements.

8.1.2 Registration Traffic Overhead

The evaluation is performed using the UCT IMS Client and the FOKUS Open Source IMS Core. The client registers with the core with the AKAv2 authentication scheme. The service profile for the user is modified to comply with each case described above. This service profile is downloaded from the HSS to the S-CSCF during a successful registration.

The open source Wireshark Network Protocol Analyser [5] is used to capture all SIP and Diameter network traffic to and from the P-CSCF, I-CSCF and S-CSCF. Only the actual registration is captured and not the subsequent subscriptions to the user's reg event by the P-CSCF and the UE.

The results of the evaluation are detailed in Table 8.1. The results show that approximately 22 kB of network traffic flows between the IMS core elements during a normal user registration. When only using the CPA this value increase by approximately half a kilobyte equating to 2.69% of the reference case. When using the full solution there is approximately one kilobyte more traffic amounting to a 5.36% increase in network traffic.

8.1.3 Session Setup Traffic Overhead

In order to determine the session setup traffic overheads a successful video call session is initiated between two IMS users in the same IMS realm. The session setup is defined as the signalling from the initial INVITE request to the 200 OK of the provisional response acknowledgement (PRACK) for the 180 Ringing response.

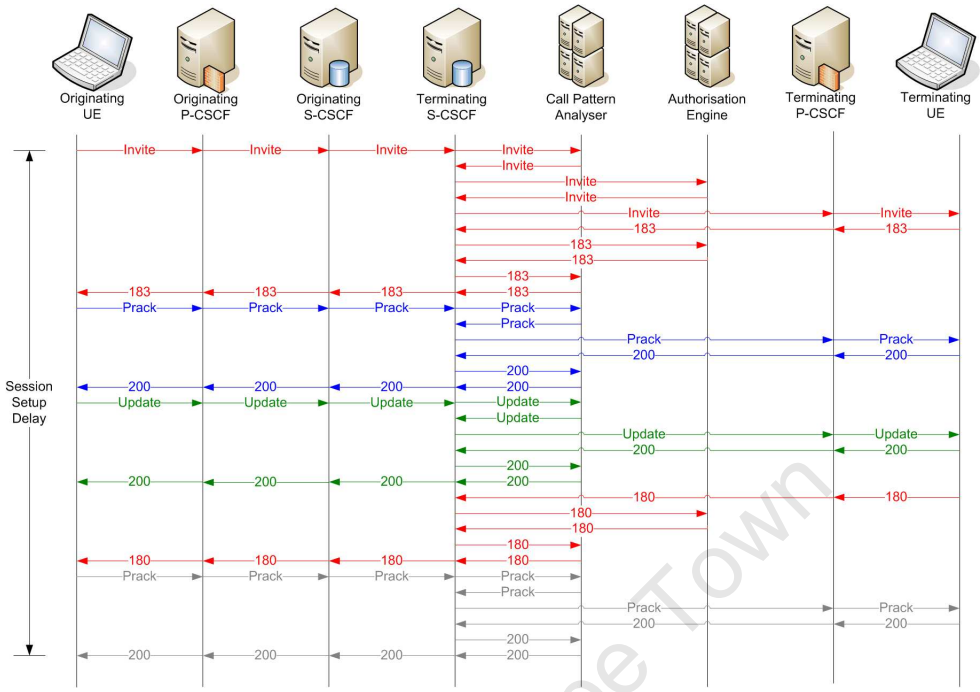

Figure 8.2: Session setup signalling (Case III).

The authorisation policy of the terminating user does not subject the caller to either a computational puzzle or a Turing test. The total traffic to and from the three CSCFs, the CPA and the Authorisation Engine is measured in terms of number of SIP packets and the number of bytes transmitted between these elements. For simplicity the XCAP look-ups are excluded from the evaluation as these do not traverse the IMS core elements. Unlike the registration procedure, there are no Diameter messages exchanged between the IMS core elements during session setup apart from charging and QoS elements that are excluded from this evaluation as they have little bearing on the results. The initial INVITE request follows the service-routes specified during registration. The measurement includes both the originating and terminating parts of the call leg as the two users are part of the same domain.

The three cases illustrated in Figure 8.1 are used again in this evaluation. In the Case II only the CPA is invoked during session setup. Recall that the CPA must record-route itself in order to stay on the signalling path of the dialog and hence all subsequent transactions within the dialog are routed via the CPA. Case III invokes the Authorisation Engine after the CPA. The Authorisation Engine does not record-route itself as it is concerned only with an initial authorisation decision and not subsequent requests in the dialog. For

Table 8.2: Session setup traffic overheads.

	Reference case	Case II	Case III
Total SIP packets	55	77	93
Total core traffic (bytes)	57042	79434	94892
Traffic overhead (%)	-	39.26%	66.35%

Table 8.3: Session setup delay overheads for LAN.

	Reference case	Case I	Case II
Mean Setup Time (seconds)	1.064	1.091	1.129
Mean delay overhead (%)	-	2.54%	6.11%
95th Percentile (seconds)	1.149	1.153	1.304

clarity this behaviour is illustrated in Figure 8.2.

The results of the evaluation are shown in Table 8.2. In order to setup an IMS call 55 SIP messages are sent between the core elements with a total traffic volume of approximately 57 kB. When only the CPA is invoked the number of SIP messages increases to 77 and the traffic increases by 39.26%. In the final case where both the CPA and Authorisation Engine are invoked 93 SIP messages are exchanged with an increase of 66.35% in traffic from the reference case. The marginal increase in signalling is less in Case II as only requests and responses that are part of the INVITE transaction must traverse the Authorisation Engine proxy.

8.1.4 Session Setup Delay Overhead

In the above evaluation it has been shown that the proposed architecture does add some additional traffic overheads to the IMS core network. The traffic overheads only affect the network operator, however the resultant delay overheads adversely influence the experience of the end user. Thus it is important to calculate the increase in call setup time when using the proposed spam prevention architecture.

Table 8.4: Session setup delay overheads for WLAN.

	Reference case	Case I	Case II
Mean Setup Time (seconds)	1.110	1.114	1.151
Mean delay overhead (%)	-	0.36%	3.69%
95th Percentile (seconds)	1.281	1.157	1.365

Table 8.5: Session setup delay overheads for HSDPA.

	Reference case	Case I	Case II
Mean Setup Time (seconds)	3.952	3.957	3.963
Mean delay overhead (%)	-	0.12%	0.28%
95th Percentile (seconds)	5.254	5.285	5.231

In this evaluation the call setup delay time is measured over three different access networks, Ethernet LAN, 802.11 WLAN and HSDPA. These CANs (Connectivity Access Networks) are used to illustrate how the access network technology affects the session setup time and consequently the proportion of overhead that the spam prevention solution adds. While the CANs change in each experiment the network elements themselves are always connected by an Ethernet LAN. The cases shown in Figure 8.1 are again applied and the results compared, establishing the delay overheads introduced.

The experiment is conducted so as emulate a operator network environment as closely as possible. The CPA is supplied a database with approximately one million call records to mimic a typical table size that might be experienced in a practical network. On receiving an INVITE or MESSAGE request the CPA must extract the last 100 session records of that type from the call record table in order to calculate the spam probability score. The three modules then run in serial and the final spam score is calculated. The Authorisation Engine on the other hand must perform an XCAP look-up on every session request. It then authorises the caller and proxies the call back to the S-CSCF.

The user terminals are emulated by an Intel 1.6 GHz Centrino Dual Core laptop running the UCT IMS Client. The core elements run on Intel 3.0 GHz Dual Core desktop machines. The measurements are performed by the built-in session setup delay timer in the UCT IMS Client.

The results for the LAN, WLAN and HSDPA evaluations are shown in Table 8.3, Table 8.4 and Table 8.5 respectively. For the LAN access network the CPA adds 2.54% delay overhead and the full solution adds 6.11%. For WLAN access these figures drop to 0.36% and 3.69%. Using the HSDPA access network the overhead falls to 0.12% for Case I and 0.28% for Case II.

8.1.5 Discussion

This section has evaluated the traffic and delay overheads associated with the implementation of the proposed spam prevention solution. It is found that in the worst case a relatively small traffic overhead of 5.36% is incurred during registration due to the service profile data being downloaded from the HSS to the S-CSCF, and this overhead amounts to only a few kilobytes of data.

In the case of session setup it is found that in the worst case scenario a 66.35% traffic overhead is incurred. This is mainly due to the SIP routing rules that dictate that all responses must follow the opposite return path as the request itself. Thus while most of these responses are not actually processed by the CPA or the Authorisation Engine they must still be proxied back and forth from the S-CSCF. In addition, the CPA must record-route itself thus several transactions are proxied but only a few of them actually provide useful information.

The large amount of traffic overhead does not result in a significant increase in session setup time. In the worst case scenario the LAN client experienced a 6.11% increase in session setup time. The WLAN client experienced a 3.69% delay increase and for HSDPA, which appears to suffer in general from large session setup times, this overhead drops to an insignificant 0.28%. These results can be attributed to the fact that the proposed architecture does not introduce additional round-trip delays to the session setup procedure.

8.2 Session Volume Analysis Module

The session volume analysis module of the CPA is tasked with determining a spam score for a particular session based on the number of previous sessions of the same type that have been initiated over a set time period. In this section the module is evaluated to determine how it performs when subjected to potential spamming scenarios, and the effects of adjusting the module parameters.

The module has two adjustable parameters that must be set by the network operator for each session type. These parameters are the sliding window period and the maximum number of sessions per minute. The window period parameter determines how quickly the module responds to the calling behaviour of the user in question. The sessions per minute parameter determines the al-

lowable call frequency before users are allocated a non-zero spam score. Once this value is surpassed the module increases the spam score in increments of ten units for every multiple of the allowed sessions until a maximum value of one hundred is reached.

8.2.1 Method

The evaluation is conducted with the SIPp traffic generator acting as both UAC (User Agent Client) and UAS (User Agent Server). The network is configured as described in Case II of Figure 8.1 in which only the CPA is invoked and not the Authorisation Engine.

In order to make reliable comparisons the maximum sessions per minute parameter is held constant at a value of 6 SPM (sessions per minute) throughout these evaluations. This value was chosen because it was found in the available call records that legitimate callers rarely exceed this limit. The evaluations run for a total of 10 minutes each from a cold start, i.e. the CPA has no previous record of the caller.

For each of these sessions the UAC sends an INVITE addressed to the UAS. The UAS immediately responds with a 180 Ringing provisional response followed by a two second delay and then a 200 OK final response. On receiving the final response the UAC sends an ACK request, followed by a five second delay and then a BYE request. The UAS then replies with a 200 OK message confirming the end of the session. Only the *application/sdp* session type is employed throughout the evaluations signifying a voice or video call.

The module outputs its results to a CSV (Comma-Separated Value) text file that is read into Matlab and the values plotted.

8.2.2 Scenarios

In order to demonstrate the effectiveness of the module under different conditions four different scenarios are considered. The scenarios are chosen so as to test the range of spam scores, therefore in each of the scenarios the UAC initially generates sessions at 10 times the allowed rate. The sliding window periods are chosen to fit within the evaluation time. The ability of the module to respond to changing calling behaviour is tested in the last two evaluations.

1. The UAC generates sessions at 60 SPM. The sliding window period is

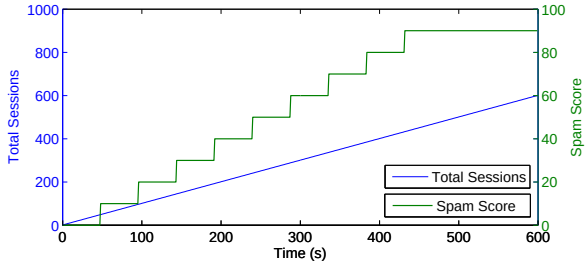

Figure 8.3: Session volume analysis module evaluation results - Scenario 1.

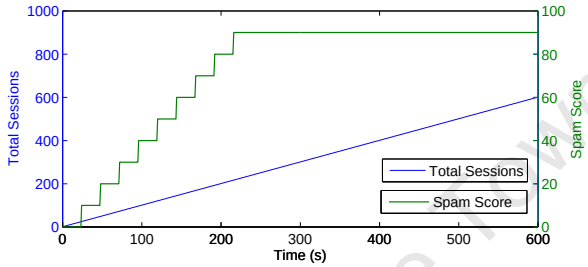

Figure 8.4: Session volume analysis module evaluation results - Scenario 2.

eight minutes.

2. The UAC generates sessions at 60 SPM. The sliding window period is four minutes.

3. The UAC generates sessions at 60 SPM. After five minutes the UAC slows its rate to 30 SPM. The sliding window period is four minutes.

4. The UAC generates sessions at 60 SPM. After five minutes the UAC slows its rate to 30 SPM. The sliding window period is two minutes.

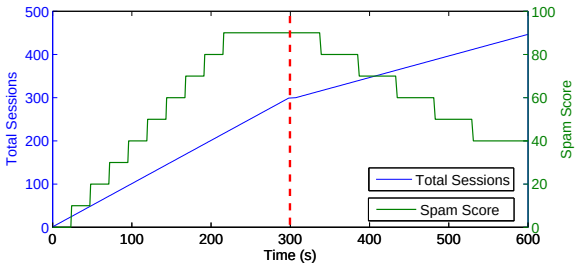

Figure 8.5: Session volume analysis module evaluation results - Scenario 3.

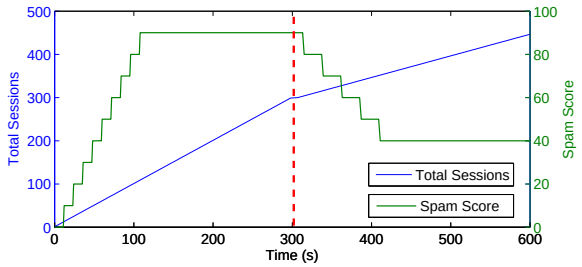

Figure 8.6: Session volume analysis module evaluation results - Scenario 4.

8.2.3 Results

The results of the evaluations for the four scenarios are graphed in Figure 8.3, Figure 8.4, Figure 8.5 and Figure 8.6 respectively. The graphs show the time scale in seconds on the horizontal axis, the total session volume scale on the left side axis and the spam score scale on the right side axis.

8.2.4 Discussion

In Scenario 1 the spam score rises to the value of 90 within eight minutes where it stays for the remainder of the evaluation. The score correctly does not rise to 100 as the spam score remains at zero while the UAC does not exceed the allowed rate. The evaluation demonstrates the module's ability to correctly allocate a high spam score to a caller with a rate much faster than the allowable rate.

In Scenario 2 the caller initiates the same amount of sessions per minute but the spam score rises significantly faster, taking four minutes to reach its maximum score of 90 due to the larger sliding window period. This evaluation demonstrates that the window period can be altered in order to reduce or increase the sensitivity of the module to changing calling behaviour.

In Scenario 3 the spam score rises to its maximum value of 90 after four minutes, and remains there until the caller reduces its sessions per minute. The score begins to drop almost immediately and reduces to 40 approximately nine minutes into the evaluation. The evaluation shows that a caller can reduce their spam score by adopting a more appropriate calling behaviour.

In Scenario 4 the caller initiates the same sessions per minute but this time the spam score reacts faster to the caller's changing behaviour. The score quickly increase to 90 after only two minutes and the spam score reduces to

40 at five minutes into the evaluation. The evaluation again shows that the sliding window affects the speed at which the module reacts to changing calling behaviour.

The results show that a shorter sliding window period is desirable for quick response to spamming attacks, although this also means that the spam score falls rapidly after an attack.

8.3 Concurrent Session Analysis Module

The concurrent session analysis module allocates a spam score to a particular session determined by the current number of simultaneous active sessions. The module has only one adjustable parameter and that is the number of allowable concurrent sessions. In this evaluation the performance of the concurrent session analysis module is tested under various different scenarios.

8.3.1 Method

The network is again configured in the topology described by Case II of Figure 8.1. The SIPp traffic generator is used as a UAS and UAC.

The allowable concurrent session parameter is kept constant throughout the evaluations at a value of four so that each scenario can be compared fairly. This value is chosen arbitrarily, but it is unlikely that a legitimate caller would be involved in more than four concurrent calls or text conversations. Each evaluation is allowed to run for ten minutes from a cold start.

The signalling is identical to the session volume analysis module evaluations in that the UAC sends an INVITE to which the UAS responds with a 180 Ringing response followed two seconds later by a 200 OK response. However, in these evaluations the length of the call is adjusted in the various scenarios in order to show how the session length affects the number of concurrent sessions and the corresponding spam score.

The module outputs its results to a CSV text file and the results are plotted in Matlab.

8.3.2 Scenarios

Four scenarios are considered for evaluation. The scenarios are constructed to evaluate how the frequency of sessions setup and length of session affects

the number of concurrent calls passing through the module and the resulting spam score. The session lengths are selected to test a wide range of spam scores. The first three evaluations are chosen to show that changing either parameters of the calling behaviour can have an effect on the spam score. The final evaluation addresses the ability of the module to respond to a changing calling behaviour.

1. The UAC generates sessions at a rate of 60 SPM. The session length is 15 seconds.
2. The UAC generates sessions at a rate of 60 SPM. The session length is 30 seconds.
3. The UAC generates sessions at a rate of 100 SPM. The session length is 15 seconds.
4. The UAC generates sessions at a rate of 30 SPM that increases by 1 SPM every 5 seconds. The session length is 15 seconds.

8.3.3 Results

The results of the four evaluations are depicted in Figure 8.7, Figure 8.8, Figure 8.9 and Figure 8.10 respectively. The horizontal axis shows the time scale in seconds, the left vertical axis shows the total session scale and the right vertical axis shows the spam score scale.

8.3.4 Discussion

The results of the first scenario show that the module has identified concurrent sessions by the user and, due to the constant session rate and session duration of caller, the spam score remains constant until the end of the evaluation period.

The second scenario shows that despite the caller making an equal number of calls the length of the session affects the number of concurrent sessions identified by the module, and consequently the spam score is significantly higher.

The third scenario shows that the session rate also affects the number of concurrent sessions, and hence the higher spam score. The minor deviations

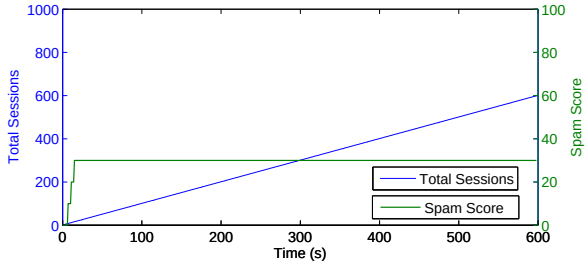

Figure 8.7: Concurrent session analysis module evaluation results - Scenario 1.

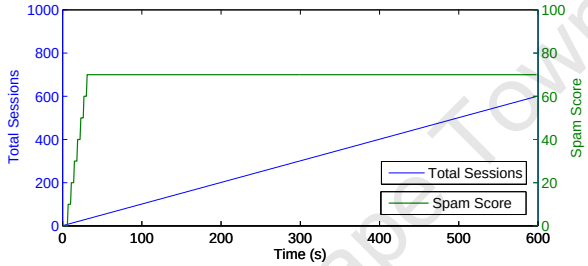

Figure 8.8: Concurrent session analysis module evaluation results - Scenario 2.

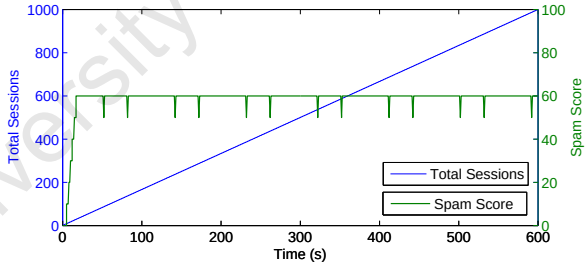

Figure 8.9: Concurrent session analysis module evaluation results - Scenario 3.

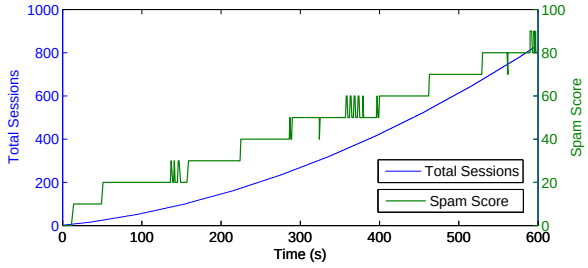

Figure 8.10: Concurrent session analysis module evaluation results - Scenario 4.

in the spam score are due to small fluctuations in the number of concurrent sessions as can be expected in a practical evaluation.

The final evaluation shows the module's capability to react to changing calling behaviour by the caller. As the call rate increases so does the number of concurrent sessions, and hence the calculated spam score. Again minor deviations from the pattern are observed as the concurrent sessions approaches a threshold value.

8.4 Session Duration Analysis Module Function Minimiser

The session duration analysis module requires a model of a known spamming pattern for comparison with observed session history data. For this purpose a PDF model must be interpolated from typical spamming behaviour and fitted to the data using a function minimiser. In this section an evaluation is performed that determines the call duration analysis module's ability to minimise a spammer model function using various initial parameter guesses.

8.4.1 Method

In Chapter 5 an experiment was performed that yielded a data set which typifies spamming behaviour. A model was proposed to describe this data set shown in Equation 8.1, for $0 \leq x \leq x_0$.

$$f(x, p, \lambda, x_0, \mu, \sigma) = \frac{1}{\alpha\sqrt{2\Pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) + \lambda \exp(\lambda(x - x_0) \qquad (8.1)$$

In this evaluation the model is fitted to a histogram of the data using various initial guesses of the five variable parameters: $p, \lambda, x_0, \mu, \sigma$. The function minimiser is allowed to run until the function cannot be minimised significantly further. At each iteration of the minimisation process the new parameters are extracted. The session duration histogram and the model curve are plotted in Matlab at the various stages of the minimisation process.

Table 8.6: Initial parameter values for the session duration analysis module evaluation scenarios.

Parameter	p	λ	x_0	μ	σ
Scenario 1	0.3	0.8	10	2	0.5
Scenario 2	0.7	1.2	10	5	1.5
Scenario 3	0.3	0.7	10	1	1.2

Table 8.7: Results of call duration analysis module evaluation scenarios.

	Scenario 1	Scenario 2	Scenario 3
Total iterations	17	44	7
Sum of least squares (initial)	149.27 x 10^{-3}	92.31 x 10^{-3}	122.38 x 10^{-3}
Sum of least squares (final)	6.73 x 10^{-3}	6.73 x 10^{-3}	22.33 x 10^{-3}

8.4.2 Scenarios

Three scenarios are considered in which the initial guesses of parameters are varied. These parameters are detailed in Table 8.6.

8.4.3 Results

The results of the evaluation are shown in Table 8.7. Table 8.8 shows the final parameter values after a correct fitting. As can be seen in Figure 8.11 in Scenario 1 the minimiser is able to obtain a good fit to the data in relatively few iterations, and the results show that the best fit is found after a total of 17 iterations. On the other hand, Figure 8.12 shows that in Scenario 2 the minimiser struggles to find a good fit, and eventually only converges to a good fit after 44 iterations. This is despite the fact that the parameter guesses for the first scenario actually resulted in a greater initial least squares difference than the second scenario. Figure 8.13 shows the final scenario in which the minimiser quits after seven iterations having not found the correct minimum.

Table 8.8: Final parameter values for a correct fitting.

Parameter	p	λ	x_0	μ	σ
Value	0.57	1.06	10.38	2.97	1.17

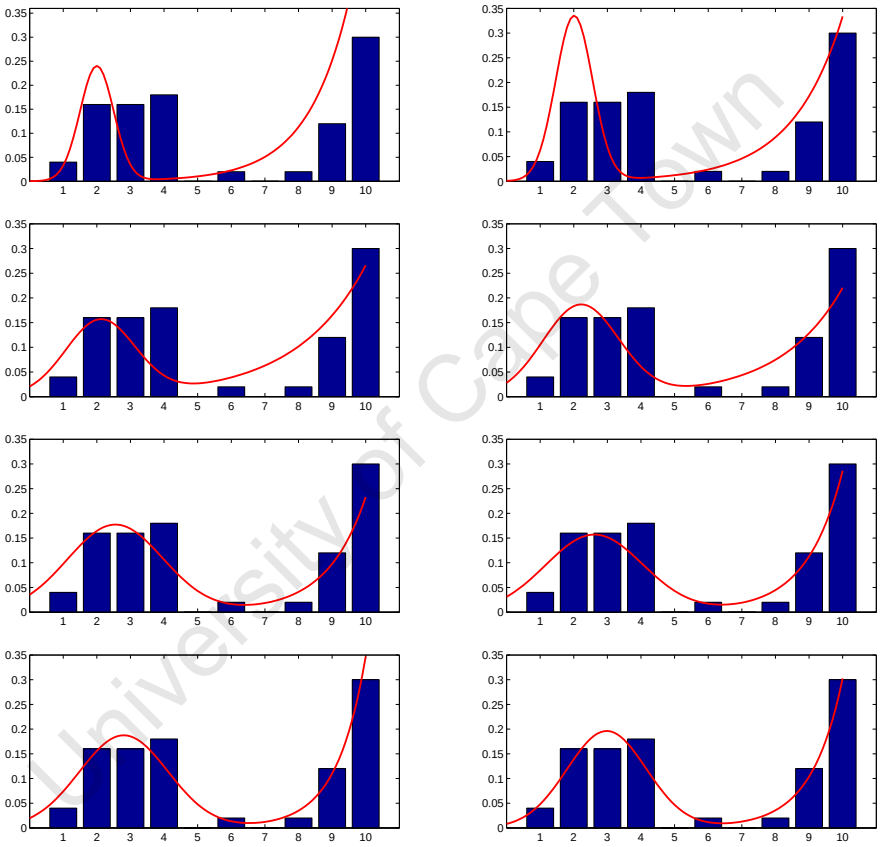

Figure 8.11: Scenario 1 - A good fit of the function f is found after only eight iterations.

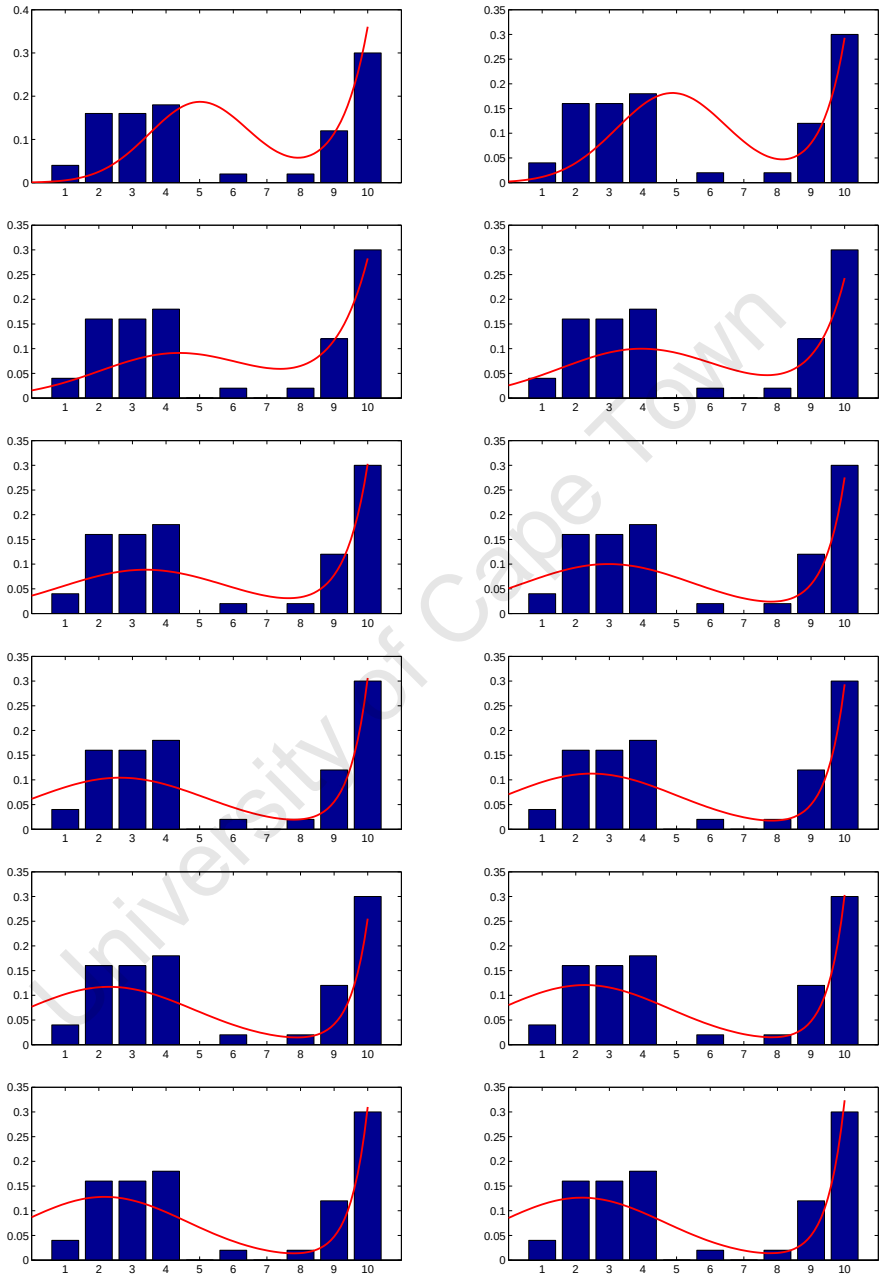

Figure 8.12: Scenario 2 - After twelve iterations shown the minimiser has still not found a good fit, but is successful after 44 iterations.

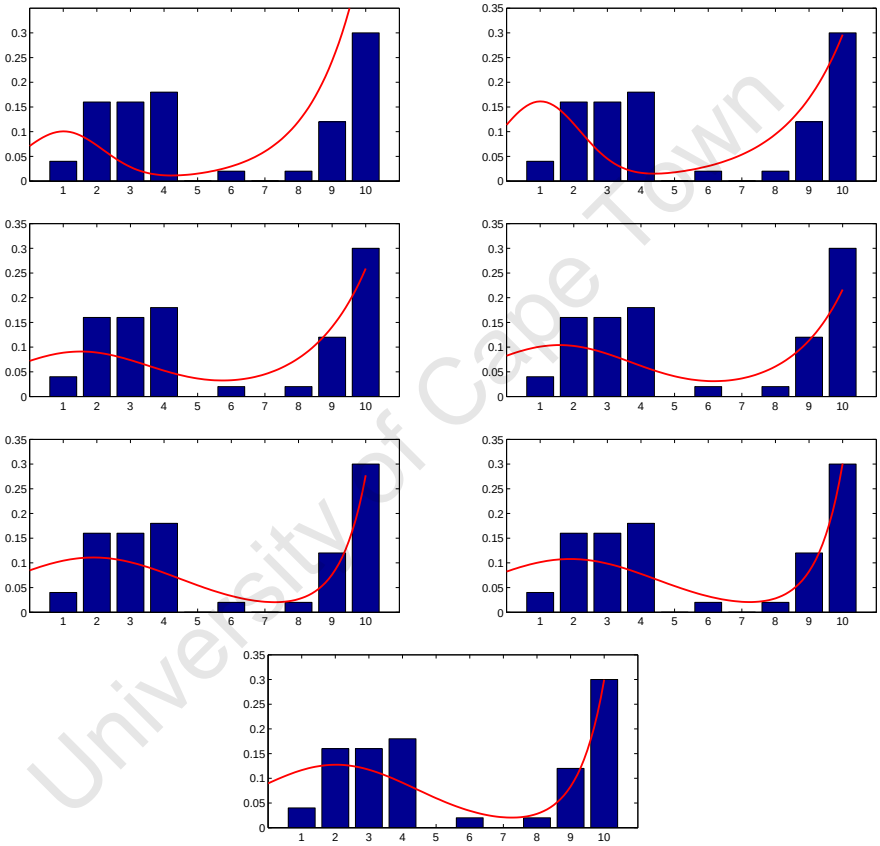

Figure 8.13: After seven iterations the minimiser finds the incorrect minimum and quits.

8.4.4 Discussion

The evaluation has shown two results. First, that the minimiser is implemented correctly and is able to find a correct fit to the data. Second, the initial guess of parameters in very important to ensure that the minimiser converges on a correct solution.

A poor initial guess of parameters can cause the minimiser to find an incorrect minimum and hence fail. Thus is it advisable to perform visual confirmation of correctness whenever applying a new model to the module.

8.5 Session Duration Analysis Module Correlation

The duration analysis module calculates the Pearson product moment correlation coefficient in order to establish an appropriate spam score for the current session.

In this section the fitted curve from the previous evaluation is correlated against the data obtained in the Chapter 5 Robo-caller experiment and several call records of legitimate callers. The aim is to demonstrate that the module correctly allocates a high spam score to the data obtained from the robo-caller experiment and a low score to the data obtained from the legitimate caller records.

8.5.1 Method

Due to the nature of the module it is not practical to perform this particular evaluation in the test-bed as it would require a great deal of time to run each experiment. Thus the evaluation is performed using the numerical analysis tools of Matlab. The code of the session duration analysis module is accurately emulated with a Matlab script.

Data points corresponding to the histogram bins are interpreted as a random variable. These points are normalised and converted to a standard score by subtracting the mean of the sample and dividing by the standard deviation to form a vector of ten values. The same process is then followed for the observed data. The dot product of the two vectors is then calculated. The elements of the vector are then summed and divided by the sample size minus

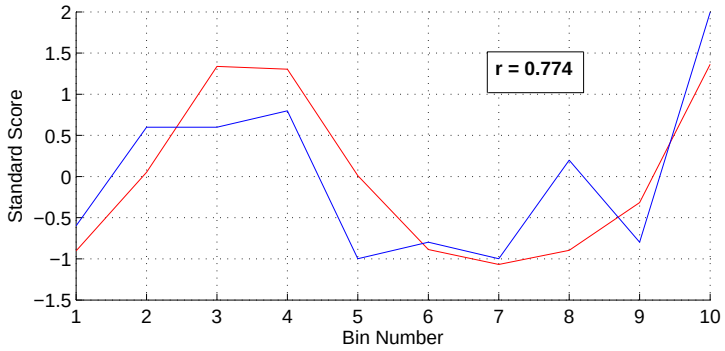

Figure 8.14: The spammer call data correlates well with the model function as expected.

one. This then yields the Pearson Product Moment Correlation value r. This process is followed for the spammer data from Chapter 5 and, for comparison, the call records of fourteen random callers extracted from the call records that are also used in Chapter 5.

8.5.2 Results

The results from the evaluation are shown in Figure 8.14 for the spammer data and Figure 8.15 for the legitimate caller data. The spammer data shows a relatively high correlation coefficient of 0.774, indicating the proposed model is a good fit for this data. The highest correlation coefficient from the sample for legitimate callers is a relatively low 0.216. Many of the legitimate callers actually showed a negative correlation coefficient with the proposed model.

8.5.3 Discussion

The results of the evaluation indicate the the proposed model describes the data observed from the robo-caller experiment fairly accurately and, according to the spam score equation proposed in the previous chapter, results in a spam score of 54.8. The evaluation also shows that of the random callers selected from the available call records, none showed a call history that corresponded closely to the proposed model. This is a good sign as it implies that there is a low chance of a false positive from the proposed model. Unfortunately, without more call history data from spammers it is not possible to comment on the overall effectiveness of the model to spamming behaviour. However, the evaluation has shown that it is possible to automatically calculate a correla-

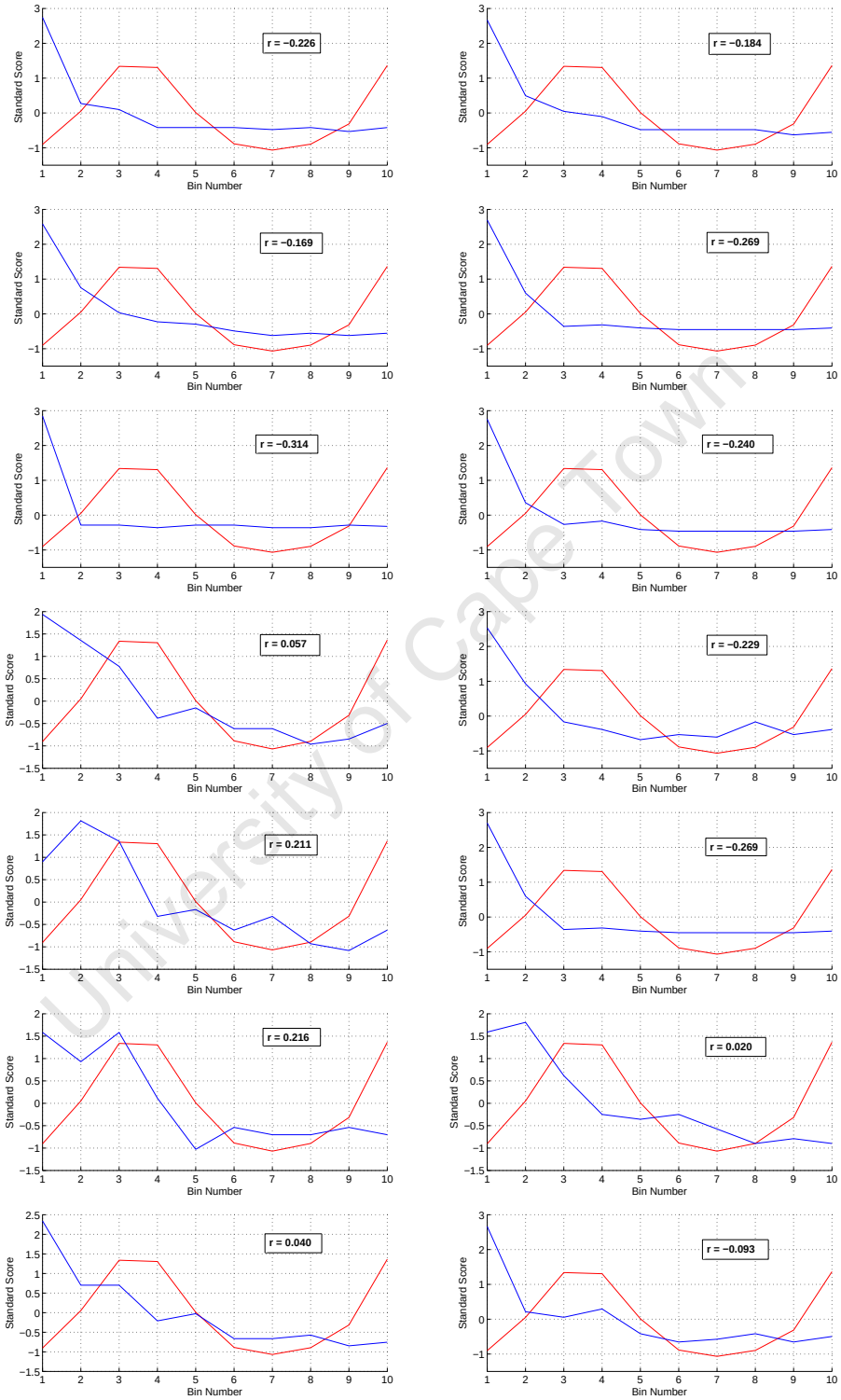

Figure 8.15: Correlation results for fourteen legitimate callers.

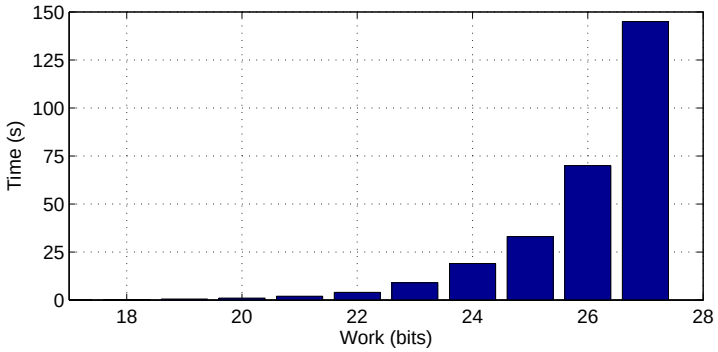

Figure 8.16: Average computation times for the hashcash puzzle.

tion coefficient for observed call history data and consequently the model can allocate an appropriate spam score.

8.6 Computational Puzzle

The computational puzzle is a mechanism to slow the rate of session initiation, and forms part of the proposed spam prevention architecture. The mechanism requires that the caller's user equipment solve a computational puzzle by brute force. This requires a variable amount of time depending on the work parameter of the puzzle.

In this section the effectiveness of the computational puzzle in slowing the rate of session initiation is evaluated for various different values of the work parameter. For reference the average times required to solve puzzles of varying difficulty on the test-bed user equipment is shown in Figure 8.16. Since the algorithm employs a brute-force attack the actual times may vary a great deal if the algorithm is particularly lucky or unlucky.

8.6.1 Method

In this evaluation the network topology is configured as in Case III of Figure 8.1, in which both the CPA and Authorisation Engine are invoked.

A modified version of the UCT IMS Client is used as a UAC, as the SIPp traffic generator is not capable of solving computational puzzles. The UAC attempts to generate sessions at a regular interval - although this might not be possible if it is currently overwhelmed with solving puzzles. However, it will attempt to catch up any missed session generations and should therefore

generate the correct average session rate required.

The authorisation engine acting in user agent mode answers all incoming INVITE requests that do not contain a correct puzzle header with the 419 With Puzzle error response. The response contains a work parameter that tells the UAC how much work must be performed in order to create an appropriate pre-image.

On receiving the 419 response the UAC sends an ACK and begins solving the computational puzzle. Once the puzzle is solved the UAC sends a new INVITE request, this time with the *Puzzle* header correctly filled, thus proving to the Authorisation Engine that the puzzle has been solved.

If the Authorisation Engine receives an Invite with a correct *Puzzle* header it proxies the request to a dummy UAS that answers the request with a 200 OK response, to which the UAC replies with an ACK and then immediately terminates the session.

8.6.2 Scenarios

Three scenarios are considered for the purposes of evaluation:

1. The UAC generates sessions at 60 SPM. The Authorisation Engine challenges the UAC with a puzzle of 19 bits difficulty regardless of the current spam score.

2. The UAC generates sessions at 60 SPM. The Authorisation Engine challenges the UAC with a puzzle of 20 bits difficulty regardless of the current spam score.

3. The UAC generates sessions at 60 SPM. The Authorisation Engine challenges the UAC with a puzzle of variable difficulty depending on the spam score, where $work = 15 + SpamScore/10$.

4. The UAC generates sessions at 60 SPM. The Authorisation Engine challenges the UAC with a puzzle of variable difficulty depending on the spam score, where $work = 16 + SpamScore/10$.

8.6.3 Results

The results from the four evaluation scenarios are shown in Figure 8.17, Figure 8.18, Figure 8.19 and Figure 8.20 respectively.

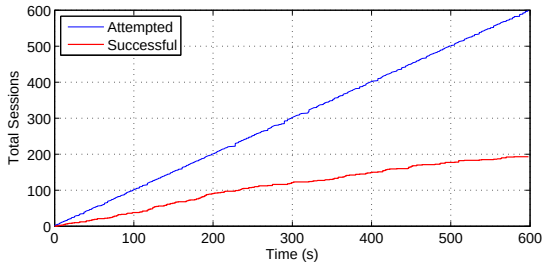

Figure 8.17: Results of computational puzzle evaluation - Scenario 1.

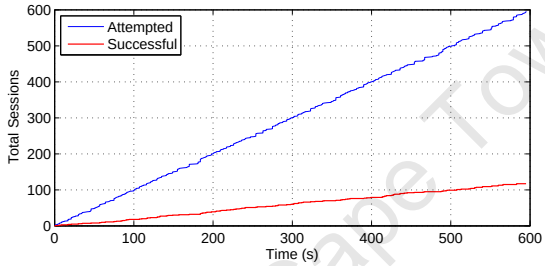

Figure 8.18: Results of computational puzzle evaluation - Scenario 2.

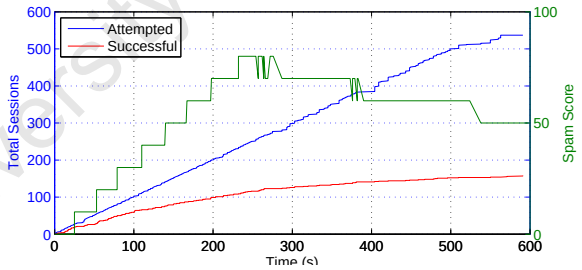

Figure 8.19: Results of computational puzzle evaluation - Scenario 3.

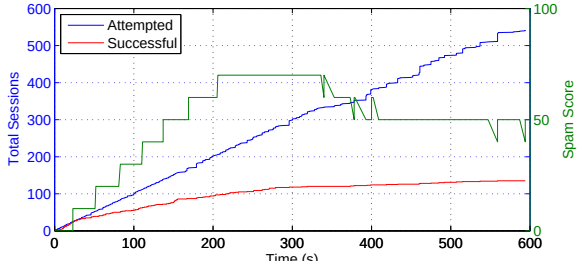

Figure 8.20: Results of computational puzzle evaluation - Scenario 4.

8.6.4 Discussion

In the first two scenarios is can be seen how the computational puzzle is able to slow down the rate of successful session generation for the potential spammer.

In the third and fourth scenarios the spam score is used to determine the computational puzzle difficulty as would be expected in a practical setting. In both cases the spam score of the caller rises quickly but as the UAC struggles to solve the computational puzzle, and consequently generates less sessions, so the spam score begins to moderate. These scenarios illustrate how the proposed architecture is able to adapt over time to changing session initiation behaviour.

8.7 Turing Test

The video and audio Turing tests are used in rare instances in which a particular session attracts a very large overall spam score - in this implementation when the score is greater than 90. The Authorisation Engine answers the call on behalf of the end user and challenges the caller with the Turing test in order to ensure that they are human and not an automated robo-caller.

Unfortunately this requires that the server that hosts this component of the Authorisation Engine is required to handle a great deal of simultaneous audio and video sessions in the process of challenging callers. In this evaluation the performance of this server is evaluated in order to determine typical bandwidth requirements for a server of this type, and hence determine how scalable the solution is for a very large network of many millions of users. The signalling traffic for this service is negligible compared to the media, therefore the evaluation is focused at the bearer level.

8.7.1 Method

The Turing test server is supplied with a CAPTCHA video with dimensions 384x288 stored in Microsoft Windows Media 9 format. The video includes a video and audio component so that it is compatible with all multimedia devices. The SIPp traffic generator initiates a session that is answered on behalf of its recipient by the Authorisation Engine's Turing test server. The server transcodes the CAPTCHA video file into the H263 video codec and MPEG layer 1 audio codec, then continuously transmits the streams to the UAC over RTP.

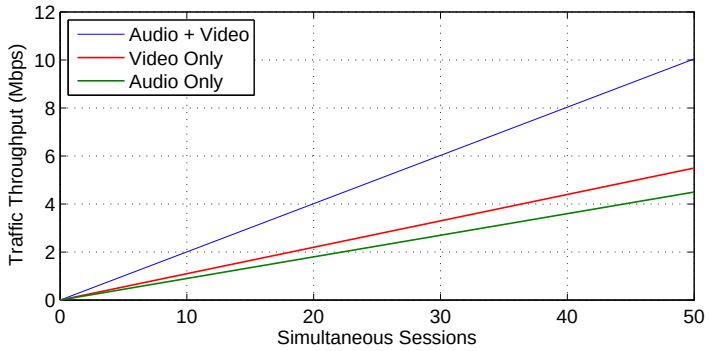

Figure 8.21: Media traffic throughput at the Turing test server.

For the purposes of this evaluation the stream continues indefinitely. The traffic throughput at the Turing test server is then measured for 30 seconds by Wireshark and an average value calculated. At each iteration of the evaluation the number of SIPp UACs is increased and again the traffic is measured. The evaluation is then repeated with audio-only and video-only versions of the CAPTCHA.

The CPU load is also observed throughout the evaluation, using the Linux *pidstat* utility that monitors individual tasks managed by the kernel.

8.7.2 Results

The results from the Turing test server evaluation are shown in Figure 8.21. The server requires 10.043 Mbps of bandwidth in order to stream the CAPTCHA with both audio and video components to 50 users. The audio-only version requires only 4.527 Mbps and the video-only version requires 5.517 Mbps for the same number of users. It was found that the CPU load fluctuated but did not exceed 10% at any time during the evaluation.

8.7.3 Discussion

The results indicate that the server requires a large amount of available bandwidth in order to serve relatively few simultaneous Turing tests, indicating that in its current form it is not very scalable. It is expected that only a very small percentage of sessions will be subjected to this type of test, however, in a very large network even a small percentage of sessions results in a large load on the server.

There are several options for reducing this load. First, the server should wherever possible use only an audio or video CAPTCHA and not the combined version. Second, it is possible that a static image can be sent to the user instead of a streaming video feed if the client terminal supports it. Third, lower quality audio and video codecs can be used to reduce the stream size, however this option should be approached with caution, as the danger exists that the test may become hard to solve if the image is grainy and the sound badly distorted.

It was found that the CPU load was not excessive. However, it is still possible to reduce this load by performing pre-processing on the input audio and video files so that they do not need to be trans-coded at run-time.

8.8 Summary

This chapter has subjected the implementation of the proposed spam prevention architecture to numerous evaluations, and in doing so has shown proof-of-concept as well as key performance data. Most importantly, however, the chapter has shown that the architecture is able to use the various modules of the Call Pattern Analyser to allocate a spam probability score to incoming sessions, and use this score in the process of making authorisation decisions. Of course authorisation decisions can be made in other ways, for example by means of whitelists and blacklists, but this behaviour is not as interesting and hence was omitted from the evaluations.

The modular design of the architecture allowed for individual components to be tested separately with the result that the performance of each module could be gauged on its own. It has been shown that the implementation of the proposed architecture was successful, and the results indicate very few shortcomings in the design.

In the next chapter the thesis is concluded and recommendations are made, based on the results of the evaluations, for possible improvements to the proposed architecture.

Chapter 9

Conclusions and Further Study Items

The preceding chapters have performed an in-depth study of the current and future outlook of the multimedia spam problem that is expected to manifest itself within the next generation network, specifically the emerging IMS framework. In this chapter conclusions are drawn from the findings of the thesis and items for further study in this field are identified.

9.1 Conclusions

9.1.1 IMS Adoption

The introduction to this work presented an overview of the VoIP revolution and the steady migration to all-IP network architectures. It was found that there are currently many network operators that have expressed an interest in the benefits of a service delivery platform, as opposed to the stove-pipe service creation model that has been adopted for many years in the legacy PSTN. As such, standardisation bodies including 3GPP, 3GPP2, ETSI TiSPAN and Packet Cable have all begun the process of compiling the specifications for a packet-switched framework that provides PSTN emulation and rich new services. One particular technology, the 3GPP IP Multimedia Subsystem, has enjoyed considerable support recently, and many standards bodies have reused substantial parts of the 3GPP's work in their own specifications. Therefore it can be concluded that in the near future several large network providers will begin deploying IMS infrastructure in their networks. It is unclear, however,

exactly how long this will take, or whether this IMS deployment will replace their existing technologies or used in conjunction.

9.1.2 IMS Challenges

The thesis found that IMS is a relatively new technology, and consequently there are still many challenges that must be overcome before it will see widespread adoption. Network providers have made huge capital expenditures in existing circuit-switched equipment and will aim to recover these costs for many years to come. However, the packet-switched network architecture of the IMS brings with it substantial cost savings, both for the operator and the end user. This results in cheaper telecommunication costs but also in a greater risk of the technology being used for marketing, fraud and phishing, as has been witnessed extensively in the modern-day Internet. Thus it is concluded that the IMS, while offering a multitude of cheap new services to its potential users, will also subject them to an invasion of their privacy and potentially make them susceptible to fraud and identity theft. Therefore network operators must consider these problems when formulating their NGN strategies.

9.1.3 Spam No Longer Just Text

It was found that roughly 80% of email worldwide is unsolicited bulk marketing commonly referred to as spam. The problem of spam has recently encompassed other technologies such as instant messaging systems and voice calls. This is a clear indication that spammers are no longer satisfied with the returns from email spamming and are looking to branch out to more lucrative forms of advertising. SIP has been identified as a prime target for spammers due its similarities with email, especially the familiar addressing schemes. It is possible to create automated SIP callers with no expert knowledge due to the availability of free SIP software and programming APIs on the Internet. Other Internet voice platforms, such as Skype, are just as much at risk as SIP providers.

Mobile network users are also susceptible to spamming, and it was found that there have been many instances of spamming behaviour by means of cellular telephone text messaging. Spam has also been targeted at multimedia handsets that can play video and audio messages to the user.

Thus it is clear that the problem of spam has grown to encompass many different forms and will continue to be distributed over many different technologies. Focusing on a single type of spam is no longer sufficient, and security providers must offer solutions that cover the wide range of VoIP and rich Internet-like services.

9.1.4 State of the Art

The review of literature in the field found that there is little in the way of research focused on the spam problem in NGNs. Although the 3GPP has dedicated a work item to combating spam, the scope of the work item is very limited and, as yet, few deliverables have emerged.

However, there have recently been several Internet drafts published by the IETF looking at the problem from an Internet perspective, mostly authored by well-known researchers including Rosenberg, Schulzrinne and Jennings, amongst others. Fortunately, the 3GPP has largely looked to the IETF to provide protocols for the IMS framework, for example SIP, HTTP and Diameter. Thus it is possible for the work performed in solving spam in the Internet to be applied to the NGN setting, albeit with a few modifications. For example, the IMS has devised specific methods for service invocation which will not necessarily be relevant in the Internet. Moreover, IMS offers stronger authentication and authorisation than normal Internet nodes and this can be leveraged in a spam prevention solution. Research into multimedia spam at this time should be founded on the specifications of the IETF as currently this is the most advanced work in this field.

9.1.5 Types of Spam

Spam has often been referred to as bulk commercial email because it has long since been used as a mechanism to market prescription drugs, pornographic content and illegal software. However, this thesis has discussed several case studies that show spam being used as a tool by fraudsters, criminals and political parties. In fact it was found that in some parts of the world spamming has now been linked to organised crime.

Nowadays, spam is frequently used to manipulate the stock market with the well-known pump-and-dump scam, where the perpetrators of these scams often

reap healthy profits. Spam has also been used to spread false business information, causing company stocks to plummet in minutes, and as a political tool particularly in the USA, in the form of electioneering, attack-advertisements and rumour spreading.

These cases demonstrate that the motives for spamming are far more varied than simple marketing. It is no longer correct to assume that all spam is sent for commercial gain, but instead may be a tool for crime, fraud and questionable political tactics. It can be concluded that spam is no longer just an annoyance for users but also a threat to their finances and security.

9.1.6 Spam Legislation

The thesis found that current spam legislation is limited, particularly in South Africa. South African legislation deals primarily with bulk commercial email, and as discussed already, this is no longer sufficient. Furthermore, the legislation does not specifically ban the practice of unsolicited communication but rather provides provisions for performing it legally.

Studies in the USA have shown that legislation has had some success in reducing spamming and it allows victims legal recourse against persistent transgressors. It is concluded that current legislation in South Africa is inadequate and should be reformed as a matter of urgency in order to take into account current and future threats.

9.1.7 Perceptions on Spam

The survey conducted as part of this thesis aimed to discover the current state of multimedia spam and the attitude of the participants towards spam in general. It was found that although email is still the mostly widely received form of unsolicited communication, many participants reported receiving a great deal of unsolicited SMSs and voice calls. Overall, the participants showed a strong negative attitude towards spam. When questioned on who was responsible for solving this problem the vast majority stated that is was the up to the network service providers.

Several conclusions can be drawn from this survey. First, there is currently a serious problem of unsolicited communications in many different forms in South Africa and the problem appears to be getting worse. Second, consumers in general thoroughly dislike receiving spam. Third, consumers hold the service

providers responsible for the problem and it is the service providers who will bear the brunt of their inevitable discontent.

9.1.8 Response to Robo-calling

A robo-calling experiment was performed in order to determine typical human behaviour on receiving an automated call. The experiment was designed to calculate two important factors: the number of people that respond to automated voice solicitations and the length of time that they are willing to listen to a pre-recorded message.

A website was created so that the response ratio could be measured. It was found that sixteen percent of the experiment participants visited the website after receiving the call. This is an impressive response ratio, compared to the response ratio to email-based spam that is typically far less than one percent. Thus, it can be concluded that voice spam can be particularly effective as a marketing tool.

The second goal of the experiment, to measure the amount of time that people would listen to the call, resulted in some interesting results. It was found that a first category of participants would listen for a period of time and then, on realising that the call was pre-recorded and not important, would terminate the call. These participants would generally terminate the call within the first ten to thirty seconds.

It was found that a second category of participants would listen to the message all the way to the end. A histogram of these call durations was compared with similar histograms of legitimate callers and it was found that there were clear differences in the call duration patterns. Thus the experiment showed it is possible that spamming behaviour patterns can be identified by performing call duration analysis.

9.1.9 Test-bed Implementation

The thesis proposed an architecture for the prevention of spam in the IMS and this architecture was implemented in a practical test-bed framework. One of the primary goals of the implementation was to make it as reproducible as possible so that future researchers would be able to recreate the testing environment for their own evaluations. In this regard a decision was made to use open source software wherever possible for the test-bed components.

The test-bed framework was constructed from the Open Source IMS Core project, the OpenXCAP project, application servers based on the open source oSIP and eXosip libraries, the SIPp traffic generation tool, and the UCT IMS Client project. Each of the elements are available freely with their source code and can be modified legally and easily. This enabled a rapid and accurate implementation of the proposed architecture in a practical setting. Thus the thesis finds that it is possible to create a fully functional IMS network emulation with the exclusive use of open source software.

9.1.10 Solution Overheads

The thesis performed an evaluation of the traffic and delay overheads associated with the implementation of the proposed architecture in a practical setting. The evaluations considered the overheads introduced to IMS-level registration with the core network, and to the session setup procedure. All experiments were performed on the network test-bed framework.

It was found that in the worst case the proposed solution added 5.36% increase in network traffic during IMS-level registration. This is considered to be a tolerable overhead. On the other hand, the proposed architecture adds a 66.35% increase in network traffic during session setup. This is a significant increase in network traffic that is partly accounted for by the use of two application servers instead of one. However, much of the blame can be imparted to the SIP routing rules that cause many unnecessary message transmissions between the IMS core and the application servers. The Call Pattern Analyser is interested in only three of the SIP messages, however, it must proxy all eleven requests and responses required for IMS session setup in order to be compliant with the SIP routing rules.

The evaluations found that the increase in traffic did not cause a proportional increase in session setup delay. In fact in the worst case scenario the increase in delay was only 6.11% when the client was attached via an Ethernet LAN and in the case of HSDPA this reduced to only 0.28%. This was primarily because no new round-trip delays were introduced.

Thus it can be concluded that although the proposed architecture increases the core network traffic significantly during session setup, the resulting delay, and hence the intrusion on the end user, is minimal.

9.1.11 Session Volume Analysis

Session volume analysis has been proposed in the literature as a method to identify spamming patterns. In this thesis a session volume analysis module was included in the Call Pattern Analyser. Unlike the previous works that relied on simulation to show the effectiveness of session volume analysis, in this thesis the concept was evaluated in a practical network setting.

It was found that it is possible to vary the parameters of the module so as to respond differently to similar calling behaviour. For example, by altering the allowable sessions parameter the module could assign a higher or lower score to a particular users calling history, and by adjusting the sliding window period under consideration the module could be tuned to react quicker or slower to changing calling patterns.

The evaluations have proven that the module is effective in a practical setting and that it is important to correctly set the module's parameters in order to assign a suitable score to a particular session. An examination of call records found that a fair limit was six voice calls or instant message conversations per minute but operators may wish to increase this slightly to ensure that legitimate callers are rarely inconvenienced.

9.1.12 Concurrent Session Analysis

The concurrent session analysis module detects the start and end of sessions and in doing so can calculate the number of simultaneous sessions in which the user is currently involved. This module was evaluated in the practical test-bed framework.

The experiments showed that the module was able to successfully assign a suitable spam score to the current session based on the number of concurrent sessions. It was found that two factors influenced the number of concurrent sessions made by an automated traffic generator. The first is the frequency of session generation and the second is the length of the sessions; with a rise in either factor increasing the number of concurrent sessions. The modules allowable concurrent sessions parameter can be adjusted according to the network operator's preference. The evaluations allowed for four concurrent sessions but this value was chosen arbitrarily. In a practical setting this value will almost certainly need to be higher.

It is intuitive that a spammer that initiates more sessions and sessions of

greater length is more troublesome than one that initiates infrequent short sessions. The evaluations showed that a higher frequency of session generation and longer session times increased the number of concurrent sessions and the corresponding spam score. Therefore, it can be concluded that a concurrent session analysis module is an effective tool in identifying the worst spamming offenders.

9.1.13 Session Duration Analysis

The session duration analysis module was designed and implemented to assign a spam score according to the past session records of a particular user. Two processes were identified in order to achieve this. A model of known spamming behaviour must be constructed and this model must be applied to a user's session history and checked for similarity.

The first process involved creating a function minimiser to fit a probability distribution function to the data from a known spamming attack. It was found that the minimiser worked well in most cases but not if the initial guess of parameters was not good. Therefore, it can be concluded that the outcome of this process must be validated by human inspection.

The second process required that the Pearson product moment correlation coefficient be calculated between the model and the observed data. The observed data was stripped of outliers and then a histogram was created. This histogram was normalised and then compared with the model. This process was successfully implemented in the test-bed framework. The implementation proved that the module could correctly perform the required calculations and hence assign an appropriate spam score. It was found that the module correctly assigned a high spam score to the sample spammer data and low spam score to the records of randomly chosen legitimate callers. Due to the promising results it is concluded that the proposed mechanism may be a viable option for performing session duration analysis.

9.1.14 Computational Puzzle

The computational puzzle forms part of the proposed spam prevention solution and is designed to slow the rate of session generation by a potential spammer. A working implementation was produced using the mechanisms proposed in

the literature . This implementation was subjected to several evaluations to validate that it is possible to reduce the volume of successful sessions by a user.

The evaluations showed that the mechanism was able to slow the rate of session setup significantly. They also showed that by adjusting the amount of work required by the puzzle that the session rate can be controlled. The solution proposed in this thesis requires that the work parameter is adjusted dynamically according to information supplied by the Call Pattern Analysis modules. This scenario was also evaluated and it was found that the work difficulty was correctly adjusted as the caller slowed their rate of successful session setups. It is concluded that the computational puzzle and dynamic work adjustment are effective tools in reducing session setup rates.

9.1.15 Turing Test

The last layer of defence in the proposed solution is the Turing test that seeks to verify whether a caller is human or robotic. This mechanism was implemented in the test-bed framework and was evaluated in order to determine how scalable the solution is in a large network environment.

It was found that the server on which the solution is implemented must sustain a high volume of media traffic in order to serve the Turing tests. However, the processing power required was not a limitation.

Therefore, it is concluded that the media component of the Turing test should be made as small as possible, while still maintaining good quality, and should be distributed across several media servers in order to ensure scalability.

9.2 Further Study Items

9.2.1 Survey Data

In this thesis it was found that there is a shortage of data regarding the severity of the multimedia spam problem. There are many reports of outraged groups and individuals who have reported spam but few studies showing exactly how many spam messages are sent each year and in what form these messages are being transmitted. Without this data it is hard to know how serious a problem is being faced and where the greatest research efforts should be focused. It is recommended that surveys be conducted that encompass a large population

in order to determine how prevalent spam is in its various forms and how it affects the utility of using electronic communications devices.

9.2.2 Spam Data

Apart from the lack of data that surrounds the severity of multimedia spam there is also a shortage of data regarding how spamming attacks are performed. This thesis proposes to use models based on previous attacks in order to identify future ones. However, due to the current lack of records from robo-calling, text message and other spam campaigns, it is impossible to create accurate models. Therefore it is recommended that there be efforts to identify spamming attacks and to make this data available to researchers so that solutions can be found.

9.2.3 Text Message Filters

The architecture proposed in this work focused primarily on real-time voice and video sessions, however, as discussed in the conclusions above, the IMS offers many new services that may be targeted by spammers. Although some consideration was afforded to text messaging in terms of limiting the volume of messages, no specific content filters were proposed or implemented. The thesis has discussed some plausible solutions, including the use of keyword detectors and Bayesian filters but these have yet to be proved in an IMS network. In future work it is recommended that these solutions be implemented and tested in the IMS environment to handle both pager-mode and session-based instant messaging. In addition to text messaging work must also be performed to address presence-based spam and other IMS services that may be targeted.

9.2.4 Call Duration Analysis

The thesis has proposed and tested a call duration analysis module that attempts to apply a known spamming model to a caller's history. However, it is possible that there are more accurate ways to identify suspicious behaviour. Another option may be to apply a model of a legitimate caller and look for deviations. Alternatively, a simpler approach could be utilised that looks for a large number of calls terminating at the same time.

From the preliminary results in this work it appears that call duration analysis can be a powerful method for identifying voice and video spam. It is recommended that future research attempts to determine the optimal technique for call duration analysis.

9.2.5 Computational Puzzles

The computational puzzle used in this work was shown to successfully slow the rate of session initiation. However, these puzzles have been criticised in the past as it is felt that fast desktop machines are able to solve the puzzles faster than slower handheld devices. This thesis proposed a solution to this problem in the form of a dynamic work parameter that increases depending on the current spam score. Thus if the device is able to solve puzzles faster, it will generate sessions faster, and hence generating a higher spam score and consequently an exponentially harder puzzle. This reverse feedback mechanism was tested in the evaluations but further research must be performed to confirm that it does indeed treat all devices fairly.

Alternatively, there exists the option to use puzzles that rely on memory bus speed, which has been shown to be fairly constant between devices, although there is still a fair amount of variability between devices. A further issue is that computational puzzles of any type drain battery power in mobile devices. Therefore, future research should aim to find fair ways to slow the rate of session setup, without affecting legitimate callers.

9.2.6 CAPTCHAs

Static CAPTCHA images were used in this thesis as a Turing test by converting them into video streams of a single repeating frame. However, CAPTCHAs were originally designed to be used in the Internet as static images and not as videos. Artificial intelligence algorithms have become more advanced at solving static CAPTCHA images and as such they have had to become more complex and consequently harder for humans to solve. A video CAPTCHA can avoid artificial intelligence algorithms by dividing the CAPTCHA string by time so that each frame of the video contains only a portion of the test. Furthermore, a few very brief dummy frames can be included in order to further confuse any attempts to solve the test by machine.

Future work in CAPTCHA technologies should look into video-based tests as this can potentially make them harder for machines to solve but easier for humans to solve. This would reduce the intrusion caused by the tests and make them a more viable option for widespread deployment.

9.2.7 3GPP Standardisation Efforts

The literature review found that there is substantial efforts within the IETF to prevent SIP spam and little work being performed by the 3GPP in this field. It was also found that IMS spam is almost certain to become problem in the near future for network operators and users alike.

Unfortunately, SIP and IMS are not equivalent and as such the works of the IETF, while relevant, cannot be applied directly in an IMS setting. Therefore, it is recommended that as a matter of urgency the 3GPP should take the work performed by the IETF and adapt it for their own specifications as well as begin their own research into solutions. The risk of leaving the problem unsolved is too great.

Bibliography

[1] Gstreamer - Open Source Multimedia Framework. *http://gstreamer.freedesktop.org*.

[2] OpenXCAP. *http://www.openxcap.org*.

[3] SIPp. *http://sipp.sourceforge.net/*.

[4] Ubuntu Linux. *http://www.ubuntu.com/*.

[5] Wireshark Network Protocol Analyser. *http://www.wireshark.org/*.

[6] Controlling the Assault of Non-Solicited Pornography And Marketing Act. *http://uscode.house.gov/download/pls/15C103.txt*, 2003.

[7] Christopher Abad. The Economy of Phishing: A Survey of the Operations of the Phishing Market. *First Monday*, 10(9), September 2005.

[8] Statistics South Africa. General Household Survey. *Statistical Release P0318*, July 2005.

[9] Open Mobile Alliance. OMA Provisioning Architecture Overview version 1.1. TS. *Open Mobile Alliance*, November 2002.

[10] Michael Arrington. Engadget Knocks $4 billion off Apple Market Cap on Bogus iPhone email. *http://www.techcrunch.com*, May 2007.

[11] Charles Babington and Alec MacGillis. It's a Candidate Calling - Again. *The Washington Post*, page A08, November 2006.

[12] Adam Back. Hashcash - A Denial of Service Counter-Measure. *Tech Report*, August 2002.

[13] Drew Bird. Think Spam Is Tough? Try Fighting Spim. *http://itmanagement.earthweb.com/secu/article.php/3365931*, June 2004.

[14] Adam Black. Hashcash. *http://hashcash.org/*.

[15] M. Terre Blanche, K. Durrheim, and D. Painter. *Research in practice: Applied methods for the social sciences*. Cape Town: UCT Press, 2nd edition, 2006.

[16] E.F. Borgetta and M.L. Borgetta. Encyclopaedia of Sociology. *New York: Macmillan*, 1992.

[17] G. Camarillo and M.A. Garcia-Martin. *The 3G IP Multimedia Subsystem (IMS)*. John Wiley & Sons Ltd, second edition, 2006.

[18] Federal Trade Commission. Compliance with Do Not Call Registry Exceptional. *http://www.ftc.gov/opa/2004/02/dncstats0204.shtm*, February 2004.

[19] Spamhaus Blocklist (SBL) Database. Spamhaus Statistics : The Top 10. *The Spamhaus Project Ltd. (dynamic report) - http://www.spamhaus.org/statistics/countries.lasso*, October 2007.

[20] S. Deering and R. Hinden. Internet Protocol, Version 6 (IPv6) Specification. *IETF RFC 2460*, December 1998.

[21] D. Eastlake and P. Jones. US Secure Hash Algorithm 1 (SHA1). *IETF RFC 3174*, September 2001.

[22] eBay Inc. eBay Second Quarter 2007 Financial Results. *http://investor.ebay.com/results.cfm*, July 2007.

[23] Deborah Fallows. CAN-SPAM a year later. *Pew Internet and American Life Project*, April 2005.

[24] P. Faltstrom and M. Mealling. The E.164 to Uniform Resource Identifiers (URI) Dynamic Delegation Discovery System (DDDS) Application (ENUM). *IETF RFC 3761*, April 2004.

[25] Fraunhofer Institute FOKUS. The Open Source IMS Core Project. *http://www.openimscore.org/*.

[26] Susannah Fox and Jean Beier. Online Banking 2006: Surfing to the Bank. *Pew Internet and American Life Project*, June 2006.

[27] C. Foxcroft and G. Roodt. *An introduction to psychological assessment in the South African context.* Oxford: Oxford University Press, 2001.

[28] Laura L. Frieder and Jonathan L. Zittrain. Spam Works: Evidence from Stock Touts and Corresponding Market Activity. *Berkman Center Research Publication No. 2006-11*, January 2007.

[29] GFi. Attachment spam - The Latest Trend. *http://www.gfi.com/whitepapers/attachment-spam.pdf*, 2007.

[30] Paul Graham. *Hackers and Painters.* O'Reilly, 2004.

[31] R.J Gregory. *Psychological Testing.* Allyn and Bacon, 1992.

[32] Miniwatts Marketing Group. World Internet Usage and Population Statistics. *http://www.internetworldstats.com/stats.htm*, September 2007.

[33] Florian Hammer, Peter Reichl, and Alexander Raake. Elements of Interactivity in Telephone Conversations. *Proceedings of the 8th International Conference on Spoken Language Processing*, 3:1741–1744, October 2004.

[34] Magnus R. Hestenes and Eduard Stiefel. Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards*, 46(6), 1952.

[35] Thomas J. Holt and Danielle C. Graves. A Qualitative Analysis of Advance Fee Fraud E-mail Schemes. *International Journal of Cyber Criminology*, 1(1):137–154, January 2007.

[36] Boris Iglewicz. *How To Detect And Handle Outliers.* Asqc Basic References in Quality Control, 1993.

[37] PUREmail Inc. Image Spam - The New Face of Email Threats. *http://www.puremail.com/wpImageSpam.php*, 2007.

[38] Harris Interactive. Do Not Call Registry Is Working Well. *The Harris Poll #106*, October 2007.

[39] C. Jennings. Computational Puzzles for SPAM Reduction in SIP - draft-jennings-sip-hashcash-06. *IETF Internet Draft (work in progress)*, July 2007.

[40] C. Jennings, J. Peterson, and M. Watson. Private Extensions to the Session Initiation Protocol (SIP) for Asserted Identity within Trusted Networks. *IETF RFC 3325*, November 2002.

[41] A.E. Kazdin. The Encyclopedia of Psychology. *Oxford University Press/American Psychological Association*, 2000.

[42] J. Kiefer. Sequential minimax search for a maximum. *Proceedings of the American Mathematical Society*, 4:502–506, 1953.

[43] Brian Krebs. Shadowy Russian Firm Seen as Conduit for Cybercrime. *The Washington Post*, 2007.

[44] Vodafone Group R&D Lab. Vodafone Mobile Connect Card driver for Linux. *https://forge.vodafonebetavine.net/projects/vodafonemobilec/*.

[45] Mary Madden and Lee Rainie. Music and Video Downloading Moves Beyond P2P. *Pew Internet and American Life Project*, March 2005.

[46] Jessica Marshall. Hackers could skew US elections. *NewScientist.com News Service*, October 2007.

[47] Raymond McConville. IMS Goes From Hero to Zero. *Light Reading*, July 2007.

[48] Aymeric Moizard. The GNU oSIP library. *http://www.gnu.org/software/osip/*.

[49] BBC News. Clampdown on 'missed call' scam. *http://news.bbc.co.uk/2/hi/uk_news/3499337.stm*, February 2004.

[50] Parliament of the Republic of South Africa. Electronic Communications and Transactions Act. *http://www.acts.co.za/ect_act/*, July 2002.

[51] Postinini White Paper. The Shifting Tactics of Spammers: What you need to know about new email threats. *http://www.spamwash.com/whitepapers/WP10-01-0406_Postini_Connections.pdf*, 2004.

[52] J. Peterson and C Jennings. Enhancements for Authenticated Identity Management in the Session Initiation Protocol (SIP). *IETF RFC 4474*, August 2006.

[53] M. Poikselka, G. Mayer, H. Khartabil, and A. Niemi. *The IMS - IP Multimedia Concepts and Services in the Mobile Domain.* John Wiley & Sons Ltd, first edition, 2004.

[54] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 2nd edition, October 1992.

[55] Third Generation Partnership Project. Internet Protocol (IP) multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3. *3GPP TS 24.229*, December 2007.

[56] Third Generation Partnership Project. IP Multimedia (IM) session handling; IM call model; Stage 2. *3GPP TS 23.218*, December 2007.

[57] Third Generation Partnership Project. IP Multimedia Subsystem (IMS); Stage 2. *3GPP TS 23.228 v8.3.0*, December 2007.

[58] J. Quittek, S. Niccolini, S. Tartarelli, M. Stiemerling, M. Brunner, and T. Ewald. Detecting SPIT Calls by Checking Human Communication Patterns. *Proceedings of the IEEE International Conference on Communications*, pages 1979–1984, June 2007.

[59] Lee Rainie. 64% of registered voters received robo-calls in the final two months of the 2006 election. *Pew Internet and American Life Project*, December 2006.

[60] Lee Rainie and Deborah Fallows. The CAN-SPAM Act has not helped most email users so far: A PIP Data Memo. *Pew Internet and American Life Project*, March 2004.

[61] Yacine Rebahi, Dorgham Sisalem, and Thomas Magedanz. SIP Spam Detection. *Proceedings of the International Conference on Digital Telecommunications*, pages 68–73, 2006.

[62] Messaging Anti-Abuse Working Group Release. MAAWG Issues First Global Email Spam Report. *http://www.maawg.org/news/maawg060308*, March 2006.

[63] J. Rosenberg. Applying Loose Routing to Session Initiation Protocol (SIP) User Agents (UA) - draft-rosenberg-sip-ua-loose-route-01. *IETF Internet Draft (work in progress)*, June 2007.

[64] J. Rosenberg. The Extensible Markup Language (XML) Configuration Access Protocol (XCAP). *IETF RFC 4825*, May 2007.

[65] J. Rosenberg and C. Jennings. The Session Initiation Protocol (SIP) and Spam - draft-ietf-sipping-spam-05. *IETF Internet Draft (work in progress)*, July 2007.

[66] J. Rosenberg and C. Jennings. The Session Initiation Protocol (SIP) and Spam. *IETF RFC 5039*, January 2008.

[67] J. Rosenberg, C. Jennings, and J. Peterson. The Session Initiation Protocol (SIP) and Spam - draft-ietf-sipping-spam-00. *IETF Internet Draft (work in progress)*, February 2005.

[68] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. SIP: Session Initiation Protocol. *IETF RFC 3261*, June 2002.

[69] H. Rust and S. Golombok. Modern Psychometrics: the Science of Psychological Assessment. *London: Routledge*, 1999.

[70] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz. A Bayesian Approach to Filtering Junk E-Mail. *Proceedings of the AAAI'98 Workshop on Learning for Text Categorization*, 1998.

[71] Claudia Sarrocco. Spam in the Information Scoiety: Buidling Frameworks for International Cooperation. *World Summit on the Information Society (WSIS)*, 2006.

[72] H. Schulzrinne, H. Tschofenig, J. Morris, J. Cuellar, J. Polk, and J. Rosenberg. Common Policy: A Document Format for Expressing Privacy Preferences - RFC 4745. *IETF Request for Comments*, February 2007.

[73] Tara Seals. IMS Reality Check. *Xchange Magazine*, January 2007.

[74] 3GPP Technical Specification Group Services and System Aspects. Study Item: Protection against SMS and MMS spam. *http://www.3gpp.org/specs/WorkItem-info/WI–320026.htm*, 2006.

[75] Dongwook Shin, Jinyoung Ahn, and Choon Shim. Progressive Multi Gray-Leveling: A Voice Spam Protection Algorithm. *IEEE Network Magazine*, 20(5):18–24, September 2006.

[76] Eulynn Shiu and Amanda Lenhart. How Americans use Instant Messaging. *Pew Internet and American Life Project*, September 2004.

[77] N.J Smelser and P.B. Baltes. International Encyclopaedia of Social and Behavioural Science. *Oxford: Elsevier*, 2001.

[78] Bulk SMS. Pricing guide. *http://bulksms.2way.co.za/w/pricing.htm*, December 2007.

[79] R. Sparks. The Session Initiation Protocol (SIP) Refer Method. *IETF RFC 3515*, April 2003.

[80] David Streitfeld. Opening Pandora's In-Box. *Los Angeles Times*, May 2003.

[81] British Telecom. BT's 21st Century Network. *http://www.btplc.com/21CN/*, 2007.

[82] The Radicati Group, Inc. & Mirapoint, Inc. End-User Study on Email Hygiene. April 2005.

[83] Garth Theunissen. Vodacom MMSs: Spam or not? *http://www.fin24.co.za*, January 2007.

[84] V. Torvinen, J. Arkko, and M. Naslund. Hypertext Transfer Protocol (HTTP) Digest Authentication Using Authentication and Key Agreement (AKA) Version-2. *IETF RFC 4169*, December 2005.

[85] H. Tschofenig, H. Schulzrinne, D. Wing, J. Rosenberg, and D. Schwartz. A Framework to tackle Spam and Unwanted Communication for Internet Telephony - draft-tschofenig-sipping-framework-spit-reduction-01. *IETF Internet Draft (work in progress)*, July 2007.

[86] H. Tschofenig, D. Wing, H. Schulzrinne, T. Froment, and G. Dawirs. A Document Format for Expressing Authorization Policies to tackle Spam and Unwanted Communication for Internet Telephony - draft-tschofenig-sipping-spit-policy-01. *IETF Internet Draft (work in progress)*, July 2007.

[87] International Telecommunication Union. Definition of Next Generation Network. *http://www.itu.int/ITU-T/studygroups/com13/ngn2004/working_definition.html*, 2004.

[88] John G. Waclawsky. IMS: A Critique of the Grand Plan. *Business Communications Review*, pages 54–58, October 2005.

[89] D. Waiting and R. Good. The UCT IMS Client. *http://uctimsclient.berlios.de*.

[90] D. Waiting, R. Good, R. Spiers, and N. Ventura. Open Source Development Tools for IMS Research. *4th International Conference on Testbeds and Research Infrastructures for the Development of Networks & Communities (TRIDENTCOM)*, March 2008.

[91] D. Waiting and N. Ventura. A Multilayered Architecture for Preventing Automated Spam in the IP Multimedia Subsystem. *50th IEEE Global Communications Conference (Globecom), Washington DC, USA*, November 2007.

[92] D. Waiting and N. Ventura. The Threat of Unsolicited Sessions in the 3GPP IP Multimedia Subsystem. *IEEE Communications Magazine*, 45(7):100–106, July 2007.

[93] D. Wing, S. Niccolini, H. Tschofenig, and M. Stiemerling. Spam Score for SIP - draft-wing-sipping-spam-score-00. *IETF Internet Draft (work in progress)*, August 2007.

[94] Karey Wutkowski. SEC cracks down on spam-driven stocks. *Reuters*, March 2007.

[95] Ho Khee Yoke and Lawrence Tan. Curbing Spam via Technical Measures: An Overview. *ITU World Summit on the Information Society (WSIS)*, 2006.

[96] Jonathan A. Zdziarski. *Ending Spam: Bayesian Content Filtering and the Art of Statistical Language*. No Starch Press, Inc, 2005.

[97] Tom Zeller. Law Barring Junk E-Mail Allows a Flood Instead. *The New York Times*, February 2005.

Appendix A

Survey Documents

This appendix holds the pilot survey that was conducted and the revised final survey. The style of the documents have been formated for brevity.

A.1 Pilot Survey

Dear Participant,

This survey aims to investigate telecommunication consumers' perceptions of multimedia spam (Junk Mail). The investigation forms part of a PhD dissertation at the University of Cape Town. Your participation is entirely voluntary. All data collected from you will be kept confidential. Please read and follow the instructions preceding the survey sections carefully.

Your participation is greatly appreciated.

Participant Age: _____

Participant Gender: _____

A.1.1 Section A

SECTION A INSTRUCTIONS:

Each statement below is followed by a series of possible responses. Please read each statement carefully and decide which choice best describes your experience. Please respond to every statement. If you are not completely sure which response is more accurate, put the response which you feel is most appropriate.

1. How often do you receive unsolicited email ("Junk Mail")?

- Daily
- Weekly
- Monthly
- Less often than monthly
- Never

2. How often do you receive unsolicited phone calls (e.g. Advertising/Sales Calls) on your cellphone?

- Daily
- Weekly
- Monthly
- Less often than monthly
- Never

3. How often do you receive unsolicited SMSs (e.g. Advertisements)?

- Daily
- Weekly
- Monthly
- Less often than monthly
- Never

4. How often do you receive unsolicited MMSs?

- Daily
- Weekly
- Monthly
- Less often than monthly

- Never

5. How many unsolicited phone calls would make you stop using your cellphone altogether?

- 10 per month
- 10 per week
- 10 per day
- I would not stop using my cellphone

6. How many unsolicited SMSs or MMSs would make you stop using your cellphone altogether?

- 10 per month
- 10 per week
- 10 per day
- I would not stop using my cellphone

A.1.2 Section B

SECTION B INSTRUCTIONS:

Each statement below is followed by a series of possible responses: Strongly Agree, Agree, Disagree, Strongly Disagree. Please read each statement carefully and decide which choice best describes your experience. Please respond to every statement. If you are not completely sure which response is more accurate, put the response which you feel is most appropriate. Do not spend too long on each statement. It is important that you answer each question as honestly as possible.

For each question answer one of the following:

- Strongly Agree
- Agree
- Disagree

- Strongly Disagree

7. The amount of unsolicited cellphone communication (calls / SMS / MMS) is getting worse.

8. I do not mind receiving unsolicited (i.e. "Junk"-) email.

9. Receiving unsolicited cellphone calls bothers me.

10. I feel bothered by unsolicited SMS'.

11. I do not mind receiving unsolicited MMS'.

12. I hesitate or do not answer cellphone calls from unidentified numbers.

13. Unsolicited calls and SMSs reduce the enjoyment I receive from using my cellphone.

14. I find unsolicited cellphone advertising informative.

15. Receiving unsolicited pornography / adult content or religious messages on my cellphone would offend me.

16. I would tolerate cellphone spam if I receive call / SMS credits, or airtime in return.

17. I am concerned with the ease at which access to my cellphone details is available to third parties.

18. I am sufficiently protected from being defrauded through unsolicited cellphone communication.

19. Cellphone-spam can trick the recipient into accepting hidden costs for these calls / SMS'.

20. Increased access-security to my phone details (i.e. number) is currently unnecessary.

A.1.3 Section C

SECTION C INSTRUCTIONS:

Each statement below is followed by a series of possible responses: Strongly Agree, Agree, Disagree, Strongly Disagree. Please read each statement carefully and decide which choice best describes your experience. Please respond to every statement. If you are not completely sure which response is more accurate, put the response which you feel is most appropriate. Do not spend too long on each statement. It is important that you answer each question as honestly as possible.

For each question answer one of the following:

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

21. My network provider should supply me with spam-blocking mechanisms (e.g. 'Junk Filter') for my phone.

22. The regulation of cellphone spam is not the responsibility of Government.

23. It is not the responsibility of network providers to regulate cellphone spam.

24. Managing the cellphone spam I receive is my own problem.

25. Government should pass laws to better protect cellphone users from unsolicited communication.

26. I would purchase spam-filter software for my cellphone.

A.2 Final Survey

Dear Participant,

This survey aims to investigate telecommunication consumers' perceptions of multimedia spam. The investigation forms part of a PhD dissertation at the University of Cape Town's Engineering Department. Your participation is entirely voluntary. All data collected from you will be kept confidential. Please read and follow the instructions preceding the survey sections carefully.

Your participation is greatly appreciated.

Participant Age: _____

Participant Gender: _____

A.2.1 Section A

SECTION A INSTRUCTIONS:

Each statement below is followed by a series of possible responses. Please read each statement carefully and decide which choice best describes your experience. Please respond to every statement. If you are not completely

sure which response is more accurate, put the response which you feel is most appropriate.

1. How often do you receive unsolicited email ("Junk Mail")?

- Daily
- Weekly
- Monthly
- Less often than monthly
- Never

2. How often do you receive unsolicited phone calls (e.g. Advertising/Sales Calls) on your cellphone?

- Daily
- Weekly
- Monthly
- Less often than monthly
- Never

3. How often do you receive unsolicited SMSs (e.g. Advertisements)?

- Daily
- Weekly
- Monthly
- Less often than monthly
- Never

4. How often do you receive unsolicited MMSs?

- Daily
- Weekly

- Monthly
- Less often than monthly
- Never

5. How many unsolicited phone calls would make you stop using your cellphone altogether?

- 10 per month
- 10 per week
- 10 per day
- I would not stop using my cellphone

6. How many unsolicited SMSs or MMSs would make you stop using your cellphone altogether?

- 10 per month
- 10 per week
- 10 per day
- I would not stop using my cellphone

A.2.2 Section B

SECTION B INSTRUCTIONS:

Each statement below is followed by a series of possible responses: Strongly Agree, Agree, Disagree, Strongly Disagree. Please read each statement carefully and decide which choice best describes your experience. Please respond to every statement. If you are not completely sure which response is more accurate, put the response which you feel is most appropriate. Do not spend too long on each statement. It is important that you answer each question as honestly as possible.

For each question answer one of the following:

- Strongly Agree

- Agree
- Disagree
- Strongly Disagree

7. The amount of unsolicited cellphone communication (calls / SMS) is getting worse.

8. I do not mind receiving unsolicited (i.e. "Junk"-) email.

9. Receiving unsolicited cellphone calls bothers me.

10. I feel bothered by unsolicited SMS'.

11. I do not mind receiving unsolicited MMS'.

12. Unsolicited calls and SMSs reduce the enjoyment I receive from using my cellphone.

13. I find unsolicited cellphone advertising informative.

14. I would tolerate cellphone spam if I receive call credits or airtime in return.

15. I am concerned with the ease at which access to my cellphone details is available to 'spammers'.

16. Cellphone spam can trick the recipient into accepting hidden costs for such calls / SMS'.

A.2.3 Section C

17. Who, in your opinion, is responsible for managing cellphone spam?

- My Service Provider (Vodacom, MTN, Cell C, etc.)
- Government (Legislation)
- It's my own responsibility

Appendix B

Robo-call Pre-recorded Message

I'm calling you today from CallRewards.co.za the instant rewards website. It is my great pleasure to inform you that your telephone number has been randomly selected by our computer.

For prize claims and more information about this exciting promotion please visit our website at www.callrewards.co.za. That's www.callrewards.co.za. Congratulations from us at the call rewards team. Please stay on the line to hear more about our terms and conditions.

Claims for prizes must be made in the manner and within the time specified on the Competition Notice. Failure to claim a prize within this time or in the manner specified may result in disqualification and selection of an alternate winner.

Employees of Call rewards or any company involved in the Competitions or any advertising agency or web company connected with Call rewards or any such persons subsidiary or associated companies, agents or members of their families or households, are not eligible to win prizes. Call rewards reserves the right to verify the eligibility of all prize-winners.

Call rewards assumes that by using its website and entering the Competition you have legal capacity to enter the competition and agree to the rules. If you are under the age of 18 your parents must consent to your entry of the competition and use of these rules.

Prize winners are chosen at random unless specified otherwise in the Competition Notice. You may not claim a prize if you have already claimed a prize from call rewards in the last three months. For the full list of terms and conditions please visit www.callrewards.co.za/terms .

Appendix C

Accompanying CD-ROM

This thesis comes with an accompanying CD-ROM containing the following items:

- Applications - The software tools used in the test-bed.
- Evaluation Results - Raw data files from the evaluations.
- Referenced Papers - A collection of literature referenced in the thesis.
- SIPp Scenarios - The SIPp XML configuration files and shell scripts used in the evaluations.
- Survey Results - Raw survey data.
- Thesis Documents - The individual chapters, the main thesis document and an abstract.

Appendix D

List of Abbreviations

3GPP - 3rd Generation Partnership Project

AAA - Authentication, Authorisation and Accounting

A-IMS - Advances to IMS

AKA - Authentication and Key Agreement

API - Application Programming Interface

AS - Application Server

ATM - Asynchronous Transfer Mode

ABNF - Augmented Backus–Naur Form

ARPAnet - Advanced Research Projects Agency Network

CAN-SPAM - Controlling the Assault of Non-Solicited Pornography and Marketing Act

CAMEL - Customised Applications for Mobile networks Enhanced Logic

CAN - Connectivity Access Network

CAPTCHA - Completely Automated Public Turing Test to Tell Computers and Humans Apart

CDMA - Code Division Multiple Access

CORBA - Common Object Request Broker Architecture

CPA - Call Pattern Analyser

CPU - Central Processing Unit

CSCF - Call / Session Control Function

CSV - Comma Separated Values

DHCP - Dynamic Host Configuration Protocol

DiffServ - Differentiated Services

DNS - Domain Name System
DoS - Denial of Service
EDGE - Enhanced Data Rates for GSM Evolution
ENUM - Telephone Number Mapping
ETSI - European Telecommunications Standards Institute
FQDN - Fully Qualified Domain Name
GGSN - Gateway GPRS Support Node
GNU - GNU's Not Unix
GPL - GNU General Public License
GPRS - General Packet Radio Service
GSM - Global System for Mobile communications
HTTP - Hypertext Transfer Protocol
HSDPA - High-Speed Down-link Packet Access
HSS - Home Subscriber Server
ICMP - Internet Control Message Protocol
IETF - Internet Engineering Task Force
IM - Instant Messaging
IMPI - IP Multimedia Private Identity
IMPU - IP Multimedia Public Identity
IMS - Internet Protocol Multimedia Subsystem
ISC - IMS Service Control
ISP - Internet Service Provider
ITU - International Telecommunication Union
JAIN - Java APIs for Integrated Networks
LGPL - GNU Lesser General Public License
MD5 - Message-Digest algorithm 5
MIME - Multipurpose Internet Mail Extensions
MMD - Multimedia Domain
MMS - Multimedia Messaging Service
MPLS - Multi-protocol Label Switching
MSRP - Message Session Relay Protocol
NAT - Network Address Translation
NGN - Next Generation Network
OMA - Open Mobile Alliance
OSA - Open Service Access
OSIMS - Open Source IMS core

OTC - Over-The-Counter

PCMCIA - Personal Computer Memory Card International Association

PDF - Portable Document Format *or* Probability Density Function

POTS - Plain Old Telephone Service

PSTN - Public Switched Telephone Network

QoS - Quality of Service

RFC - Request For Comments

RTCP - Real Time Control Protocol

RTP - Real-time Transport Protocol

RTT - Round Trip Time

SDP - Session Description Protocol *or* Service Delivery Platform

SER - SIP Express Router

SGSN - Serving GPRS Support Node

SHA1 - Secure Hash Algorithm 1

SIP - Session Initiation Protocol

SMS - Short Message Service

SOA - Service-Oriented Architecture

SPIM - Spam over Instant Messaging

SPIT - Spam over Internet Telephony

SQL - Structured Query Language

SS7 - Signalling System 7

TCP - Transmission Control Protocol

TISPAN - Telecoms and Internet converged Services and Protocols for Advanced Networks

TLS - Transport Layer Security

TS - Technical Specification

UAC - User Agent Client

UAS - User Agent Server

UCE - Unsolicited Commercial Email

UCT - University of Cape Town

UDP - User Datagram Protocol

UE - User Equipment

UML - Unified Modelling Language

URI - Uniform Resource Identifier

URL - Uniform Resource Locator

VoIP - Voice over Internet Protocol

WLAN - Wireless Local Area Network
WSDL - Web Service Definition Language
XCAP - XML Configuration Access Protocol (XCAP)
XML - Extensible Markup Language