1-1-2003

Non-business use of the World Wide Web : A study of selected Western Australian organisations

Craig Valli
Edith Cowan University

USE OF THESIS

Non-Business Use of The World Wide Web - A Study of Selected Western Australian Organisations

by

Craig Valli Dip.Teach, B.Ed, M.Manag.Inf.Sys

A thesis submitted in partial fulfillment of the Requirements for the Award of Doctor of Information Technology

At the Faculty of Computing, Health and Science, Edith Cowan University, Mount Lawley Campus

Submitted 22nd October, 2003

ABSTRACT

Employees undertake a wide range of activities when they use the World Wide Web in the work place. Some of these activities may leave the modern Internet connected organisation vulnerable to undue or unknown risk, potential productivity losses and expense as a result of misuse or abuse of the Internet provision. Much of the existing literature on this subject points to a purported epidemic of misuse in the workplace.

If this practice is so prevalent and widespread, what can modern Internet connected organisations do to identify the abuse and reduce the risks and losses that these abuses represent? To what extent is the World Wide Web used by employees for non-business related activities in organisations and can filtering or organisational policies impact on this activity?

This research specifically examines contextually, the level of misuse with respect to the use of the World Wide Web in three selected Western Australian organisations using multiple interpretive case study as the vehicle for the study. The research is significant internationally to all organisations that use Internet in their everyday work.

The research has discovered anomalous behaviour on the part of non-business users who have employed a variety of techniques and tactics to mask their activities. Also, organisational management in the cases examined had demonstrated shortfalls in their perception of misuse within their organisations and, the implementation of effective policy.

Keywords: *Internet, World Wide Web, misuse, abuse, employee*

DECLARATION

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

Signature [redacted]

Date 22nd October, 2003

ACKNOWLEDGEMENTS

I would like to acknowledge with particular gratitude the assistance and encouragement of my principal supervisor, Associate Professor William Hutchinson. I am also indebted to my second supervisor Dr. Thomas O'Neill for his assistance in completing the thesis.

Furthermore, I would also like to thank all of my colleagues and friends who have encouraged me or helped me in anyway with completion of this thesis.

Finally, I would like to thank my family that has provided unconditional support in this endeavour. In particular, I would like to thank my wife Robyn for her patience, understanding and belief during the long journey that this has been.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

GLOSSARY OF TERMS

- **AVI** - Audio Video Interleave, the file format for Microsoft's Video for Windows standard.
- **Encryption** - The translation of data into a secret code which is unreadable. Encryption is the most effective way to achieve data security. To read an encrypted file, you must have access to a secret key or password that enables you to decrypt it. Unencrypted data is called plain text ; encrypted data is referred to as cipher text.
- **MP3** – Is a common abbreviated form of MPEG, audio layer 3. Layer 3 is one of three coding schemes for the compression of audio signals. Because MP3 files are highly compressed, they can easily be transferred across the Internet.
- **Proxy server** – a server that sits between a client application or device and a real server intercepting all requests to the real server. The proxy server sees if it is able to process the request itself before forwarding it onto real servers.
- **SSL (Secure Sockets Layer)** - A session layer protocol that provides authentication and confidentiality to applications via the use of encryption technologies.
- **warez** – commercial software that has been pirated and distributed illegally.
- **WAV** – A format for storing sound files developed by IBM and Microsoft

CHAPTER 1 - INTRODUCTION

1.1 Overview

This thesis will examine the level of non-business use of the World Wide Web in selected Western Australian organisations. The thesis will examine data from a range of sources within the selected organisations to obtain a rich and balanced view of non-business usage of the World Wide Web.

There is much hype currently in the media and computer industry press with respect to non-business or abusive use of the Internet, and in particular the World Wide Web, in the workplace. The claims made in these reports are most disturbing and alarming when taken within the context of loss, risk and liability to most modern Internet connected businesses:

- "Almost 25 percent of employees' online time is given over to surfing the Internet for recreational purposes." (Ohlson, 1998)
- "Nearly two-thirds of companies report that employees access sexually explicit Web sites at work" (Hickins, 1999)
- "...online trading during working hours is costing corporate America $11 billion annually" (Ballard, 2000)
- "According to Nielsen/Netratings, 21.1 percent of adult sites are accessed at work. Viewing time is an average of four minutes per person" (Anonymous, 2001)
- "Xerox Corp. said it has fired about 40 employees this year in the U.S. for "excessive misuse of the Internet." The main problem was

> among employees who, while at work, viewed pornographic Web sites" (WSJ, 1999)

As can be seen by the previous quoted literature relating to the level of Internet misuse and abuse the actual level reported is widely varied. Furthermore, the range of activities that misuse or abuse can be includes but is not limited to viewing pornography, accessing share trading sites, recreational surfing, on-line banking, accessing racist or hate based sites and downloading of copyright materials such as MP3.

Some of these Internet misuse reports are levelled at a lack of compliance with corporate ethos or policy such as the cited Xerox example (WSJ, 1999). Yet, other reports are trying to provide quantified analysis in terms of lost productivity and actual Internet bandwidth consumption as result of non-business use or misuse by end-users (Ballard 2000, Ohlson 1998). These reports, if anything, highlight the difficulty of measuring and defining non-business usage within organisations. The usage of the World Wide Web in organisations is a complex, socially constructed phenomenon that occurs within the context of an organisational structure and supporting systems.

The literature relating to Internet misuse is often based on limited survey and blatantly commercial bias (Holtz, 2001). The literature does not contain any empirical evidence on which to base any assumptions or findings made by the reports. The current literature and information on Internet abuse and monitoring suffers from similar problems identified by (George, 1996) with regards to computer monitoring:

> "The advocacy literature devoted to computer-based monitoring is marked by three criteria: (1) its foundation on small (and usually biased) sample sizes; (2) the monolithic character of its findings; and (3) the consistency of its conclusions." (George, 1996)

There is sometimes no real attempt to measure or accurately rate the level of non-business use that does occur within an organisation. Many of the reports do not contain any empirical evidence taken from system log files for example to substantiate claims made by the authors or participants of the survey. Furthermore, most statistics provided are founded on the basis of survey of one stakeholder group, typically the organisation's management. The approach taken by some surveys has management estimating levels of misuse frequently with no sound empirical basis on which to base such an assessment. Further bias could come into the survey sample as a result of the highly contentious nature of the content being assessed. This situation could cause participants to manipulate directly responses due to the vested interests of the participant. One view would be to inflate the levels of misuse by a manager to gain access to more resources for managing the "problem". The second view is in reducing estimation of the problem to promote the perception that the situation is under control.

Not all non-business usage may be counterproductive or destructive to the corporate good. It may in fact provide some tangible and intangible benefits to the organisation as a result of employees using the World Wide Web for non-business related activities. Take the example of an employee who, as a result of non-business use of the World Wide Web to search for unauthorised material, has acquired and honed his/her search skills to a higher level of proficiency than a compliant user. When this employee now performs search related tasks they are more efficient and effective as a result of the non-business use. Another example of a benefit from using the World Wide Web is the use of on-line services for banking, paying bills or other personal errands that would normally be undertaken during a lunch break. By using the World Wide Web in this manner an employee rather than having spent a break period in a line waiting to pay a bill can instead relax and de-stress, making for happier employees.

The World Wide Web may be simply replacing other more conventional work shirking behaviours such as long lunch hours or frequent toilet breaks. This substitution could be because accessing the World Wide Web is easy, provides immediate, almost anonymous access and subsequent gratification. Some employees may see the use of the Internet as a means of redressing perceived injustices such as longer working days and unpaid overtime. Non-business use may be a method of redressing the increasing invasion of the working day into employees' personal lives as a result of modern work practices or technologies such as mobile phones.

If this non-business usage or misuse of the World Wide Web is so prevalent and widespread, what can modern connected organisations do to identify the abuse, and reduce the risks and losses inherent in these activities? To what extent is the World Wide Web being used by employees for non-business related activities in quantifiable and measured manner? Do commonly used countermeasures such as content filtering or organisational policies impact on the non-business usage of the World Wide Web by employees?

1.2 Significance of the Research

The research specifically examines the level of non-business usage that occurs with respect to the use of the World Wide Web in the work context from a variety of viewpoints and analysis. The research uses a case based approach and involve the extensive analysis of several data within the selected case organisations. The research is significant to all organisations that use the Internet and, in particular, the World Wide Web in their everyday work. The extensive analysis of log files provides empirical evidence for usage patterns and behaviours of employees.

Non-business activities that employees or users undertake whilst using the World Wide Web may leave the organisation open to undue risk or liability. Some examples are the accessing of pornography, creating a hostile workplace, or accessing copyrighted materials such as software. In light of the computer industry and media reports on non-business usage, is pornography the pariah that some reports say it is in the Internet connected workplace? Are there other risks that are not apparent or are being overlooked by these reports?

Organisations that have a connection to the World Wide Web have to manage responsibly their usage from an ethical and commercial viewpoint (Block, 2001). This research is significant in that it will examine management perceptions of what is occurring within an organisation and compare these perceptions with the empirical results of the log file analysis.

The analysis of the log files allows for the examination of what are the actual direct costs associated with the provision of World Wide Web services to employees who are undertaking non-business activities. What indirect costs, such as productivity losses, are due to employees accessing non-business related material and causing network latency for instance? Is the direct loss of productivity due to non-business usage outweighed by gains in other areas such as improved skill levels, or well being and happiness of employees?

The effectiveness of countermeasures to combat non-business usage is an area of significance and one this thesis will examine. There are a variety of countermeasures that are pertinent and applicable to preventing the inappropriate usage of the World Wide Web by end-users. There are simple measures such as monitoring of network usage through the use of device (router) or program based (proxy server) activity logging mechanisms. These log files are then processed by programs that produce a variety of executive reports for analysis and interpretation by organisational stakeholders. Content filtering is an approach that is often promoted by companies who conduct some of the surveys in the literature as a panacea to misuse ills. Organisational policy is another mechanism by which organisations attempt to control or modify behaviour of end-users. Each of these countermeasures' effects on an organisation goes beyond the simple technical implementation of them.

Do these countermeasures such as active monitoring, content filtering or policies have an effect on employee behaviour when using the World Wide Web? The actual outcomes produced as a result of countermeasure deployment may not only be that of the intended effect for instance compliance, but also produce other effects such as users becoming stressed, distrustful of management or masking problem behaviour. The research helps uncover effective practices and undetected deficiencies relating to the deployment of these common countermeasures to non-business usage of the World Wide Web.

1.3 Research Questions

The aim of the research is to provide research into the extent of non-business usage of the World Wide Web in organisations that is based on empirical, quantitative and qualitative data. The research questions that analysed in this thesis relating to non-business use in the workplace are:

1. To what extent is the World Wide Web service used by employees for non-business related activities in selected organisations?
2. What is the perceived versus measured reality of non-business use in an organisation?
3. Does the enforcement of countermeasures, such as policy, affect the behaviours of users within an organisation?

In Chapter 2 an examination of the body of literature relating to the thesis is undertaken. Chapter 3 deals with the selection of an appropriate research methodology in this thesis which is multiple interpretive case studies. Chapter 4 details the appropriate data and methods for analysis of this data and case study design. Chapters 5, 6 and 7 each contain an individual case study as part of the conducted research. Chapter 8 considers the implications of the findings of the three case studies. Chapter 9 contains concluding remarks and identifies further research.

CHAPTER 2 - LITERATURE REVIEW

2.1 Defining Non-Business Usage or Misuse

The term non-business use is not definitive as highlighted by Lim (2002). Terms such as “misuse”, “abuse”, “cyber loafing”, “cyber slacking”, “cyber slouching” are used to describe the inappropriate use of the Internet within the workplace context. For the purpose of this study the term non-business use will be adapted from articles by various authors (Anandarajan, 2002; Block, 2001; Greengard, 2000; Lim et al., 2002; Mills, 2001; Roberts, 2000; Detmar W. Straub & Nance, 1990; Zaiton, 2000) on misuse and non-business use of computers and the Internet.

Hereafter, non-business use for this study will mean any activity conducted whilst using the organisational World Wide Web connection that is not directly related to the fulfillment of a person’s job description, or is not done in the course of processing a workflow related task, or is contrary to established organisational policy or relevant Australian laws such as State and Federal Criminal Codes, Copyright Law, Industrial Relations Law and the Equal Opportunity Act.

The use of the above definition is too rigid and problematic for use in most organisations; what one organisation would consider non-business or unacceptable use, another organisation might consider acceptable. An illustration of this is the use of on-line banking where one organisation bans all types of access to financial institutions, another allows staff to undertake personal bill pay and banking transactions, as extended working hours by staff are required.

There is material that is forbidden by law, human rights and industrial relations acts such as the viewing of pornographic, racist, sexist, hate or copyright-based materials. Such viewing is a known and obvious misuse of an organisation’s Internet connection by employees. However, for a vast majority of remaining

Internet traffic, whether it is deemed appropriate or a misuse depends largely on the organisational context and policy framework within which it occurred. Thus measurement of Internet misuse by blanket short term surveys is questionable.

2.2 Current Industry Surveys

The non-business use of the Internet in a 1000 seat organisation in the United States of America is estimated to cost that organisation as much as $US35 million in lost revenues and wages (Holtz, 2001). Recent surveys have figures quoting between 60 to 80% of employees accessing pornographic material in the workplace (Hickins, 1999). Other surveys point to staff sending private e-mails and conducting personal business (Holtz, 2001). A January 2000 study by the Saratoga Institute (U.S.) found that nearly two-thirds of U.S. firms have disciplined employees for Internet abuse, and that 56% admitted they knew employees used the Internet to gamble, to look at pornography, and to engage in other activities that were not work-related (SIA, 2000). A 1999 study by the American Management Association reported that more than 50% of all Internet activity taking place within companies was not business-related (Greengard, 2000).

Most of the existing surveys appear to be produced with a management bias in terms of either participation or the focus of the actual survey. As Holtz (2001) points out, many of these existing surveys are dubious or problematic at best. Review of this literature indicated vendors, who have vested interests in the sale of filtering software or other countermeasures to organisations, also sponsor many of these surveys or reports. Due to such bias much of the literature reporting misuse should be viewed with scepticism and this bias was one of the principal motivations for this thesis. However, from these reports it might be surmised that the management perception is that they believe that Internet connections are being used primarily to download pornography, to gamble and to engage in non-business activities.

2.3 Why do people misuse?

Excepting those instances where an organisation or individual may be held liable for damages from such accesses, the responses from organisations and research conducted in the literature varies as to whether non-business use results in lost productivity or has any detrimental effect. Not all non-business access is construed as a bad event, threat or risk. Some organisations see it as a means of educating their user base to help increase knowledge acquisition skills or aiding in the adoption of the technology itself.

> "Complicating things further is the fact that personal use of the Internet might actually provide a positive benefit. An August study conducted by Xylo, Inc., a work/life programs provider based in Bellevue, Washington, found that 56 percent of employees who use the Internet for personal reasons report that it helps them do their jobs better or simply makes them happier or less stressed-out employees. About 43 percent claim that the Internet has no real impact -- positive or negative -- on their performance." (Greengard, 2000, p.22)

A thesis central to a paper by Belanger & Slyke (2002) is that adults have better learning experiences through a meaningful, experiential, self-directed approach. The author goes on to say that the wholesale condemnation of web browsing or application play, may be short sighted and creating corporate disadvantage by hampering employee IT skills acquisition.

"Constructive Recreation" is a term used by Oravec (2002, p.61) to describe activity by an employee that is synchronous with their impending work tasks. This recreation purportedly allows the employee to equip themselves with skills they can utilise in the future. This skill acquisition is presumed to be done within the confines of legal, technical and organisational limits. Other values of constructive recreation are that it supports intellectual and psychological stimulation, thus providing the ability for workers to take on greater challenges. By engaging in bursts of constructive recreation, workers can purportedly bring back balance to a stressful or hostile environment using the diversion as a release or balance mechanism. However, Oravec(*ibid*) does point out that the notion of play or constructive recreation within an adult context has not been the subject of thorough investigation or research. Later Oravec's paper argues that this constructive recreation is the virtual form of what we refer to in the current face-to-face physical world as work shirking or, colloquially, as "slacking off".

There is little argument that the working week has increased in most developed nations. More and more people are working longer, irregular hours and, often use electronic networks of one form or another to sustain these efforts. This increased communication from the workplace has become more invasive for employees and now often extends into their personal lives. Consequently, many people now have to conduct errands, pay bills and arrange other non-business related activities during these extended business hours.

Some people may also see the non-business use of the World Wide Web as a means for redressing some perceived injustice regarding remuneration or, poor treatment, by the organisation they work for (Greenberg, 1990; Lim, 2002). The paper by Lim et al. (2002) cites examples of employees expressing these sentiments:

> "It is all right for me to use the Internet for nonwork reasons at work. After all, I work overtime without receiving extra pay from my employer." (*ibid*, p.69)
>
> "I don't see anything wrong with using the company's Internet access for nonwork purposes as long as I do not do it too often and complete my work as required by my boss." (*ibid,* p.69)

Other findings showed that the majority of respondents in the study saw it as acceptable practice to use the Internet for non-work related activities if they put in extra effort or worked unpaid overtime. The employees saw the use of this resource as appropriate informal compensation for their perceived extra efforts in the work place. The authors postulated that there is transference of this activity to the Internet from more traditional forms of loafing, which concurs with Oravec (2002). One possible reason for this behaviour was that the Internet was attractive as it has the ability to provide instant gratification to the employee.

2.4 Content Filtering

Increasingly, organisations are deploying content filtering systems to counteract some of the threats and risks that non-business and inappropriate use of the Internet poses. The effectiveness of filtering systems alone is, in some cases, proving ineffectual to preventing this sort of abuse (Hunter, 2000; Neumann & Weinstein, 1999; Nunberg, 2001).

A study conducted by Hunter (2000) of popular commercial Internet filtering software investigated how under-inclusive (failing to block the worst pornography) and/or over-inclusive (blocking non-sexual, non-violent content) the filtering software was. According to (Hunter, 2000) "Overall, filters failed to block objectionable content 25% of the time; on the other hand, they improperly blocked 21% of benign content". This finding brings into question the claim by some vendors that their filtering software will block 90-95% of undesirable content from the World Wide Web.

It is also questionable whether these tools in their present form and mode of operation, can effectively work as advertised. The top search engines on the web can only reliably map 15-20% of all known web content (Lawrence & Giles, 1999). Furthermore this percentage of coverage and overall efficiency drops as the Internet continues to grow. "Search engine coverage relative to the estimated size of the publicly indexable web has decreased substantially since December 97, with no engine indexing more than about 16% of the estimated size of the publicly indexable web" (*ibid,* p.107). The figures presented by the authors raise the question what quantum leap of technology do these filtering and categorization tools possess that are not in existing search and cataloging technologies already used to map the World Wide Web?

The World Wide Web is a dynamic, chaotic environment where websites and locations of material are in a constant state of flux. Some of the material that is problematic such as pornography or "warez" – (illegal copies of software) are known to be housed on sites that will change location every 24 hours in order to evade detection and/or prosecution under US copyright law (McCandless, 1997). It is known that habitual or criminal users will also frequent and participate in their own online communities or covert channels to exchange information relevant to procurement of materials. This level of activity has been highlighted in many cases involving serious crime such as child pornography and copyright infringements. The perpetrators are using Internet technologies such as chat and file sharing tools to exchange files or information on sites and materials. Many of these non-web browser based applications and activities can be successfully tunneled through an organisational proxy server without the administrator being aware that it is occurring.

Content filtering systems typically work by denying access based on a list of keywords, simple semantic heuristic or rule set. Words may often be used in a different semantic context, which entirely changes the meaning of those words from offensive to non-offensive, business to non-business. For example the word “breast” could be used to search for “chicken breast”, “Robin red breast” or “breast cancer” all of which would be a legitimate use of the word breast for various organisations or individuals. A second example is if the word “sex” is blocked. Thus, the sequence of letters “sex” is blocked no matter where it occurs in the word so the words “Middlesex” and “sexing” would be blocked. A legitimate HTML page could have the innocuous sentence “Matthew was sexing day old chickens in Middlesex in the summer of 2002”. Thus, that page would be banned or blocked as a pornographic site by content filtering systems that are reliant on keyword rule sets.

One of the other problems with content filtering systems is that they possibly promote an impenetrable fortress mentality. For example, if we have content filtering systems deployed, then they will block all misuse, therefore, we do not have to worry. The software vendors, who claim to have developed a replacement for adequate supervision of an Internet connection via the use of their software solution, could help to perpetuate this belief. This style of behaviour by organisations and individuals is similar to the magic bullet syndrome, as described by (Markus & Benjamin, 1997) in their article on IT-enabled transformation, when it is believed that the acquisition of the filtering software will magically transform an organisation from a cesspit of Internet iniquity into a paragon of morality. The reality is that these programs do not ostensibly achieve the claims that they make.

The Australian government in late 1999 passed amendments to the Censorship and Communications Acts. On 1 January 2000, the Australian Broadcasting Services (Online Services) Amendment Act came into law with the specific purpose of addressing the nature and accessibility of online content on Australian sites. This amended Act has made it compulsory for ISPs, for instance, to block or remove material classified as R (material requiring an adult perspective), X (sexually explicit material) or RC (refused classification) from Australian based sites when directed to do so by the Australian Broadcasting Authority (ABA). The system will only process or examine content that has been formally subjected to the ABA's complaint system.

It was clear that by 2001, this Internet content filtering attempt by the government, was largely not working (Brown, 1999; McAuliffe, 2001; Taylor, 2001). The government in particular, is accused of inaccurately portraying the number of complaints in a report published by the ABA in April 2001. Out of the 491 complaints received in the Year 2000 by the ABA only 22 Australian sites were found to have contained material requiring removal by ISPs. The remaining sites were housed offshore and consequently, did not come under Australian jurisdiction.

Akdeniz(1998) mentions that one of the shortcomings of content filtering is that "Each system is extremely subjective and affected by cultural assumptions..." (Akdeniz, 1998). Even within western societies, there is divergence in thinking as to what is acceptable/unacceptable content due to varied cultural, religious and political viewpoints. To illustrate this point, would a system developed by liberalist atheists be markedly different from one produced by conformist, fundamentalist theists? Thus cultural, religious and political biases must surely be realised in the development of content filtering systems by vendors.

Furthermore, Akdeniz(1998) highlighted the fact that many of these systems are targeted at protecting children and, as a result, treated adults as children. This underlying premise in the construction of content filtering systems presents a problem for organisations and adults. Take the case of when adults may be legitimately trying to find material that is necessary to process a workflow, but is classified as objectionable for a child and, therefore, is blocked by the system. The use of content filtering technologies as a quick fix or silver bullet to reduce Internet misuses may in fact prove counter productive within an adult workplace context. The rectification of this problem may also take considerable technical and human resource to achieve, hence impacting any benefit from using the system in the first place.

The paper by Akdeniz states "Government-imposed censorship, over-regulation, or service provider liability will do nothing to keep people from obtaining material..." Akdeniz(1998, p.42). This statement made in 1998, could almost be considered prophetic and was largely proven to be an accurate assessment. This has been exemplified by the failure of Australian governments, both State and Federal, attempts to regulate content via legislation (Brown, 1999; McAuliffe, 2001; Taylor, 2001).

2.5 Monitoring of Employees

There is concern that overt, oppressive regimes of monitoring do in fact reduce employee productivity. Previous studies in the general area of electronic monitoring of employees would concur with this concern (George, 1996; Phipps, 1996). Some of the more severe cases cited in the existing literature base, points to employee health suffering as a result of the stress caused by monitoring of activity. In some cases examined, this suffering included the manifestation of physical illness due to increased stress that the monitoring caused.

A study by Chalykoff & Kochan (1989) provided a framework for studying the effects of monitoring on employees which Urbaczewski & Jessup (2002) adapted and applied to a modern work context including Internet usage. This latter study states that monitoring has two main purposes, one is for providing feedback, the other is for monitoring for control. Feedback is simply a mechanism to improve performance about jobs and tasks performed by employees. The main motivation of feedback is to result in improved performance or outcomes for the individual being monitored and the organisation. This surveillance of course can have a negative effect by raising stress levels in employees. At a subliminal level the production of behaviours as a result of Hawthornian effects (Mayo, 1933) or Pavlovian operant conditioning, would produce artificial performance due to the activity of the monitoring: that is, the employees being monitored will perform as specified whilst being watched.

Monitoring for control is when employees are checked for compliance against organisational policy. Most of the content filtering software in existence is based upon an attitude of compliance monitoring. As pointed out in paper by Chalykoff & Kochan (1989) and others within the literature, this type of monitoring tends to be a negative experience. Monitoring that is used in this manner may impact on trust relationships, employee well-being and overall organisational performance.

One of the implications of this study is that electronic monitoring is antagonistically dichotomous and may force an organisation to make a decision between productivity and employee satisfaction. Another implication is that positive forms of monitoring are more instructive than destructive to employees and their sense of well being and contentment in the workplace.

2.6 Policy and Points of Law

Policy is touted as one of the keystones of building a system that has low non-business use. However, even though implemented, policy is often not enforced and, consequently, its efficacy is reduced. A study on a code of ethics for computer use by (Harrington, 1996) notes that they have some effect on user behaviours. This study found the effectiveness of policy is increased if management actively enforces it, which is substantiated by (Straub, D.W et al., 1993). If there is overt, deliberate publicising of policy breaches and of what actions were taken against transgressors, then the effectiveness of policy to reduce the misuse of computer systems increases.

Policy is intended as both a mechanism to protect an organisation from liability and legal issues, and a conduit for the effective and efficient dissemination of information to stakeholders. Policy formulation is done within the socio-political structure of the organisation. Policy should outline the rights and responsibilities for each stakeholder with the context of the organisation and its legal environment. From various cases and papers (Foster, 2001; Kallman, 1993; Lichtenstein & Swatman, 1997; Mirchandani & Motwani, 2002; Solicitor, 2001), policy should be explicit and easily understood by all stakeholders. The following sections will examine pertinent case law as it relates to Internet misuse within organisations with an Australian context.

2.6.1 Case Law: An Australian Context

The Australian Industrial Relations Commission (AIRC) has long held the view that companies and organisations have a right to protect their assets from inappropriate use. It also has held the view that employers should be able to take whatever legal and reasonable steps necessary to protect themselves. Many of the cases handed down by the AIRC that dealt with Internet misuse had not progressed beyond this level within the Australian Legal Framework until recently.

One of the key legal cases presided over by Justice Merkel in the Australian Federal Court was the case of Ansett versus Australian Municipal Administrative, Clerical and Services Union (AMACSU) on the 7th April 2000. It set precedent case law specifically relating to steps that an organisation must take, should it wish to terminate an employee for unauthorised use of e-mail or the Internet (Cooper & Martin, 2000). The decision does not disallow employers from engaging in monitoring activities, nor does it prevent them from terminating an employee should he/she transgress well delineated and explained policy. What the decision does is to reinforce previous findings by the AIRC that upheld the view that an employee's transgression of an organisational access policy was sufficient grounds for termination of employment.

In 2001, the Australian Government Solicitor (AGS) released Briefing 58 entitled Misconduct in E-mail and Internet Use at Work. This briefing outlines some of the risks that inappropriate usage of the Internet left an organisation vulnerable to, including claims that could be made by staff viewing and distributing objectionable material and, subsequent productivity losses. The document also outlines some forms of sanctions that could be imposed upon employees for misconduct, termination and the application of Anti-Discrimination legislation.

Concrete cases of misuse are highlighted with examples taken from Federal and State Case law or AIRC rulings. The first example highlighted is that of pornography and the briefing states, "Simply accessing pornography may not be enough to justify dismissal, although it may be where the employer has a clear policy and directions warning employees of the prospect of dismissal" (Solicitor, 2001). The highlighted cases illustrate mainly AIRC rulings with regards to unsuccessful reinstatement attempts by employees.

The requirements for adequate collection of evidence is outlined and is particularly relevant where there are breaches of a serious nature that require sanctions by an organisation or further investigation by authorities. The briefing also touches on privacy issues and directs the reader to The Privacy Commissioner's recently released Guidelines (FPC, 2002).

In summary, the briefing outlined 12 points that organisations should take to ensure that they meet the requirements of good management practice and any subsequent legal requirements, these are:

1. Create a clear internet and e-mail policy and issue directions concerning compliance
2. Distribute the policy and directions to all staff and ensure that they are understood
3. Identify with specificity the type of personal access allowed, both as to the time, duration and sites permitted – statements such as for 'work-related purposes' may be too general
4. Explain to staff how misuse impacts on productivity
5. Inform staff that e-mail and internet use is not secure or private even if it is deleted and explain how it is being monitored
6. Explain how misuse may impact on the legal liability of the organisation and the reputation of the organisation
7. Explain enforcement methods, that is, if all e-mail and internet use by all employees will be audited at regular intervals, whether it will be by random checks, or whether it will be only when a complaint is received
8. Warn staff about the consequences of contravention and the possibility of termination of employment
9. Gather evidence of possible breaches carefully
10. Administer the policy and directions in a consistent way and update the documents when necessary
11. Explain the obligations under anti-discrimination legislation giving examples
12. Ensure staff attend appropriate training and keep records of who attends

(AGS, 2001)

It is important to note in issues covered by the AGS and from cases that have come before the AIRC, most of these indicate that simply stating a policy initiative or direction is not enough. What is important is that the employee or intended audience is well aware of what the actual policy direction is and what the consequences are for a transgression of the directive. Furthermore, information, education and subsequent review of the policy must also be an on-going process conducted by the employer. Within this overall process of policy enforcement, you must also be able to show due diligence and proper auditing or monitoring of these activities. These directions from the AGS and AIRC concur with (Wood & Lo, 2001) who state that organisations who have a policy and who do not update policy and or review employees knowledge of same is not necessarily protected from any liability issues.

It should be pointed out that viewing of some material, while not directly illegal under Australian Federal and State Criminal Laws, may be in breach of Equal Opportunity, Anti-Discrimination Statutes and Acts. The viewing of certain types of pornography for instance, partial nudity by adults is legal within Australia in most criminal statutes. This legality does not mean, however, that material is not offensive to people of certain religious or philosophical followings and, hence, is then a potential issue under the relevant Anti-Discrimination or Equal Opportunity Acts. These factors are important and complex and organisations must take them into account when constructing policy regarding the usage of the Internet and, in particular, the World Wide Web.

2.7 World Wide Web Proxy Log File Formats

There is a multiplicity of proxy servers available in the marketplace and many develop proprietary methods of producing log files for the programs. There are standards as prescribed by the W3C and de-facto standards that have emerged as a result of the development and popularity of certain World Wide Web proxy servers.

The variance in information that log file formats embed into a log file is marked. Basic logging systems or formats will use a timestamp, URL and delivery status model with little information being stored in the log file itself. More verbose log file models may embed latency measures to enable calculation of relative performance and other metrics for detailed analysis by computer professionals. The following sections review the common existing log file formats and their possible shortcomings, with extensive descriptions of the particular logging formats to be found in Appendix A.

2.7.1 W3C Common Log File (CLF) Format

Most, if not all World Wide Web proxy servers have the ability to produce the W3C Common Log File (CLF) formatted log files (Luotonen, 1995) either natively or via the use of a conversion script or program. This format has become the de-facto industry standard for tools analysing World Wide Web usage. A CLF log entry has a simple tuple that has a structure that is defined as

remotehost rfc931 authuser [date] "request" status bytes

Some example CLF lines follow

```
10.101.45.177 – jerry [02/Jan/2002:09:26:23 +0800] "GET
http://musiccity.streamcastnetworks.com/desktop.htm HTTP/1.0" 200 916

10.101.45.177 – jerry  [02/Jan/2002:09:26:26 +0800] "GET
http://musiccity.streamcastnetworks.com/headertop.htm HTTP/1.0" 304 59

10.101.45.177 – jerry  [02/Jan/2002:09:26:35 +0800] "GET http://ads.musiccity.com/
HTTP/1.0" 200 2931
```

The format allows for the ability to track the IP address for the device accessing the network, the logged in users name, the time of the request, the URL requested, the status (i.e. was it successfully downloaded) and finally the file size that was transmitted to the client device or application. This tuple provides sufficient data to calculate a wide range of statistics about the World Wide Web activity of users and the organisation as a whole.

2.7.2 Squid Proxy Cache Log File Format

Squid, by default stores all access attempts to the proxy server in a file called *access.log* using the basic Squid log file format (See Appendix A). The basic format is as follows:

time elapsed remotehost code/status bytes method URL rfc931 peerstatus/peerhost type

Squid can write a log file in CLF format natively if directed to do so by a configuration setting in the *squid.conf*. However, for a complete analysis it is best practice to use Squid native log format due to its higher level of detail. The Squid format as distinct from the CLF format embeds latency information into the log file for analysis. This latency information allows for a more extensive analysis of a user's activity on the World Wide Web.

2.7.3 Novell BorderManager Logfile Format

By default, the Novell BorderManager logs user access in CLF format and as mentioned before, this format has adequate data for analysis. It can also log in Extended Format and/or Indexed Format (See Appendix A). These extended formats allow for some extra analysis to occur with regard to a user's web access. These formats are not often utilised as CLF has sufficient information embedded in it.

2.7.4 Microsoft Proxy 2 Logfile Format

This is a proprietary format that has 23 fields that are used to describe a log file tuple (See Appendix A). Much of the information provided in these 23 fields is largely superfluous to needs for even an exacting analysis. Consequently, Microsoft Proxy 2 logfiles are larger and bulkier files when compared to CLF to process and manipulate in native form. Due to this extraneous information in the log file format it is not included in many log file analysis tools. These log files often need a customised script that will convert the file to a CLF form for easier analysis of the World Wide Web usage. Otherwise, any analysis conducted is limited to a narrow band of tools that are capable of natively supporting the Microsoft Proxy 2 format.

2.8 Conclusion

In this chapter there has been an examination and outline of the issues relating to the misuse of the Internet and the World Wide Web in organisations. The examination found that the misuse is a complex issue and has many factors that affect its construction and existence within organisations. The review also examined and uncovered various issues with the valid and reliability of information systems industry surveys relating to the use of the World Wide Web.

The literature review also entailed looking at material in parallel areas such as conventional computer monitoring and related Australian law to cover the full gamut of issues relating to non-business usage of the World Wide Web in organisations. An overview of technology relevant to tracking and recording World Wide Web usage was also presented.

CHAPTER 3 - RESEARCH METHODS

3.1 Choice of research method for conduct of the study

The research method used is that of multiple interpretive case study (Benbasat & Weber, 1996; Cavaye, 1996; Davis, 1992; Klein & Myers, 1999; Lee, 1989, Klein, 1999 #59; Walsham, 1995). Such a *modus operandi* was used as it concurs with the author's belief that meaning is both socially constructed and interpreted. Meaning is achieved through the author's perceptions and interpretations of the data and actions of participants within each case. The complete thesis, as a result, is filtered through the author's perspective whilst he acknowledges, addresses and reduces any issues of bias that may arise.

Guba & Lincoln (1994) identified four paradigms of case study research namely positivist, post-positivist, constructivist and critical theory. In this thesis the research framework of Orlikowski & Baroudi (1991) will be adopted that determines research as either positivist, interpretivist or critical. According to (*ibid,* p.5) positivist research is fixated around the *priori* of determinate relationships within phenomena which typically are investigated using structured methods and techniques. Interpretivist approaches (*ibid,* p.5) assume that people create and associate their meanings as a result of interaction with the world around them. Critical studies (*ibid,* p.5-6) aim is to critique the status quo through the exposure of embedded contradictions within social systems. It is then hoped that the exposure of these contradictions will enable their emancipation.

The authors' epistemological viewpoint for this research is further reinforced by the following qoutation "IS research can be classified as interpretive if it is assumed that our knowledge of reality is gained only through social constructions such as language, consciousness, shared meanings, documents, tools, and other artefacts."(Klein & Myers, 1999, p.68). In the book by (Walsham, 1993, p. 4-5) it states that interpretive methods of research in information systems are "aimed at producing an understanding of the context of the information system, and the process whereby the information system influences and is influenced by the context".

The author believes that in the process of researching non-business usage, any such uncovered usage, is as a result of the way that the World Wide Web systems are utilised within the given organisational context. The use of the World Wide Web systems is as a result of the social interactions within, among and between stakeholders in the case organisations. Stakeholders, with multiple agendas and worldviews, construct, control and use these World Wide Web systems within a specific organisational context. All stakeholders use tools, documents and artefacts to access or gather information from the World Wide Web systems. The underlying infrastructure such as TCP/IP protocol, networks and communications conduits are artefacts that allow this access to the World Wide Web. Within these organisations the meaning of the term "non-business" or "misuse" is a shared meaning generated by the various stakeholders interacting with the system within the given organisational framework. Due to the number of variables affecting the use of the World Wide Web within such contexts, it is necessary to use an interpretive approach that encompasses all possible views within the organisation. Therefore, one can rationalise that the use and misuse of these systems is a constructed phenomenon specific to their contexts.

The research is not purely positivist as many of the relationships between the data are not determinate and can not be readily and sufficiently analysed with pre-determined rules or logic. Some of the data collected will be ideal for positivist examination and investigation namely the log file data. Clearly the data will allow for positivist observations to be made about users' activities within an interpretivist framework.

The research is not of the critical paradigm. While the study may uncover embedded contradictions within the existing system, this is not it's singular purpose. Furthermore, while the research examines the status quo of World Wide Web use within the organisations, it does not seek to uncover contradictions. This assertion assuages any bias the study generates as a result of the presumption that there may be underlying idiosyncratic behaviours in the system at the outset.

Multiple interpretive case study has been chosen by the author to verify that findings are not merely as a result of the idiosyncrasies of a particular case [Miles & Huberman, 1984] cited in (Cavaye, 1996). Multiple cases also allow for replication of logic when conducting the cases and for the author to draw implications common in the cases examined. The case study approach allows a researcher to draw implications without generalising the findings due to the multi-variate nature of the cases.

3.2 Application of the Seven Guiding Principles for Interpretivist Case Studies in Information Systems to the Research

Klein and Myers (1999) define seven guiding principles for conducting interpretivist case studies in Information Systems. These guiding principles will be used to frame the research. They are:

- The Fundamental Principle of the Hermeneutic Circle
- The Principle of Conceptualisation
- The Principle of Interaction Between the Researchers and the Subjects
- The Principle of Abstraction and Generalization
- The Principle of Dialogical Reasoning
- The Principle of Multiple Interpretations
- The Principle of Suspicion

(Klein & Myers, 1999) stress, “that all seven principles depend upon each other and form an interdependent whole”. In the following sections, the author will discuss each of the principles and how they will apply or affect this particular research.

3.2.1 The Fundamental Principle of the Hermeneutic Circle

The Principle of the Hermeneutic Circle suggests that all human understanding is achieved by iterating between considering the meaning of parts and the whole that they make up (*ibid*, 1999). In the article by (*ibid*, 1999, p.71), via the use of a sentence about football, they assert "The process of interpretation moves from a precursory understanding of the parts to the whole and from a global understanding of the whole context back to an improved understanding of each part".

The research method in the cases is such that there must be an understanding of parts of the whole, namely: the non-business users' usage and behaviour patterns. This requirement allows for a greater understanding of the traffic patterns as a whole, which reciprocates with a greater understanding of each part in this case, the individual traffic patterns of use. The concept of the hermeneutic circle also suggests a cycle or series of iterations to achieve an interpreted meaning. The authors (*ibid,* p.73) state "During repeated cycles of the hermeneutic circle, all of the suggested principles can be applied iteratively, forming a complex web of interpretations". The research employs several phases of analysis and the use of multiple analysis tools aid in enabling the iterations.

Interpretivist viewpoints assert that the principle of human understanding is fundamental and rudimentary to all the other principles of interpretivist research. This research is highly interpretivist, as the author believes the study of abuse is contextual and localised within a particular organisation. There are factors affecting the notion of non-business use within an organisation that are wide ranging from hard technical issues such as available bandwidth, and proxy performance through to soft social issues such as management style or policy enforcement. As such a researcher must consider each of these in isolation and as a whole to obtain an understanding of the problem situation.

3.2.2 The Principle of Contextualisation

The Principle of Contextualisation requires the researcher to reflect critically on the social and historical background i.e. the context of the research setting. This principle allows the intended audience to see how the current situation under investigation has evolved and emerged (Klein & Myers, 1999). As previously, stated the author believes that the notion of abuse is highly contextualised to the organisation within which it occurs. The research will gather a wide variety of research artefacts such as policy documents, interviews and computer data to build a rich understanding of the context within which the non-business usage is occurring. Through a deeper and richer understanding, it is hoped that a more critical and intense reflection can be undertaken on the social and historical background of the research situation.

3.2.3 The Principle of Interaction between the Researchers and the Subjects

The Principle of Interaction between the Researchers and the Subjects sees the need for the researcher to be placed within a historical framework of interaction between his/herself and the subjects. The researcher should be aware that his/her actions do affect history via their interaction with the research subjects (*ibid*, 1999). The construction of the facts, evidence or artefacts is done because of interpretation and investigation by the researcher and the interactions within, among and between subjects and research data. This effect is reduced in this study as it relies primarily on secondary data sources: that is, the log files for the interpretation of the key results and findings by the author. Any of the new primary sources generated because of the research discourse such as interviews have a comparative empirical counterbalance in the form of the data extracted from the log files. The log files gathered will, where possible; pre-date the researcher's engagement with the target organisations to remove the possibility of Hawthorne effects (Mayo, 1933) by the participants.

3.2.4 The Principle of Abstraction and Generalisation

The Principle of Abstraction and Generalisation implies there is a philosophical basis for abstraction and generalisation in interpretive studies. This principle is specifically where the data or events that having gone through the application of The Hermeneutic Circle and The Principle of Contextualisation are left unexplained. The important concept argued by (Klein & Myers, 1999) is that theory and its construction play an important role in interpretivist research. The authors distinguish between positivist and interpretivist approaches to theory in that positivists use theory to prove or disprove categorically, whereas interpretivists use it as a lense with which to view the world or the subject matter.

3.2.5 The Principle of Multiple Interpretations

The Principle of Multiple Interpretations requires sensitivity to possible differences in interpretations of the research situation among the different participants in the study (*ibid*, 1999). This principle is because people will interpret events and settings in a variety of ways even when examining or experiencing the same events and environment. This will be addressed by comparing qualitative interview data with quantitative data from the log files that essentially provide demarcation between the perceived reality and actual reality. People within organisations will also have different views of a problem situation based upon their *weltanschauung* or philosophical view of life. Their *weltanschauung* will affect how they interpret a situation and, hence, draw any conclusions, thoughts or create axioms they would hold to be “truth”. For example, a support technician will have a different view of the situation than that of a high level manager and their collective views will have to be taken into account when examining and reporting the findings to the organisation. Another example, would be that an atheist will have different views on World Wide Web content when compared to a religious fundamentalist of any religious faith.

3.2.6 The Principle of Suspicion

The Principle of Suspicion requires that the researcher is sensitive to possible biases and systematic distortions in the narratives and the data collected from the participants in the study (*ibid*, 1999). This principle should be taken from a variety of perspectives for example the participants potentially being deceptive or misleading in conduct, personal biases of the researcher, and/or that data being incorrect.

Due to the nature of the current research and the potential highly sensitive and topical nature the author is suspicious of data that will be derived from interview. This suspicion is due to many factors including Hawthornian effects (Mayo, 1933), the participants trying to shield their behaviours or cover up possible activity that could be illicit, illegal or result in the possible termination of their own employment and/or that of their colleagues. The collection of existing log file data that pre-dates the author’s engagement with the organisations will also help alleviate any distortions in the data gathered. One of the major reasons for undertaking this research was the authors' suspicion of existing literature that suggests there is misuse of organisational World Wide Web connections during business hours.

3.2.7 The Principle of Dialogical Reasoning

Similarly, the Principle of Dialogical Reasoning requires the researcher to have sensitivity about the possible contradictions between the theoretical preconceptions that are guiding the research design and the actual findings of the research (*ibid*, 1999). The preconception that may be contradictory is that the author believes that organisations no matter how well policed or administered will have occurrences of misuse within them.

3.3 Other Issues

The research used appropriate quantitative and qualitative tools so as to increase the richness and reliability of the study. Quantitative tools used were World Wide Web proxy server log file analysis tools; qualitative tools did include open interviews and textual analysis of policy documents. The eclectic use of qualitative and quantitative methods was intended to create a richer, robust understanding of the phenomenon under examination (Cavaye, 1996). Qualitative and quantitative methods can also be used alternatively to highlight or further examine unexpected findings [Sieber, 1973] in (Cavaye, 1996). The use of quantitative tools can also be used to triangulate and verify any findings that are brought out as a result of using qualitative instruments.

The notion of misuse or non-business use is essentially constructed because of the social interactions that occur within a particular organisation. In particular, the relationships within the organisation will influence how the enforcement and implementation of many of the procedures and policies relating to the use of the World Wide Web occurs. The social relationships will also select which stakeholder or stakeholders are the determinants of the construction of appropriate use or inappropriate use within that organisation. There are also other factors and influences such as pressures from external stakeholders such as governments, regulatory bodies and shareholders.

As pointed out in Cavaye (1996) there is no pre-determined number of cases that must be studied. It was anticipated there would be 3 to 5 organisations that might be studied. The author believes it was possible to achieve this number, due to the high level of automation in the collation and collection of data and its subsequent analysis.

CHAPTER 4 - CASE STUDY DESIGN

4.1 Selection of Organisations

Organisations used for this research were selected from a variety of government and non-government institutions from Western Australia. The organisation selection was limited to Western Australia due to the level of access needed to the organisations concerned. These organisations are connected to the Internet and allow people within the organisations to access the World Wide Web. The organisations have access provided via World Wide Web proxy caching servers that have logging enabled. The log files contain the details of content that users within the organisation have accessed on the World Wide Web.

Organisations selected have a user base that will generate sufficient data for meaningful analysis of the log files. This organisation is an organisation of more than 200 users, who collectively download more than 1000MB in raw volume of traffic per calendar month from the Internet. From the author's professional experience this level of activity constitutes sufficient data for interpretation. The more longitudinal the log files the better, but a 3 months collection should be sufficient to have a reasonable indication of usage patterns of World Wide Web users.

For the purposes of the data collection, organisations are classified on the number of World Wide Web users, the size of connection to the World Wide Web and volume of traffic downloaded per calendar month.

4.2 Document Review

A review of any existing network policy documentation was undertaken on initial engagement with each organisation. Copies of any existing documentation and procedures were taken with particular emphasis on acceptable usage policies or network use guidelines for that organisation. These documents were analysed to determine inappropriate and appropriate behaviours as specified by the organisation policy. Along with initial interview data, this document review data guided the categorisation of misuse for each particular organisation. In turn, this process guided the selection of each category's rating of Acceptable or Unacceptable within the Cyfin tool.

4.3 Log File Analysis

The data initially was analysed using a logfile analysis tool called Cyfin. This tool was chosen as it has the ability to:

- use predefined and custom categories to identify World Wide Web traffic (See Appendix B for full listing of predefined categories);
- produce highly customised usage reports enabling tailoring of outputs to individual contexts;
- read a variety of cache log files from commonly used caching engines such as MS Proxy, Squid, Microsoft ISA without the need for conversion to common formats such as W3C Common Logfile Format;
- produces anonymised outputs that will protect the identity of organisational participants.

Furthermore, the data was analysed to determine to what extent the volume of the World Wide Web traffic was used for non-business usage. This analysis was then used to provide an initial estimation of the volume of the misuse and its costs based on content consumed. This estimation is not an exacting measure as each organisation had different infrastructure provision costs. The main fixed cost sources would be staffing, hardware and software licensing for connection to the Internet.

The use of Cyfin also allowed the author to target problem users of the systems for more extensive analysis. This targeting was done by identifying users with high usage and also users who had high accesses of illicit or banned material. This selection then allowed the author to examine extensively the patterns and usage of problem misusers within the organisations, and allowed the author to gain a greater understanding of a whole organisation's non-business usage and provide for the maintenance of the hermeneutic circle (Klein & Myers, 1999).

Providing triangulation of data in research helps to increase the validity of results. The log files, in particular, are the largest data set in the study and should be subject to verification and triangulation. Several log file analysers in addition to Cyfin were used to analyse the log file data sets. This multiple approach was used primarily to verify the initial analysis of the log files by the Cyfin tool for accuracy and completeness. The other analysers also presented that data in different views. For example, some tools presented a highly textual output as opposed to a highly graphical one. The types of extended analysis done by each package varied, but all packages provide basic comparative statistics.

Some analysis packages attempt to provide a timeline or histogram of usage whereas another package uses a series of statistical measures to illustrate usage patterns. The use of several tools to analyse the problem situation allowed, the author to use different views of the data to obtain a richer interpretation and understanding of the data.

4.4 Interviews

A guiding questionnaire was used as a framework for open-ended response with key organisational stakeholders responsible for the management of the World Wide Web proxy server systems, namely Organisational Management and IT/IS Support Staff responsible.

The management group was chosen for investigation, as existing literature points to management support for initiatives being pivotal to their success. This choice concurs with other findings of similar success factors for information systems implementations (Pinto & Slevin, 1989; Robey, 1979). Management arbitrates and enforces organisation policy. In Harrington (1996), it is stated that "Moreover, top management support may be more effective than codes of ethics because the effect of legal sanctions is not as great as the effects of other sources of control, such as future employment chances and opportunities" (Harrington, 1996, p.273).

IT/IS Support Staff were selected, as they are the key stakeholders who are responsible for the day to day running and supervision of the systems that allow others to access the World Wide Web. Such staff are responsible typically for the technical implementation of management guidelines based on organisation policy for the use of the system. How well the systems are installed, monitored and reported should affect end-user and management perception of misuse within the organisation. An examination of practices within this sphere of organisational influence enabled a better understanding of the issues facing the organisation. For instance, if there is no formal reporting structures in place for World Wide Web usage the users will possibly realise that their chances of being detected using the system for non-business usage will be minimal. This lack of reporting could then in turn affect how users interact with the system.

The interviews were conducted at two stages of each case study with management and IS/IT staff. These were the interviews before analysis of the log file data, and a post analysis presentation and interview designed to gain, feedback on the findings of the log file analysis.

The initial interview was an attempt to establish the participants level of policy awareness through direct questioning. Furthermore, an attempt to gauge the perceived level of policy enforcement and initial understanding of the problem situation by key stakeholders was undertaken.

After the presentation of the initial findings the post log file analysis interviews were conducted with the same organisational management and IT support staff. The interviews were a valuable tool in analysing any differences between perceived use of the World Wide Web system and the actual use determined by the analysis of the log files. The interviews were a means to discover some of the more "soft" issues or human elements of the study. It enabled the author to see what reactions there were to any complimentary or adverse findings.

4.5 Limitations of the Research

This research was limited by several factors. For example, the case study method has limitations in terms of generalisation, however, the findings can be subjected to the logic of replication. A second limitation of case research is that it may establish relationships between variables but it cannot always indicate the causality of them.

One of the main sources of the analysis was the proxy server log files whose integrity would be considered to be reliable. However, there may be problems as follows: staff altering the log file directly, staff altering their web use patterns and, log files that may not pre-date the research engagement with the selected organisation. These files are very large in storage terms and they are typically not held permanently by most organisations. It was preferred that the log files for analysis were pre-existing. Otherwise, any log files collected post initial interview with the author would have been subject to Hawthorne effects (Mayo, 1933) by staff changing their behaviour patterns.

Another limitation was that identified organisations which were aware of existing problems with World Wide Web usage would possibly be unwilling to allow an analysis of their data. In this instance, an analysis of the World Wide Web usage may uncover or confirm already grave held suspicions about the severity of the problem with World Wide Web access within the organisation. This situation may have brought some bias into the selection and subsequent examination of the cases.

4.6 Conceptual Framework

The following is the conceptual framework for each interpretive case study within the selected organisations:

- **Interview and Survey of Selected Organisation**

  This involved the interview of key stakeholders in the deployment of the World Wide Web systems in the organisation.

- **Document Review of Selected Organisation**

  This involved the review of existing policy documents and literature relating to the use of the World Wide Web resources.

- **Analysis of Log files**

  - **Macro Analysis of Organisation**

    This generated macro organisation statistics using various log file analysers

  - **Generic Category Analysis of Organisation**

    This involved the use of the Cyfin log file analysis software to determine the level of misuse in particular categories of usage.

  - **Individual Misuser Analysis**

    This involved the use of all log file analysis tools and textual pattern searching tools to examine each specific user's activities.

- **Reporting Details of Analysis Back To Organisation**

  This is where the analysis of the log files was reported back to the key stakeholders in the organisation

- **Post Analysis Interviews**

  This step involved the post analysis interview of key stakeholders about their perceptions and thoughts about the log file analysis findings.

The conceptual framework provides for the gathering of data from a variety of sources and utilizes several techniques for data gathering. The use of interviews, document review and log file analysis allows the author to contextually investigate and interpret the level of non-business from a variety of viewpoints.

4.7 Conclusion

This chapter has explained the case study design utilized for this research. The framing for the research is that of multiple interpretive case study using Klein and Myers (1999) paper as a paragon of design. The design has incorporated where possible a selection of appropriate qualitative and quantitative techniques to increase the richness and reliability of the study. The author has also outlined any potential weaknesses in case design and how redress of these has been attempted.

CHAPTER 5 - CASE STUDY 1

5.1 Organisation Description

This organisation is a University department where students and staff have access to the World Wide Web. The client's devices have a variety of connectivity from access across a 100Mbit switched Ethernet connection on the departmental LAN linking a variety of machines from 100Mbit Ethernet connected Intel PCs down to remote computers using a 28.8K modem for connectivity to the proxy caching engine.

The proxy caching engine is a Squid Proxy server running on the Linux platform that provides web connectivity for the student laboratories and staff studies. External users need access to services within the departmental LAN, via the proxy server which is accessed by students and staff who use the university modem pool. The proxy server is managed by the technical support staff and one of whom is allocated the job of maintaining it.

5.2 Pre-Analysis Interviews

Initial open-ended interviews with key organisational staff were undertaken. The relevant staff were the Head of School, the IT Services Manager and the Technician responsible for administration of the proxy server.

All parties admitted that there was a lack of monitoring of misuse within the system and welcomed the investigation. The Head of School stated that there had been infrastructure and work practice change priorities undertaken in the last 12-18 months by the department; that had taken precedence over monitoring World Wide Web usage. When asked as to what level of misuse they thought was going on, the Head of School said he was unsure and would not be willing to guess. The Head of

School did point out that misuse, however, was now going to become an issue that had to be addressed as the charging for Internet bandwidth within the University was changing. The change was that bandwidth charges were no longer to be levied from consolidated general revenue, but redirected to school budgets on a user pays system. Consequently, the Head of School was aware that misuse could cause this cost to spiral out of control if left unchecked.

The Head of School expressed some levity in allowing students to explore and use the Internet for purposes other than strictly academic pursuits. For instance, overseas students who would be accessing foreign language news services to keep up to date with local news, while away from home. However, The Head of School did suggest that if any of the analysis indicated significant or problem misuse by users that counselling would be required. He further stated that anyone accessing material contrary to other University policies on Equity and Diversity would also need to be counselled.

When asked why no proactive monitoring was taking place the IT Services Manager stated a general lack of time to perform the required analysis and the unavailability of a suitable tool that performed easy, quick analysis of traffic patterns and readily identified misusers of the system. The IT Staff had tried several attempts at analysis in the past, but found the volume of data generated by the log files too large to handle and analyse efficiently. The IT Staff had attempted to use several freely available tools, plus some commercial trialware but did not find one that fitted requirements. IT Staff had also tried to develop some internal tools using an SQL database backend with little success.

The IT Services Manager indicated that he was aware that there was a small group of students who were misusing the resource. However, due to technical and organisational constraints, it was felt difficult to pinpoint who exactly was committing misuse. The IT manager also pointed out there was little perceived benefit in spending time pursuing misusers. This perception was due to the poor level of disciplinary action that would be taken against the transgressors of policy based on prior history of similar incidents. The pending upgrade of the proxy caching software that would allow the inclusion of usernames against log file records was something the IT manager believed would help them track misusers of the system. It was also pointed out there was no real impetus or benefits to monitor the World Wide Web usage apart from enforcement. This viewpoint was because there was no current measurable direct cost saving to the department hence, the benefit was perceived as largely intangible.

The Technician said that his responsibility was to make sure that the server was running and not the analysis of log files. The technician further proffered that, due to time constraints much of the installation of the server was not optimised and use default setting for many of the key features. This non-optimised configuration in turn would have affected the overall performance of the proxy server under times of high load.

5.3 Document Review

Several documented policies relate to the use of the World Wide Web system at the University. The first policy that relates to the misuse of the Internet is titled – 'Acceptable use of Computing Facilities'. Its first page states:

> "All users must agree to comply with the University's "Computing Facilities - Conditions of Use" before being granted access to any University computing facilities or systems"
>
> and
>
> "Students must agree to these conditions at enrolment and re-enrolment, via (enrolment system)"

The document titled 'Computing Facilities – Conditions of Use' states that "Facilities may only be used for the purposes for which they have been provided and not be used for other projects, games, 'hobby computing', private or consulting work". This statement indicates the policy intentions of the University as being one of where usage of the facilities is only for activities that directly relate to work or study. The document mentions specifically harassment of others via e-mail but makes, no mention of penalties for accessing illegal sites or sites that could constitute a threat in terms of content such as pornography, hate sites and racist material.

With regards to Copyright issues, the document states "Only legally obtained software is to be used on University computing equipment - the penalties for breaching copyright are very high. All users of University equipment are warned that any such breach is the liability of the user. The University will not be liable for any breaches made by users." This section of policy specifically relates to software and it could be argued does not cover MP3, streaming audio and video traffic that potentially contains illegally acquired content. The assumption that is made in the policy is that each user is aware of all freeware, shareware and commercial products and brands currently being used in the world today.

There is no mention of penalties for specific breaches, such as the viewing of pornography or accessing copyright materials. The document makes two references to transgression of the policy, these are:

> "Failure to adhere to the above conditions will be considered an act of grave misconduct and cancellation of enrolment may result.
>
> Breaches which involve security and/or access violations may be referred to the Australian Federal Police."

The operative word in these clauses state "may", which does not imply that severe consequences will be affected for what could be considered the top level of the breaches scale. There is no delineation of level of offence and the policy document does not outline levels of severity of breach and resultant penalties. There is no apparent mechanism for the reporting of policy breaches, nor apparent direction as to whom one would report such a breach. No mention made for instance of any of the more recent legislation in Australia such as the 2001 Cyber Crimes Act which has significant changes as to what now constitutes an offence.

Within the University, there have been no publicised policy breaches, and outcomes of breaching policy, which Harrington (1996) points out as a major success factor for implementation of policy. The policy, however, is overtly displayed in large print format in each of the student laboratories, and, as a requirement before re-enrolling, students must acknowledge that they have read and understood the conditions of the access policy. One legal point that the university has obviated by this process is that some students in their first year will be minors and, as such, this contract is not enforceable until the students reach majority at age eighteen.

University staff members do not have any reminder of policy and they are expected to have acknowledged and understood any policies when signing their initial employment contract. There is no other formal training or updating of skills for staff on the policies that relate to network access. As (Wood & Lo, 2001) and the (Australian Government Solicitor, 2001) point out, this potential failure could see any dismissal that results from breaches of the policy ultimately fail. This is because the University must be able to demonstrate due diligence on the maintenance and implementation of the policy. Take the case of a staff member who has been with the University for 6 years, he/she will have potentially not seen nor been presented with a new copy of the policy for acknowledgement and, hence, could be unaware that their practices may be in breach of the University policy.

5.4 Method of Log File Analysis

The log files analysed in this case are Squid proxy cache server log files (for full details of format see Appendix A - Squid Proxy Cache Logfile Format) . The log files covered the period 1st November 2001 to 8th March 2002 inclusive. The total volume of traffic served by this proxy server during this period was 121 Gigabytes (Gb) to 846 uniquely identified client devices producing 11,171,790 unique log file entries.

The log files were analysed by 5 different log file analysis tools these were Cyfin(Wavecrest, 2002), Squid Analysis Report Generator (SARG) (Orso, 2001), Webalizer (Barrett, 2002), pwebstats (Gleeson, 2002) and Analog (Turner, 2003).

Cyfin (Wavecrest, 2002) is a commercially available tool that allows for extensive analysis of a wide range of log files and the adaptation of reports and outputs for analysis. Cyfin allows for a high level graphical reporting of the traffic usage patterns. It also, categorises traffic into 55 preset categories (see Appendix B) and 10 customisable categories of URL's. The author used all of the preset categories and created 2 custom categories, one for Chat sites that were unclassified Web based chat engines and another for unclassified MP3 sites to aid in pin pointing problematic traffic patterns or trends relating to unclassified MP3 traffic.

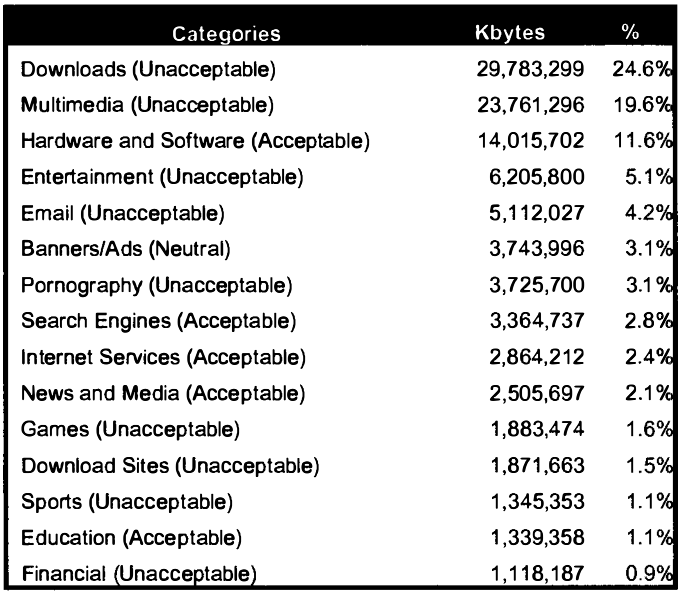
SARG (freeware) and pwebstats (freeware for educational institutions) both have the ability to process Squid proxy log files natively. They produce general statistics with a high level of granularity making it possible to determine, down to the user and file level, any activity that is generated. SARG and pwebstats, in particular, each produces a detailed 24 hour histogram of users traffic usage, which allowed the author to analyse readily hours of web usage by users.

Webalizer and Analog do not handle Squid formats natively and the log files had to be converted into Common Log file Format (Luotonen, 1995) using a perl based script. Webalizer allowed for a reasonable degree of granularity, but it was not as comprehensive as the SARG outputs. It did, however, confirm general statistics generated by the other log file processors on a user and overall basis.

5.5 Overall Results

All analysers produced the same generic statistics with a transmitted traffic volume of 121Gb and total log file line entries of 11,171,790. Everyone, identified the same top users for the content downloaded.

Analysis done with Cyfin strictly applying the guidelines for the University posited that 74.6% (90.4 Gb) of traffic was unacceptable and 25.4% (30.8 Gb) of traffic would be acceptable. The top 15 categories of usage as produced by Cyfin are displayed in Table 5.1.



Categories	Kbytes	%
Downloads (Unacceptable)	29,783,299	24.6%
Multimedia (Unacceptable)	23,761,296	19.6%
Hardware and Software (Acceptable)	14,015,702	11.6%
Entertainment (Unacceptable)	6,205,800	5.1%
Email (Unacceptable)	5,112,027	4.2%
Banners/Ads (Neutral)	3,743,996	3.1%
Pornography (Unacceptable)	3,725,700	3.1%
Search Engines (Acceptable)	3,364,737	2.8%
Internet Services (Acceptable)	2,864,212	2.4%
News and Media (Acceptable)	2,505,697	2.1%
Games (Unacceptable)	1,883,474	1.6%
Download Sites (Unacceptable)	1,871,663	1.5%
Sports (Unacceptable)	1,345,353	1.1%
Education (Acceptable)	1,339,358	1.1%
Financial (Unacceptable)	1,118,187	0.9%

Table 5.1: Usage by Categories

Analog (Turner, 2003) was used to extract the files sizes of each request made to the proxy server by the clients see Table 5.2. Analysis of file sizes and number of requests sees that 1Mb or larger files account for 66.12% of the downloaded material. This statistic would indicate quite a significant level of potential misuse by users.

File Size	Number of requests	Percentage of the total bytes
0	110,185	0.00%
1b- 10b	0	0.00%
11b- 100b	4,218	0.00%
101b- 1kb	8,204,985	2.42%
1kb- 10kb	2,019,761	5.88%
10kb-100kb	771,015	15.46%
100kb- 1Mb	33,981	10.12%
1Mb- 10Mb	14,502	33.15%
10Mb-100Mb	632	14.65%
100Mb- 1Gb	98	18.32%

Table 5.2: File sizes of requests

Web pages are rendered in HTML, or as dynamically generated content from data base driven systems such as Microsoft's ASP platform or PHP; these types of pages are largely text and highly compressed graphics. The design philosophy of most World Wide Web sites is that minimal bytes are transferred to increase speed of delivery to the end user. Single files that are over 100Kb in size would be a rarity as most web pages regardless of development are normally multi-part, multi-media depictions.

There was 14,502 files (33.15%) or 40.1 Gb that were 1Mb to 10Mb in size which is considerable. These files are not web page impressions but would indicate downloading of MP3, binary files, large archive files containing compressed files or patches to operating systems or applications software. Furthermore, 632 files (14.65%) or approximately 17.7 Gb of all downloaded content were files 10-100Mb in size. These file sizes would indicate that these were large archives, applications or sizable video or sound files. The final category had 98 files in the 100Mb to 1Gbyte range that consumed 18.32% or 22.16 Gb of the traffic. These files considerable in size, were further investigated and found to be either large system patches, games files, movie files or CDROM images of operating systems.

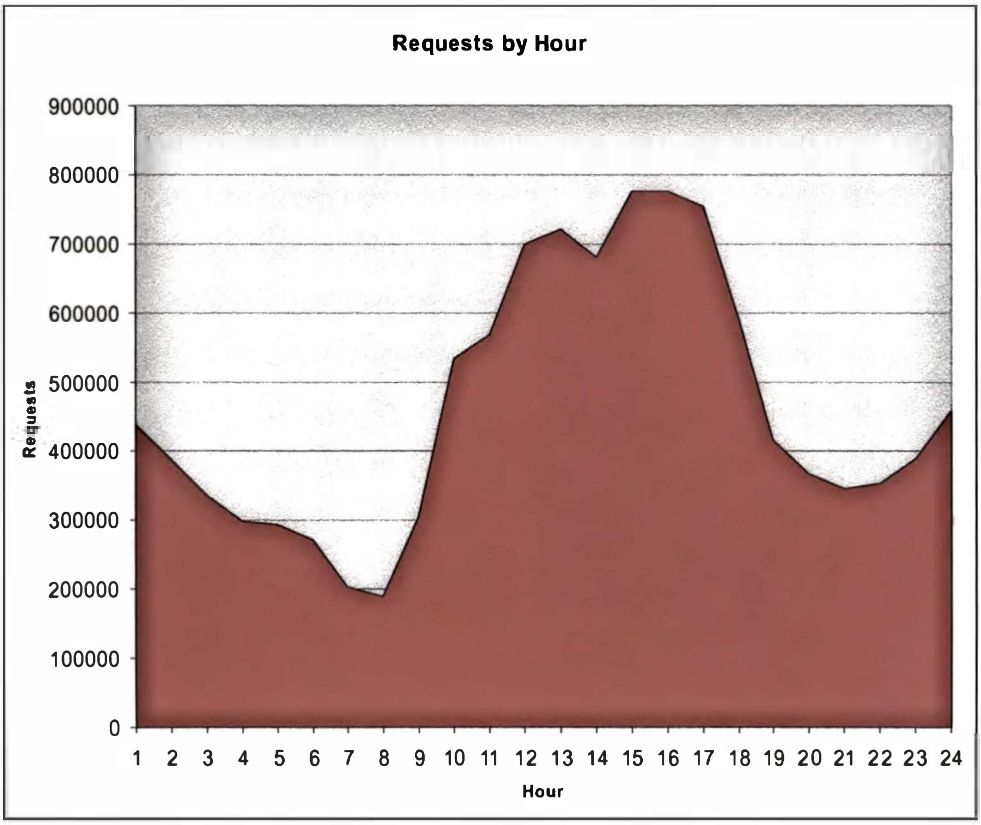


Figure 5.1: Hourly usage

Figure 5.1 is a graph showing the number of connections per hour for the proxy server for the period examined. The University has formal lectures and laboratory sessions from 9am to 9pm, Monday to Friday with Friday typically being a period of low activity for formal classes and laboratory usage. As would be expected, activity rises from the start of the day at 9am and peaks at 3pm and degrades until 9pm. The standout pattern in this traffic usage is that there is an increase in requests after 9pm and this pattern continues until 1am in the morning. Levels then remain relatively high until 4am where they directly correlate with levels at 9am in the morning. These patterns would indicate that some of the potential misuse is occurring well outside of designated laboratory hours or teaching times. It could also indicate that users are possibly cognisant of the fact that they are breaching policy and are accessing the World Wide Web at a time when the chance of physical detection by staff is almost nil.

The usage over the Christmas period dropped, as would be expected, as the University campus was shutdown from 19th December 2001 until the 6th January

2002. There was still the possibility of access to the proxy server via dialup services which was used during the closedown period. However, due to technical constraints of analogue modem connections this usage did not figure largely in consumption of traffic.

In Figure 5.2 November 2002 and February 2002 would point to student activity contributing to the higher proxy server request rates. Further detailed analysis showed that the immediate post exam week indicated a marked decline in accesses and in the week preceding the start of semester a corresponding increase in requests occurred. It should also be noted that the March month had only 8 days in the total month that could be analysed.

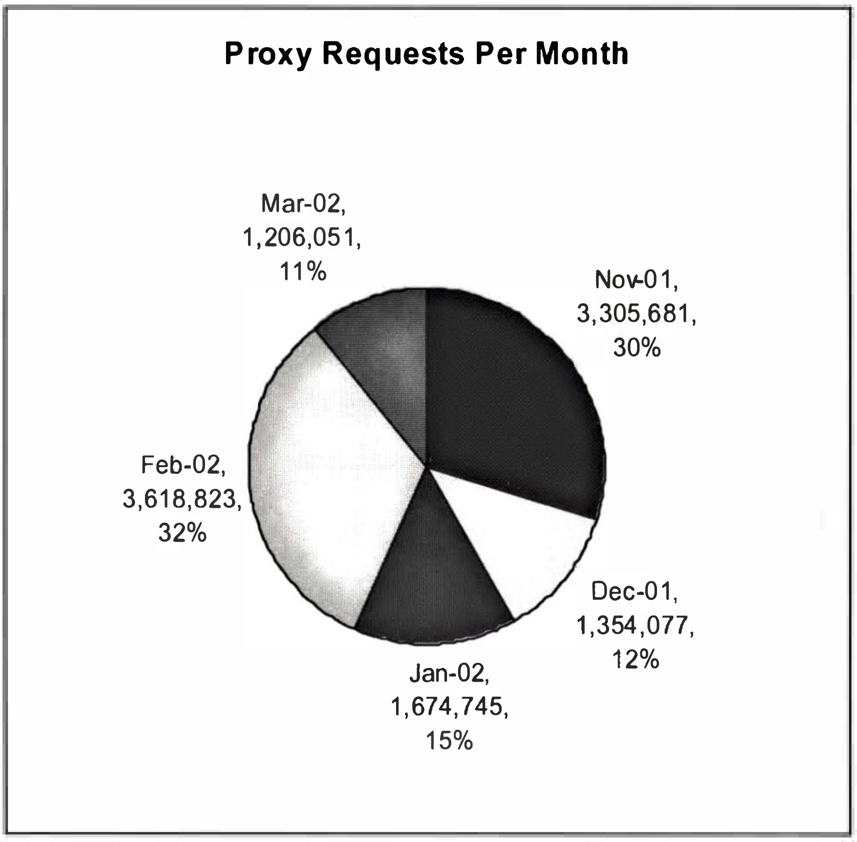


Figure: 5.2 Number of Requests per month

5.6 Analysis of Top 15 Categories

In the following section an analysis of the top 15 categories of use as indicated by the Cyfin tool will be undertaken.

5.6.1 Downloads (Unacceptable)

While as a rule, the Downloads category would be unacceptable there are occasions where the usage of the proxy server for downloads is necessary. For example, when technical staff are downloading vendor provided patches or updates to software that is deployed by the university. It was found that 133 (15.7%) of the 846 identified users accessed this category accounting for 29,783,299 kB (24.6%) of all traffic downloaded. This usage is a substantial part of the traffic usage for the department. What is interesting about this statistic is that only 15.7% of users have accounted for this consumption, which indicates that a small group of users are potentially misusing this category.

5.6.2 Multimedia (Unacceptable)

The overall patterns indicate that there is extensive downloading of entertainment or recreational related content. In all of this activity there is a heavy emphasis on streaming and sound media files and, in particular, the ASF format.

This use is surprising given that access for nearly all clients is via the caching proxy server only. The use of file sharing tools such as Napster, GNUTella was blocked via firewall rule sets. In addition to this countermeasure all laboratory based computers had strict policies that did not allow for the installation of programs onto the desktop computers. The only method of accessing streaming media was for the clients to use conventional World Wide Web services to access the Internet, or to use clients that allowed tunneling of protocols through conventional World Wide Web services. Files contained in this category are MP3 and illegal copies of movie files, the downloading of which is in direct contravention of existing university policy and copyright law.

5.6.3 Hardware and Software (Acceptable)

Due to the emphasis of the School this area was considered acceptable. The total number of users accessing this category was 536 or 63.3% of all users. These users downloaded 14,015,689 kB or 11.6% of the downloaded traffic. The top twenty users downloaded 10,593,414 kB or 75.6% of the downloaded volume in this category. This traffic indicates potential misuse and further weight is given to this argument when the top five users have downloaded 8,539,994 kB or 60.9% of the content. The top five users in this category all were student based users of the system and this statistic would add further credence to the argument that this is a potential heavy misuse of the World Wide Web by these users.

5.6.4 Entertainment (Unacceptable)

The total downloaded volume for this category was 6,192, 827 kB by 387 or 45.7% of users. This category had the top 20 users consuming 3,507,873 kB or 56.6% of the downloaded content. This statistic would indicate that there is some potential misuse of the system occurring with student users consuming a large proportion of this bandwidth.

5.6.5 E-mail (Unacceptable)

The web based e-mail downloaded was 5,112,027 kB produced by 420 users or 49.6% users of the system. The top 20 users in this category consumed 2,191,231 kB or 42.8% of the volume downloaded. The usage by the top 20 users in particular, would indicate a very high volume of e-mails being sent. The top e-mail user sent 211,277 Kb (4.1%) which is considerable when taken in the context that many e-mail messages are less than 2Kb in size.

This metric indicates significant misuse by the users of the system as all users in the system are allocated conventional mail services and accounts as standard network practice. None of the URLs in this category were pointing to any of the designated university-based e-mail servers. The amount of traffic that is downloaded is high, given that many free Web based e-mail services only allocate a 1 MB sized account for the transmission and receiving of e-mail, which makes the use of any attachment based e-mail virtually impossible. While the use of this form of traffic is not strictly against policy it is deemed unacceptable, as students are all provided with ready access to e-mail services within the University.

5.6.6 Banners/Ads (Neutral)

A total of 3,743,996 kB (3.1%) of material was downloaded in this category. This material is typically included in web pages or in some cases, may utilise push technology to deliver customised content to the users desktop.

5.6.7 Pornography (Unacceptable)

A total of 219 users (25.8%) downloaded 3,725,700 kB (3.1%) of material. This measurement clearly demonstrates misuse of the system as the accessing of pornography is in direct contravention of policy. The top twenty users (9.1%) consumed 2709027 kB (72.7%) of the material which is considerable. The last 199 users were responsible for only 366,167 kB which could possibly indicate unsolicited e-mail containing embedded HTML links to pornographic sites or materials.

With the top 20 users in this category consuming 72.7% this would indicate potential problem behaviours and definite misuse. Upon further investigation by the author many of these top 20 users revisited the same files or sites on several occasions. This pattern of behaviour would indicate definite and deliberate misuse of the system as their access of the material is not strictly a singular occurrence. Further evidence was evinced by the fact that many of these accesses had different casual linkages/paths followed by the students to access them each time. It would be highly doubtful that these identified URLs were as a result of opening e-mails each time. In some cases the complex URLs often appeared in the log file without any searching activity by the users indicating the use of bookmarks or pre-researched URLs.

5.6.8 Search Engines (Acceptable)

A total of 418 users (49.4%) accessed material in this category for a downloaded volume of 3,364,737 kB (2.8%). The top 20 users consumed 1957396 kB (58.1%) of material, this may be a precursor to misuse as the users are actively searching for content. What is of some concern is that potentially 428 (50.6%) users are not using search engines to access materials.

5.6.9 Internet Services (Acceptable)

It was found that 422(49.8%) users accessed services related to this category for a total download volume of 2,862,393 kB. The top 20 users consumed 1,557,796 kB (54.4%) of the total downloaded volume.

5.6.10 News and Media (Acceptable)

In this category there were 306 (36.1%) users who downloaded 2,504,334 kB of material. The top 20 users in this category consumed 1,207,088 kB (48.2%) of all material accessed. These accesses are considered acceptable as students and in particular overseas students, would be accessing foreign news services to access their local news.

5.6.11 Games (Unacceptable)

This category of usage is in direction contravention of established policy, however, 189 (22.3%) users downloaded 1,883,311 kB of such material. The top 20 users downloaded 1,489,699 kB (79.1%) of all the content. This activity is extensive and would warrant further investigation. The 169 remaining users were responsible for only 393,612 kB of download which could easily be as a result of embedded HTML content in unsolicited e-mails.

5.6.12 Download Sites (Unacceptable)

Download sites saw 183 (21.6%) users download 1,871,256 kB of content with the top 20 users responsible for 1,392,090 kB (74.3%) of this. The top 50 users consumed 94.2% of all traffic volume with the remaining 133 users responsible for 107,653 kB of material. This unacceptable activity again could possibly indicate misuse of sites within this category.

5.6.13 Sports (Unacceptable)

By strict definition this category would be unacceptable usage as described by the University's access policy. Only 196 users (23.1%) accessed such material for a total amount of 1,345,284 kB. The top 20 users accessed 769,106 kB (57.2%) of the total download. Most of the activity generated was the result of students and staff checking for up to date sports scores.

5.6.14 Education (Acceptable)

This category had 246 users (29.1%) accessing such systems with a download content of 1,345,353 kB (1.1%). This metric seems low at first glance seeing as the proxy server is located within a university and the school makes extensive use of e-learning technologies. However, the proxy server does not proxy the server local domain or server IP address space and hence, this only represents access of educational sources external to the University. A potential issue was the top user who accessed 106,598 kB (8.0%) of content in the category.

5.6.15 Financial (Unacceptable)

Of the 846 users accessing the proxy server 221 (26.1%) of them consumed 1,118,096 kB (0.9%). The top 20 users of this category were responsible for 763,245 kB (68.2%) of the downloaded materials. Most of the accesses were to share trading and personal on-line banking sites. A large amount of the accesses were occurring from staff areas of the network and were primarily focussed at share trading sites, which would indicate non-business related use.

5.7 Analysis of Top Users

In the log files examined only 14 users out a total of 846 (1.6% of users) were responsible for consuming over 50% of this bandwidth, 56 users (6.6% of users) used 75% and the other 750 users consumed the remainder. The focus of this section closely examines the actions of the top 14 users of the system for that period their consumption is listed in Table 5.3

	Requests To Proxy	*Traffic Megabytes*	*kB Per Request*	*% of Download Total*
PGStudent1	691,766	11418.8	16.5	9.24
PGStudent2	137,076	8379.7	61.1	6.78
Staff1	10,114	8141.7	805.0	6.59
Student1	269,346	7242.6	26.9	5.86
External1	125,827	5125.2	40.7	4.15
Student2	180,161	4020.6	22.3	3.25
Student3	325,111	3626.6	11.2	2.93
Staff2	74,111	3171.4	42.8	2.57
Student4	134	2501.9	18,670.8	2.02
Student5	78,778	2150.1	27.3	1.74
Student6	117,794	1899.1	16.1	1.54
Staff3	514,021	1623.6	3.2	1.31
Student7	113,411	1593.8	14.1	1.29
Staff4	46,367	1525.5	32.9	1.23

Table 5.3: Usage statistics for users consuming 50% of bandwidth

Of the top 14 users, one IP address was an external address to the university, this fact at first appeared to be problematic. Upon investigation with the IT manager it was determined that this IP was a peered proxy cache server, which is a mutually beneficial arrangement whereby organisations benefit by transferring already cached data between their servers. Thirteen actual users are left for analysis seven general students, two postgraduate students and four staff members.

For each of the top 13 users, their World Wide Web access information was filtered from the log files and the resultant data stored in 13 unique files. These files were then processed to enable extensive analysis of their World Wide Web traffic usage patterns. The degree of granularity was such that the author could drill down to which user transferred individual files at what time.

5.7.1 Technical Staff

Three of the staff members Staff1, Staff2, Staff4 were technical support staff whose usage was considered business usage as they were downloading patches and fixes for installed systems. Over 90% of downloaded volume for all three parties was either patch files or major version revision upgrade software for the operation of devices in the system. The remainder of the traffic was either based directly with work for example scouting for patches and upgrades or replying to e-mail traffic via web based systems. The only detected non-business usage was the occasional checking of progress scores from various sporting events.

5.7.2 Postgraduatel & Postgraduate2

Both postgraduate students displayed behaviours that can only be considered as naïve for their inappropriate usage of the resources. Combined, they used 20 Gigabytes of traffic volume causing considerable throughput for the cache. They did not directly download a large percentage of the traffic from the Internet, but transferred it via the proxy server across the local area network.

The students who generated this traffic could be described as non-malicious, ignorant users of bandwidth. Some 11,263 Mb PGStudentl passed through the proxy server, a total of 10,151 Mb of this traffic was video based traffic, and directed from the local area network. Similarly, PGStudent2 produced 8,502Mb of traffic of which 8,002 Mb was video based traffic and was generated on the local area network. The fact that this traffic was on the internal local area network running across a fibre based 1GBit redundant backbone, to a switched 100Mbit network connection at the desktop should have not presented a issue if it was typical network traffic. However, the video camera would have effectively bombarded the cache memory on the Squid proxy caching server which has some severe implications for cache serving ability. To compound the problem the content being sent on video camera occurred within 31 hours, over 5 days and typically in peak usage times for other legitimate users of the cache.

One of the main problems is the volume of traffic that this activity would have generated. In the top 8 hours of the usage, there was around 10Gbits of network throughput per hour. This extra traffic would have meant that the cache would have been severely hampered in its ability to service legitimate traffic due to the additional loading placed upon it. Furthermore, the traffic was streaming video and the number of cacheable objects that this generated was substantive.

The effect on the cache would have been to flush legitimate objects off the semi-permanent disk cache and also from the cache memory of the caching server.

This effect would see previously accessed material from the Internet being wiped from the cache, which means that several gigabytes of already cached relevant and reusable objects would have been deleted. By doing so any benefit gained by using a cache would have been lost as considerable re-caching of content would have to occur before any performance improvement is noticed again.

5.7.3 Student 1, Student 2 and Student 3

There was 15,888 Mb of World Wide Web content that was accessed by these 3 students through this proxy. It was found that 90% of this content was unacceptable and not adhering to University policy. A high proportion of that 90% was downloading and viewing streaming media, such as pirated versions of new release movies, pornographic movies or song files. The problem here is not really one of money but the liability that these abusive behaviours present for the University.

Students accessing this material were doing so during hours 1am – 6 am where the only potential for detection was essentially an audit of the proxy server log files. This behaviour would indicate that the students were aware that their actions were potentially abusive, even though they may have been unaware of the University's acceptable usage policy. Furthermore, the fact that they are potentially breaking several criminal codes by commissioning of these activities would further strengthen this argument.

5.7.3.1 Student 1

Studentl downloaded 7,242 Mb of traffic of which, 5357 Mb or 74% was streaming movie files mostly using the .ASF format. Of the remaining bandwidth, 1652 Mb were MP3 song files or MPEG and AVI based pornographic movies and, less than 300Mb of this could be classified as legitimate traffic.

The student perpetrated most of these accesses between the hours of 11pm and 6am. This usage on the 2 worst days was 2,762 Mb on a Wednesday and Thursday in December 2002, well outside of standard semester time. There were only 19 hours idle out of the 48-hour period of accesses during this time. The accesses during this period were a selection of sex, sound and cinema files.

5.7.3.2 Student 2

Student2 downloaded 4,020 Mb of data from the proxy server of which 89% or 3,522 Mb was unacceptable traffic. This student was more dedicated to the download of streaming media be they movie files, pornography or MP3 song files.

Of the remaining 498 Mb, there was 170Mb of e-mail traffic, which is large when these students are provided with conventional POP3 accounts for e-mail. What is of further interest is 45MB of chat traffic, which generated approximately 1800 visits to various chat sites. What makes this figure large is that the student had only generated this traffic in 295 hours, which equates to 6 visits per hour to a chat site. The inappropriate accesses by the student were perpetrated mainly during the hours of 11pm and 6am. The accesses by this student though were more evenly spread which showed a consistent rather than erratic pattern of misuse.

5.7.3.3 Student 3

The analysis for Student3 followed similar patterns to the other aforementioned students with 3,200 Mb of traffic dedicated to the download of movies and MP3 files. The behaviour patterns of waiting until the small hours of the morning to perpetrate much of this misuse was also demonstrated in the log file analysis reports for this user.

5.7.4 Student 4

In 2 calendar days Student4 downloaded 2501Mb of data, of which 2481 Mb were copies of an Oracle database implementation for a Linux based Intel server. Students were able to get a copy of this product and the Linux distribution legally through the school for educational purposes. The download session was targeted at getting the latest version of Oracle Database server for Linux and the latest patches to run it on a Linux Redhat 7.2 server based implementation. What makes this traffic suspicious is its excessive volume of data and that it was downloaded some 10 days after the end of examinations for the year.

5.7.5 Student 5 and Student 6

Student5 had 665Mb (30.9%) of legitimate traffic out of 2150Mb. Student6 had 253Mb (13.3%) that was legitimate traffic out of a total download of 1899Mb for the period examined. In both of these cases, the remainder was primarily turned over to downloading programs and utilities to burn writable CD-ROM media. Other downloads included utilities that could be used for the manipulation, download, and conversion of audio files to and from popular formats, for example WAV to MP3.

Both students were downloading a wide range of copyrighted song files that it is assumed were listened to and/or stored on CD. This assumption is due to the fact that both students downloaded the Nero CDROM burning tools, and downloaded various tools for “ripping” CDs which is the process of converting conventional CD tracks into digital file formats such as MP3 or WMA. Upon checking with the technical staff, it was verified that a person using a USB based CD Writer could attach this device successfully to a laboratory machine and, because of Windows XP in-built CD software, would be able to burn files to a CD for later use.

5.7.6 Student 7

Student7 appears to access an on-line storage system for files with this traffic representing almost 25% of the 1,593 Mb of traffic accessed. A 551MB file named Netware6.zip was downloaded not long after the official release of Netware6, which could reasonably treated as suspicious. There is a smattering of MP3 song files and small amounts of pornographic files in the traffic. The remaining content was wide in scope representing 24 out of 55 different categories that the Cyfin (Wavecrest, 2002) analysis tool had as preset categories.

5.7.7 Staff3

Staff3 has what appears to be a modest 1,623 Mb of bandwidth of which 1,200MB was downloading of patches for his/her systems which is legitimate traffic: however, this is where legitimacy ends.

The remaining bandwidth, while only 423Mb is problematic. MP3 and movie traffic accounted for 190Mb of this 423Mb, which is against policy, and some of the traffic was in breach of copyright.

Of the remaining non-legitimate content accessed, pornographic material accounted for 45MB of this activity. What made this case potentially problematic was that Staff3 had paid memberships to several hardcore pornographic sites that they were accessing from this connection, providing a further audit trail for the activity.

A potential productivity issue was that Staff3 had a high number of requests to a stock market quote engine on the web that accounted for 7.45% of their on-line time. The rest of their surfing pattern accessed a wide category of sites representing 26 different categories from a selection of 55 in the Cyfin (Wavecrest, 2002) analysis tools set of base categories and presented no real problem.

5.8 Post Analysis Interview

Post analysis interviews were conducted with the Head of School and the IT Manager. The Head of School when asked about the level of misuse indicated that there was some surprise at the result. The Head of School further pointed out that the misuse by some students particularly the pornography may involve some need for individual counselling. However, the Head of School did say that there was the expectation that the level of pornography would have been higher.

Students who were accessing excessive volumes of material from the World Wide Web would be sent a letter informing them that they were over allocated and that they should arrange an interview with the Head of School. When asked as to what the benchmark for this level would be the Head of School proffered that it would be, similar to what an active research staff member would use.

The Head of School was concerned at the level of copyright materials that the students may be accessing and any liability that this may cause for the students. The Head of School was cognisant of the fact that some of the usage issues brought as a result of study need to be addressed before the new charging regime came into effect, so that they would not impact on the school.

The IT Manager was not surprised by the outcome and appeared cynical of any action that would be taken against students or staff who would be identified if they conducted the same tests. The IT Manager pointed out that the level of pornography was lower than expected and, that statistics generated concurred with tests they had run in the past from the log files, when no action had been taken.

5.9 Discussion of the Case Results

One of the patterns emergent in this case is that the unacceptable categories have the top twenty users in each category consuming between 60-90% of the bandwidth for the particular category. This pattern would indicate that these users are misusing the system as the volume of material downloaded is substantive and disproportionate.

Accessing unacceptable categories is an identifiable high-risk behaviour for users. Such access of material, for instance pornography, could see possible punitive measures against users for breaching policy. These high-risk categories have low numbers of users accessing them due to the attendant risks and they also tend to have the top 20 users consuming much of the bandwidth for that category. The pornography example shows that the number of users deliberately accessing such material is 219 users (25.8%) reflecting that one in four users are potentially viewing pornography. When you take into account 72.2% of all material downloaded in the pornography category was accessed by the top 20 users and the remaining 199 users access only accounted for 366,167 kB this situation would point to significant misuse by the top 20 users.

The analysis of all the results highlighted the ubiquity of MP3 and video based traffic in this system. Apart from the fact that it is against established policy for users to be generating this sort of traffic, there is a major risk in terms of copyright violation. Many of the MP3 and video files downloaded were pirate copies of original works. The filenames typically indicated the performing artist, quite often the album, its order of play and the song or video title in full.

In addition to the downloading of these copyright files it appears that many of the students were arming themselves with a variety of utilities for recording the downloaded files to CD-ROM for possible later use/distribution. These utilities included several shareware versions of CD-ROM burning software that function for a given time period or number of sessions before expiring.

Not all of the traffic analysed had malicious or abusive implications. In fact the largest detected misuse of the proxy was that of the 2 postgraduate students unwittingly placing a video camera on the network. This action generated sufficient traffic over five days during peak usage periods to overload the proxy server to such a point that it would have hampered legitimate use of the Internet connection for other users.

If reports by (Foster, 2001; Hamilton, 2002; Hickins, 1999) on World Wide Web misuse and in particular, pornography are taken at face value, then the pornography issue in this case is minimal. However, the downloading of pornographic material, although not large in volume when compared to other traffic, does represent quite a significant problem for the University. The students who were accessing this sort of material were doing so at times where the detection of this activity would be highly unlikely except via an audit of log files. The volume of material that the top 20 students downloaded, coupled with the fact that they requested some of these files more than once from the system, would indicate possible problem behaviours on their part.

Staff3 who was accessing pornographic material was not doing so in an ad-hoc or opportunistic fashion like the students were. Staff3 was actually a paid member to several hardcore pornographic sites this behaviour indicates a habitual usage of pornography inside and out of work. The area of concern about Staff3 is that he was deliberately accessing these sites during the working day in breach of existing University policy and potentially jeopardising his employment. Staff3 could have a difficult time defending his actions should it come to a dismissal situation.

There is University policy that indicates the downloading of pornography is considered a serious breach. The only defence possibly to these actions is that there is not regular training and updates on the computer access policy within the University.

The actions of Staff3 are further brought into question when 7.5% of web browsing time is spent requesting quotations from an on-line stock quoting facility. This level of access means that at least 7.5% of the time during the normal working day was taken up with unacceptable activities. It should be noted that this measure is the actual download time not download plus viewing time which would be part of the overall activity. This behaviour breaches existing policy within the University and further exacerbates the position of the staff member.

Some of the commercially available misuse detection tools failed to detect effectively misuses by the users from different linguistic and cultural backgrounds. The software failed to detect and appropriately categorise many Southeast Asian and Asian sites. This flaw appears mainly due to many of the undetected sites containing native language in their URLs. Many of these websites were unclassified by Cyfin and were placed in the Non-Categorised category. On the initial runs of the Cyfin tool the size of the traffic that was in the Non-Categorised category was substantial and warranted further investigation. The author used a method whereby the entire concatenated log file was searched for trigger strings of typically large file format extensions such as AVI, MP3 in the text stream. The subsequent files containing these triggers were then written to a file for checking and re-categorisation by the Cyfin tool which left most of the known English language sites categorised and the non-English sites again uncategorised.

Upon investigation many of the sites were found to be either MP3, download, movie or other illicit sites. This pattern was particularly evident on pornographic web sites that some of the misusers frequently visited because the sites used colloquial terms to access these sites rather than brand or technology specific tagging such as MP3.

Previous study of similar issues by (Hunter, 2000; Nunberg, 2001; Weare & Lin, 2000) have found shortcomings in content filtering technologies' ability to detect inappropriate content with natural language based approaches. The problem of non-English websites uncovered in this case would indicate that there are further difficulties with relying solely on a content filtered approach for your web content management and policy enforcement.

CHAPTER 6 – CASE STUDY 2

6.1 Organisation Description

This organisation is a large state government agency that has a 1500 plus user base that accesses its wide range of information systems. All of the desktop personal computers used by the employees have access to World Wide Web. The desktop PC's all run the Windows suite of family as the operating system ranging from Windows98 through to Windows2000. The PCs configuration was no lower than Pentium II Class machines with 128Mb of RAM and sizable hard drives. The desktop PCs connected to the organisational LAN either through 100Mb Ethernet or redundant 16Mb Token-Ring; the latter being phased out and replaced with Ethernet. The employees are expected to save any documents onto the organisational servers.

The organisation has multiple sites and locations across the state of Western Australia. However, all sites within the Perth Metropolitan area connect to the World Wide Web via WAN links to central proxy services provided on a centrally administered caching proxy server in the Perth Central Business District. These services are provided via a Novell Netware Border Manager Proxy server. The remote sites access the World Wide Web proxy services on the same conduit as other organisational traffic. The bandwidth capabilities of these remote WAN connections vary from 256Kb ISDN circuits to 155MB ATM fibre based connections.

The organisation uses content filtering software in the form of CyberPatrol software (SurfControl, 2002) as this software is optimised for use on the Novell Netware based proxy server. The organisation also has a range of written policy to control inbound content and usage of the network resources.

6.2 Pre-Analysis Interview

An initial open-ended interview was conducted on 12th June 2002 upon engagement with the organisation, with the participants being (in order of authority), the Manager for Information Services, the Information Technology Manager and the Network Manager.

When asked to outline any perceived problems the Network Manager responded that MP3 traffic was of concern. The Network Manager explained as a result of standard software audit functions within the organisation they had recently found one user's PC with over 5GB of MP3 files being stored upon it, which was against established organisational policy. The Network Manager also gave a case of where one of the institutional servers was also running out of disk capacity mainly due to 11GB of MP3 files stored by users on their network home drives which were subsequently removed. The other managers in the room did not really respond to this question and did not see any apparent or obvious problems in Internet usage patterns. All three however, acknowledged that there was no proactive or systematic monitoring of the World Wide Web usage occurring.

In response to a query regarding why they were not proactively monitoring the World Wide Web usage, the Network Manager and IT Services Manager cited inadequate resourcing as an issue in terms of human resource. They also coupled lack of human resource with the high cost of commercially available software for the purpose of monitoring. These Managers felt they would have had difficulties in tracking a lot of the usage due to the sheer volume of traffic and, they mentioned that they used the Cyber Patrol (SurfControl, 2002) content filtering suite to filter traffic. All of the Managers present indicated that they were reasonably confident that problem traffic, which they perceived to be pornography, would not be an issue.

The Network Manager was aware that some users were defeating the content filtering suite and accessing sites using dotted quad IP numbers (e.g., 203.59.227.226) instead of a URL. The users achieved this by keeping ahead of the content filter updates, downloading files with mangled or deceptive extensions and the use of dotted quad IP numbers. When challenged by the Manager of Information Services as to why this information had not filtered up the management chain, the Network Manager proffered that it was not of significant concern.

The organisation has several remote sites that access the main services and servers at the head office located in the CBD. The connection to the head office from these remote sites is provided via TCP/IP based links that are used to transport mission critical institutional data from conventional information systems, as well as Internet traffic. All Managers identified the issue that the access to the organisational systems from the remote sites was perceived to be slower in the afternoon. There had been several problems noted with the institutional helpdesk regarding this matter. The Manager of Information Services particularly noted this is as an area of concern, which needed some attention.

6.3 Document Review

The organisation's main vehicle for dissemination of policy that relates to World Wide Web usage is titled IS Acceptable Usage/Conduct. The document is broken into 3 main sections namely, 'Introduction', 'IS Acceptable Usage/Conduct' and 'Definitions'.

The Introduction section states that the policy document is issued under the Information Security Policy and that this policy should be read in conjunction with IS Acceptable Usage/Conduct Policy. It directs employees to Section 5.4 of the aforementioned Information Security Policy as it relates to Roles and Responsibilities in Information Security. Specifically, it stipulates that the employees need to adhere to and promote the IS Acceptable Usage/Conduct policy.

The IS Acceptable Usage/Conduct has 16 subsections many of which overlap. It is not the intention of the author to review all 16 subsections but only those that pertain to World Wide Web and acceptable usage.

The Page 1 of Section 2.1 places responsibility of breaches of policy back with the employee by clearly stating:

> "Any breach of ------ acceptable use policies shall constitute a security violation. A user shall be held personally accountable for any breach and may be subject to disciplinary action or criminal prosecution in relation to any such breach."

It is of interest to note that at no stage does the policy explain what a breach is. It is also naïve of the organisation to believe that it's statement will be upheld and that it relieves itself of any liability or responsibility for incidents, if the organisation can not show due process and due diligence on it's part for the education, enforcement and oversight of it's policy as outlined by Australian Government Solicitor (Solicitor, 2001), then it can be held liable for an employees action.

Page 1, Section 2.2 Personal Use has contradictory and confusing statements:

> "------ corporate information resources must be generally used for ---- business activities. Incidental personal use of a limited nature is permissible provided that use is not abused as defined by their line manager."

This section's statement that information resources must be "generally used", is problematical as there is not a definitive measure or scale. However, it could be argued the second sentenence makes the line manager the sole arbiter of acceptable usage.

The next relevant section of the policy document is Section 2.4 which outlines activities that are not appropriate when using the systems states "Any user caught producing, accessing, requesting or disseminating pornographic material will be subject to internal discipline, including dismissal, and/or legal prosecution" (p.2). Similarly Section 2.5 of the policy document Offensive Material is equally lucid: "To meet EEO requirements, a user who accesses offensive material (as defined by the EEO Policy), reproduces it or distributes it may have disciplinary or legal action taken against them." However, this statement does not provide linkage or directions to the EEO materials.

In the policy document, Section 2.6 Malicious or Illegal Activities states that "Users shall not undertake malicious or illegal activities". Following this statement there are 8 points outlining types of activity that are regard as malicious or illegal.

Section 2.8 of the policy document deals with Copyright and alerts users to use software in accordance with licensing and copyright restrictions. This section explicitly mentions music, literature, pictures as being materials that also may be subject to copyright.

Section 2.10 Personal Software strictly forbids the use of any software other than which has been provided by the organisation.

Section 2.12 Personal Computer Settings/Problems specifies that "Only authorised IS staff are permitted to adjust settings or install software on ----- Information resources". This combined with Section 2.10 makes any installation of software or modification of the organisational standard operating environment contrary to organisational policy.

Section 2.15.2 Content Filtering relates to e-mail usage and that any attachments are filtered. Nowhere else in the document does it mention content filtering of the World Wide Web traffic.

Section 2.16 deals with Internet Usage/Conduct which is inappropriately named as all references in this section are referring to World Wide Web usage. The section mentions that known inappropriate sites are blocked and that accesses to these sites are monitored. In Section 2.16.3 it states that administrators are not allowed to perform site monitoring of any employee and, the Administrators must seek authorisation before proceeding, or furnish copies of unsanctioned searches/monitoring within 24 hours to management! This type of statement is confusing and problematic and belies one of the fundamental weaknesses of the entire policy document. The weakness is how can non-business or inappropriate usage to be detected if active monitoring of the World Wide Web is not normally permissible.

6.4 Method of Log File Analysis

The log files produced by the Novell BorderManager proxy server were formatted in CLF format and were able to be readily analysed by the analysis tools. The Cyfin analysis tool was used first to categorise traffic and identify any macro-problems with usage using the standard 55 categories. Cyfin was then used to identify the misusers and categories that had anomalous behaviour for investigation. Other log file analysers namely Pwebstats, Analog, and Webalizer were each used individually to confirm the measurements from Cyfin tool, and were used to investigate individual users non-business usage patterns.

The log files examined for this organisation cover the period from 1st January 2002 through to the 16th June 2002 inclusive. The log files produced 27,006,253 unique records that saw a resultant 142,861,451 kB of data served to the desktop computers during this period.

6.5 Overall Results

All of the analysers produced the same generic statistics with 1995 unique users who accessed a traffic volume of 142,847,552 kB and created total log file line entries of 27 million for the period 1st January 2002 until 16th June 2002. Initial analysis was again done with Cyfin as the author was looking for trends and potential macro problems that were occurring with usage. If the organisation's policy was applied, 56% (80.1 Gigabytes) was Unacceptable and 43% (62.7 Gigabytes) was Acceptable traffic.

The usage per user was more homogenous in this organisation with the top 50 users consuming only 41.2%(58.3 Gigabytes) of traffic. If the 2 heaviest users who were the IT staff downloading legitimate service packs and updates are removed this drops to 35.7% (51.1 Gigabytes) of traffic used by the next top 50 users. The top 15 users in this organisation, except the IT staff, consumed 18.8% (26.9 Gigabytes) of traffic which again could be indicative of misuse of the World Wide Web.

The top 15 Cyfin categories accessed by the user are presented in Table 6.1

Categories	Kbytes	%
Banners/Ads	11,905,985	8.3%
News and Media	8,355,024	5.8%
Hardware and Software	8,197,424	5.7%
Sports	8,129,407	5.7%
Government	7,028,047	4.9%
Entertainment	6,596,408	4.6%
Email	6,131,972	4.3%
Internet Services	6,079,527	4.3%
Reference	5,010,800	3.5%
Financial	4,828,394	3.4%
Search Engines	3,118,425	2.2%
Society and Culture	2,764,378	1.9%
Real Estate	2,740,696	1.9%
Download Sites	2,352,960	1.6%
Travel	2,336,626	1.6%

Table 6.1: Usage by Cyfin Categories

The Analog file analyser was used to produce the data in Table 6.2 containing the number of requests per file size range.

Number of requests	Number of requests	Percentage of the total bytes	Avg Size Per Request Kb
1b- 10b	32	0.00%	0.0
11b- 100b	3,128,857	0.12%	0.1
101b- 1kb	9,784,574	3.40%	0.5
1kb- 10kb	8,989,668	20.93%	3.3
10kb-100kb	2,818,809	49.45%	25.1
100kb- 1Mb	66,351	10.32%	222.1
1Mb- 10Mb	5,345	9.63%	2572.8
10Mb-100Mb	282	4.07%	20609.8
100Mb- 1Gb	10	2.08%	297024.0

Table 6.2: File sizes of requests

The distribution of this traffic in Table 6.2 would indicate that much of the traffic is potentially HTML traffic because 73.9% of all objects transmitted are 100 kbytes or less. However, 15.8% of traffic was 1Mb or larger in size with 9.63% being in the 1-10 Mb category with an average of 2572 kb. Only 282 downloads were responsible for 4.07% of the files in 10-100Mb range which means that each file in this category was 20Mb on average in size these figures considerable and could indicate possible misuse of the system by staff downloading patches or large software archives at work.

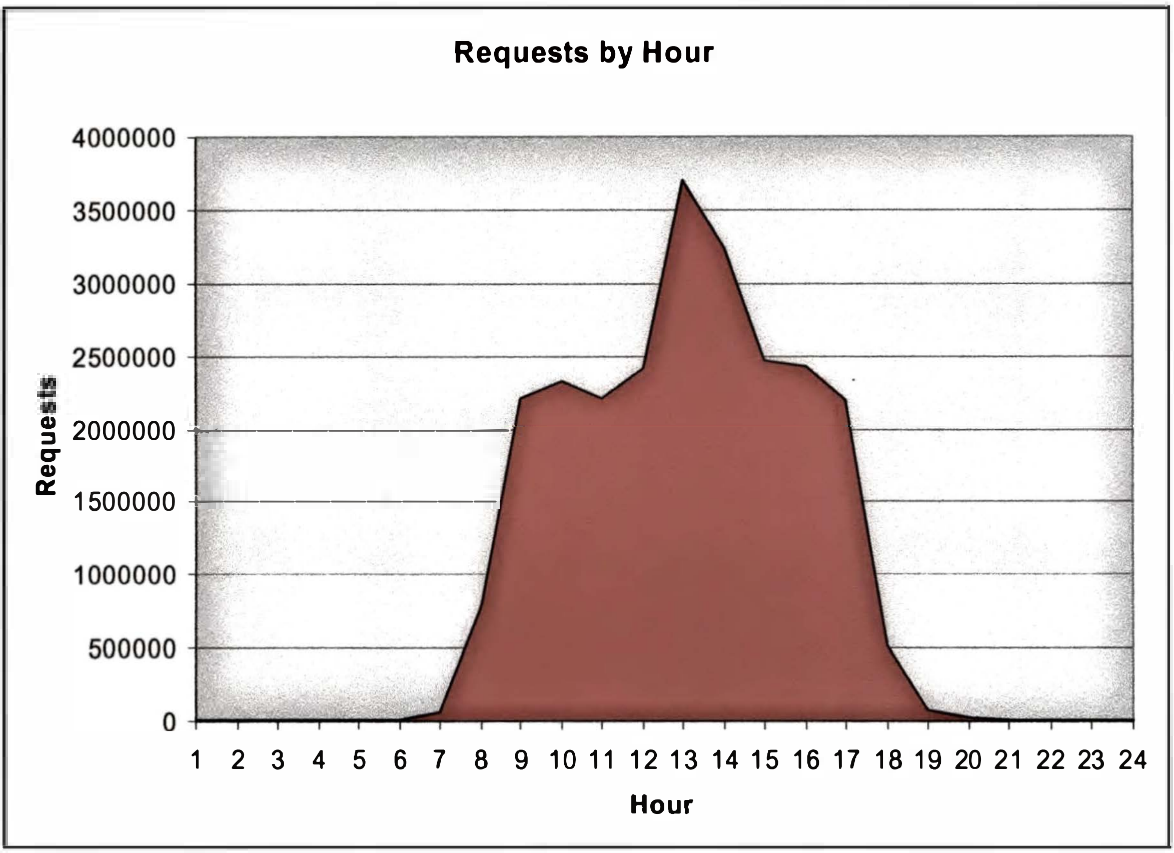


Figure 6.1: Number of Requests to the Proxy Server by the Hour

From Figure 6.1 it can be seen that World Wide Web usage throughout the day indicated increasing use from 8 a.m. peaking between 12 – 2 p.m. This peak coincides with lunch breaks between 12 – 1 p.m. and 1 – 2 p.m. It would be a reasonable expectation that there would be a noticeable reduction in the use of the network during this time as the users are on a scheduled break.

The usage spikes by 53% in the 12 – 1 p.m. timeslot and if you take the 1 – 2 p.m. timeslot and calculate this spike a percentage change from 11 – 12 a.m. timeslot, there is a 34% increase in usage. The hours, outside of the "lunch spike", average within a ± 5% variance of the average which is 2,343,174 requests. The fact that the spike occurs during the lunch hour would point to staff utilising their breaks to conduct non-business activities on the World Wide Web.

This usage pattern concurred with the interview data which indicated the network access from remote sites was slower in the afternoon. The spike is considerable and certainly impacts on the organisations capacity to serve data to remote sites during this time along the corporate LAN.

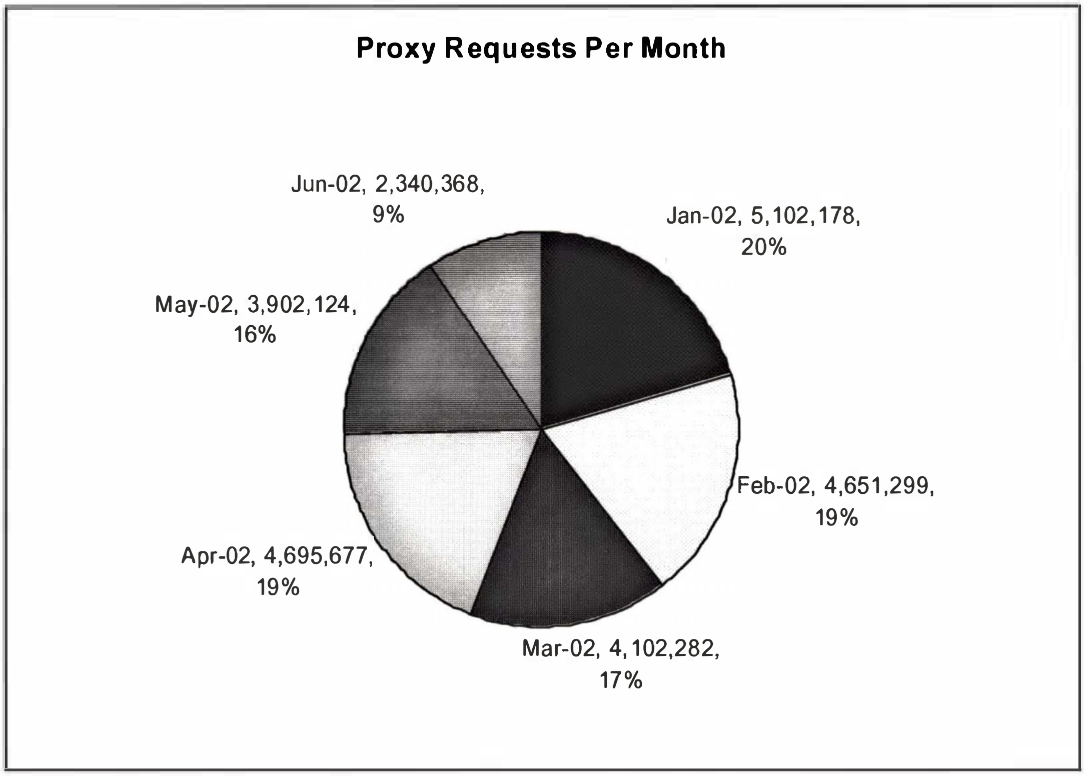


Figure 6.2: Proxy Server Requests Per Month

The range of requests for the 5 full months (January to May) as displayed in Figure 6.2 had an average of 4,490,712 requests per month. The variance in these months, compared to the average, was minimal. This trend would indicate that usage patterns as a whole are fairly well established within the organisation.

6.6 Analysis of Top 15 Categories

This section 6.6 examines the top 15 categories as calculated by the Cyfin log file analysis tool.

6.6.1 Banners/Ads (Unacceptable)

The downloaded content in this category represents 11,909,985 kB of download volume or 8.3% of total bandwidth. This traffic is considerable and all the users on the system generated activity in this category. Due to the typically push based nature of this traffic, much of this content would have been unsolicited advertising and banners from websites that users have accessed during the course of using the World Wide Web.

Upon further investigation this situation was found to be exacerbated by system wide installation of Gator.com software that presents as an innocuous client but is actually Spyware. The purpose of this particular Spyware software is to monitor a person's usage of the Internet. The monitored and collated information about the users' Web surfing patterns is then sent back to the gator.com servers which store them as user profiles. Gator.com may then sell this profile based on the user's accesses and surfing patterns to third parties. These third parties can send targeted banners and advertisements to the users desktop from selected sites via the innocuous software installed on the users PC.

In this way the use of the Gator.com software produces a positive self feedback loop. Due to the nature of the Gator.com client software, this was not really a misuse of the system. However, it was a misunderstanding of the technology or ignorance of the threat that the Gator.com spyware software represented to the organisation before its installation across the network.

6.6.2 News and Media (Acceptable)

In this category 8,336,045kB (5.85%) of content was downloaded by 795 users (39.8%). This category was determined to be acceptable as staff were expected to be aware of current events and public opinion relating to policy that the department was responsible for. The top 20 (2.5%) users accessed 3,871,135kB (46.4%) of the total download for this category. This usage is considerable when taken in the context that the remaining 53.6% of material was downloaded by 775 (97.5%) users. Some of the activities or practices undertaken by the top 20 users need further scrutiny as their activity in the category appears to be abnormal and excessive.

6.6.3 Hardware and Software (Acceptable)

The total downloaded volume for this category by the 1274 users (63.8%) was 8,197,424 kb (5.7%). The number of users accessing this category is high: however, it should be remembered that many programs retrieve updates or at least check vendor sites for updates on a regular basis, often without user knowledge or interactions.

The top 20 users accessed 4,018,072 kB (49%) of all material in this category. Unlike other categories, this usage would not possibly indicate misuse as it would be expected that the network administrators and other relevant information technology staff should be extensively downloading material in this category. However, less than 20 people would be tasked with such a job to perform and, as such, this usage would require further scrutiny.

This category was determined to be acceptable by the organisation as it believed that such access was necessary to facilitate updates of corporate equipment and software. From the authors, IT administrative experience this volume would be average for an organisation of this size and range of operating platforms and is acceptable.

6.6.4 Sports (Unacceptable)

In this category 8,129,407 kB (5.6%) of the total volume was downloaded by 655 (32.8%) users. The top 20 users consumed 4,816,144 kB (59.2%) with the remaining 635 users responsible for 40.8% of the content downloaded. This statistic reflects considerable usage by the top 20 users and warrants further examination.

The top user consumed 1,866,758 kb (22.9%) which is considerable given that it is almost four times the size of the next top user's volume. This fact would indicate a potential habitual misuser of the system when related to sports and would involve extensive downloading or viewing of content from the World Wide Web relating to sports.

The problem was of non-business usage was institutionalised in this category by an external trainer being unaware of the organisational policy, who encouraged the access of cricket sites as part of Internet training exercises. Such training would have given de-facto authorisation to examine crickets scores as a method to reinforce skills acquired in corporate training.

6.6.5 Government (Acceptable)

This category is applicable to the organisation's mission and goals and as such is acceptable. There was 1,023 users (51.2%) were responsible for downloading 7,028,047 kB of traffic in this category. The top 20 users downloaded 2,217,069 kB (31.5%) of material from the World Wide Web. The remaining 1003 users downloaded the remaining 68.5% of the volume. This pattern would indicate homogenous use and as such the category that has not been subject to any noticeable misuse. Many of the accesses recorded were downloading on-line forms for processing or accessing on-line databases.

6.6.6 Entertainment (Unacceptable)

By enforcing organisation policy this category is considered unacceptable However, 935 users accessed materials to produce a total download of 6,596,807 kB. The top 20 users downloaded 2,766,710 kb (42.7%) of all material accessed. This usage is excessive and would indicate sustained misuse by these individuals.

The top user accessed 1,015,706 kB (15.4%) of all material for this category. The next highest user accessed 222,160 kB (3.3%) of content. This clear discrepancy would indicate potential habitual and deliberate misuse of system by the top user.

6.6.7 E-mail (Unacceptable)

There were 826 (41.4%) users who accessed material in this category for a total download of 6,131,972 kB. The top 20 users downloaded 2,408,432 kbytes (39.2%) of material. This usage is significant misuse considering that all users are allocated institutional e-mail addresses for business use. The volume of information that has been transmitted via this mechanism is not opportunistic but, sustained, misuse by these users. What makes these statistics significant is that most web based e-mail systems typically only allow around 1-5 Mb for a users mailbox size. Due to such limitations, the use of e-mail with large attachments is significantly reduced by technical constraints. Of particular note is that fact the top 5 users in this category that have each transmitted over 150Mb of e-mail via these services for the period examined.

These users who are accessing and transmitting e-mail via web based services are doing so as a direct and deliberate violation or avoidance of organisational policy. This practice may pose a significant threat in terms of security for the organisation in two ways. Firstly, the inbound e-mail from these sources will have the desktop PC's virus scanning software as last resort for detecting any malicious code. Secondly, there is no possible control of outbound content in terms of viruses and actual content.

6.6.8 Internet Services (Acceptable)

A total of 1,147 users accessed materials in this category to produce 6,079,527 kB (4.3%) of download. The top 20 users consumed 2,359,067 kB (38.8%) of downloads. The top user consumed 473,018 kB (7.8%) of all volume which was over twice the amount the of the next user at 218,059 kB the latter was identified as an administrator. This usage by the top user was suspicious and warranted further investigation. Upon, which it was determined that the user was accessing e-mail based services through a local ISP. Most of the accesses in this category were users accessing web based e-mail which is noted as unacceptable.

6.6.9 Reference (Acceptable)

In this category there were 1,015 users (50.8%) who accessed and downloaded 5,010,800 kB (3.5%) of material. This access is considered acceptable as staff often need to check definitions or refer to reference materials to verify their work. The distribution of traffic was widespread as the top 50 users only downloaded 1,767,166 kB (35%) which would not indicate misuse.

6.6.10 Financial (Unacceptable)

There were 796 users (39.8%) responsible for downloading 4,828,394 kb of material in this category. The top 10 users consumed 2,338,736 kB (48.4%) of content which is considerable usage, given the fact that the remaining 796 users consumed the other 51.6% of content.

The top 2 users were responsible for 656,466kb (13.5%) and 483,932 kB (10%) respectively. This activity conducted by both top users can only be considered as significant misuse of the resource.

Upon further examination of the usage patterns of the top 10 users in this category their activities would be consistent with persistent on-line share trading during business hours. Furthermore, popular banks and share trading sites featured in the top 10 sites visited by proxy server requests for this organisation. This statistic points to widespread organisational misuse.

6.6.11 Search Engines (Acceptable)

This category is considered acceptable and 1,287 users accessed material that accounted for 3,118,245 kB (2.2%). The top 20 users were responsible for 841,550 kB (26.9%). This usage is considerable given that these are search requests to a search engine. The top 3 users downloaded 93,254 kB, 83,304 kB and 75,875kb respectively, which would appear to be an excessively high amount of searching for materials. This pattern could be a precursor for abusive activity by the users concerned or indicate staff members that need training in effective web searching to lower this high usage.

6.6.12 Society and Culture (Unacceptable)

In this category 554 users (27.7%) accessed material to download 2,764,378 kB (1.9%) of content. The top 20 users consumed 1,483,550 kB (53.6%) of the downloads, the top 5 users alone consumed 657,585 kB (23.7%) of material. The use of this category is deemed unacceptable and the usage by the top 20 users, in particular, would indicate misuse with the top 5 demonstrating sustained and deliberate malpractice of this category.

6.6.13 Real Estate (Acceptable)

This category is considered acceptable as it is related to one of the core business functions of the organisation concerned. There is a small section within the organisation whose function is the monitoring of rental properties and advertisement of same.

The 2,740,696 kB (1.9%) of material was downloaded by 466 users in the organisation. The top 20 users downloaded 1,388,656 kB (50.6%) of material, with the top 5 users downloading 756,688 kB (27.6%). This pattern would indicate that there is possibly some misuse of the system by top users or they are the result of unsolicited e-mail generating this traffic.

6.6.14 Downloads (Unacceptable)

Material in this category was accessed by 236 users (11.8%) for a total download volume of 2,352,960 kB (1.6%). The top twenty users in this category downloaded 2,224,690 kB (94.5%) of all material. The top user was responsible for 1,530,026 kB (65.0%) of all traffic downloaded which is considerable comparative usage. Upon further investigation, this user was not an administrator nor a technician responsible for updating software: this would indicate possible misuse by the user concerned. The volume of material downloaded by the top 20 users would also point to potential misuse of the system.

6.6.15 Travel (Unacceptable)

In this category 654 users (32.7%) accessed content to produce 2,336,626 kB (1.6%) of downloaded material. The top user downloaded 392,984 kB (16.8%) and should be considered a misuser. The top 20 users consumed 1,016,387 kB (43.3%), while the top 5 users consumed 643,349 kB (27.5%), these figures should be considered as consistent misuse. The 604 users outside the top 50 consumed only 927,838 kB (39.7%) which could easily be accounted for as a result of unsolicited e-mail traffic.

6.7 Analysis of Top Users

Overall, any use of the World Wide Web within this organisation is relatively homogeneous in terms of downloads volumes for individual users. The examined cases in this organisation demonstrated that the misusers were typically abnormal singular standout cases as opposed to a small nucleus of habitual misusers. The identified malpractices in this organisation also tended to have singular themes of misuse, for instance, previewing a holiday destination or downloading bandwidth intensive files.

6.7.1 Misuser 1

The highest user of the bandwidth excluding the 2 previously identified IT staff was the user who consumed 2.64 Gigabytes (Gb) of volume. Of which 1.53 Gb (58%) was consumed in March 2002 with this content being the Mandrake Linux 8.2 ISO CDROM images for installation. The remaining bandwidth was primarily consumed via the download of multimedia streaming files such as MP3, MOV, and AVI (13%) files. Over 90% of the user's activity was against established policy within the organisation.

The size of the three Mandrake Linux ISO files at 650Mb each would indicate that user either had access to CD-ROM burning equipment or some other form of portable mass storage device. This displayed behaviour in particular would indicate serious and intentional misuse of the system.

6.7.2 Misuser 2

This user primarily accessed motor racing sites and some Sony Playstation related games sites, and had a volume usage of 2.01 Gb for the period examined. The access involved the regular viewing of the sites amounting to 10 - 20 visits a day. This user's established behaviour pattern was the downloading of various product trailers and video clips from the World Wide Web related to personal interests. The content of these files was either demonstrating the latest games available for the Sony Playstation gaming console or live footage of automotive racing related activities.

6.7.3 Misuser 3

This user consumed 1.6 Gb of content as a result of their activity on the World Wide Web. Over 85% of the download volume was dedicated to the preview of sites relating to a holiday in Bali. A wide range of sites were visited all which related to various resorts and tourist facilities in Bali. This activity occurred either during scheduled lunch break or late in the afternoon.

6.7.4 Misuser 4

This user consumed 1.88 Gb of traffic for the period covered. Of the requested traffic 60% was consumed in the hours of 12:00 a.m. - 1:00 p.m. hours which is the traditional lunchtime slot for the organisation.

This user appeared to have been downloading World Wide Web access logs for an external website related to Formula 1 car racing. The user had 24.6% of his/her total usage resulting from accessing files that were 100kB or larger, 51.19% of their total usage involved 10-100kB size files. Such files were either extensive written reports or a large graphics files. This activity is suspect and could possibly indicate that this user is managing an external web site during business hours.

6.7.5 Misuser 5

The eighth highest user of bandwidth for this organisation did so only in the first 2 months of the six examined. The staff member's bandwidth consumed stopped on Friday 22nd February 2002 it was also later confirmed at the post analysis interview that this person had left the organisation. The total downloaded by the user was 1.50 Gb.

What is abnormal about the traffic usage is that it rapidly accelerated in the last four weeks of the person's engagement with the organisation. This period is significant in that you must give exactly four weeks notice of intention to leave the organisation's employ. In the notice period the employee appeared to ignore the policy countermeasures such as termination of employment and disciplinary action.

Upon further analysis, over 90% of the traffic downloads were either full programs or updates to various shareware or freeware offerings for the Windows platform. Prior to this period of usage the normal pattern for this user indicated usage below 100 MB per calendar month. The increased activity can reasonably be attributed to either boredom or targeted malice by the user after acceptance of the resignation.

6.7.6 Misuser 6

This user consumed 1,576,037 kB of traffic during the period examined. This user's Web surfing habits can only be described as erratic in nature. The user visited various sites with no apparent set pattern of a daily usage other than most of the activities were conducted in the latter half of the working afternoon.

Many of the non-business sites visited were personal user sites and personal contact page websites. Other content related to this user was varied and would indicate casual World Wide Web browsing activity by the user.

6.7.7 Misuser 7

The period examined for this user covers from 25th February, 2002 until June 16th, 2002. During this period the user downloaded 1,406,433 kB of traffic. The user's highest period of usage is the 8 a.m. to 9 a.m. timeslot where the user on average accesses 25% of his content downloaded for any given day. The user is a heavy user of the Hotmail e-mail service. He/She also extensively accessed sites relating to the TV program Big Brother and other Web camera based sites. Accessing web camera based sites follows no pattern other than they are web camera based sites. The range of topics and activities covered by the user these types of sites is extensive covering Alaskan weather cameras to Zimbabwe roof mounted cameras looking out across a city.

The remainder of the activity is related to MP3 content whether they be MP3 song files, utilities or programs that are related to the viewing and display of such materials. Over 90% of the activity by this user is categorised as unacceptable by the organisation. The user's page views per day peaked in June 2002 with a 519 page per day output which is the highest of any of the misusers in this organisation.

6.7.8 Misuser 8

For the period examined this user consumed 1,809,816 kB of traffic. Most of the user activity was accessing travel and holiday related websites. The user also heavily accessed news and sports sites during a days viewing. The range of sports sites accessed was extensive, and included but was not limited to soccer, rugby, wrestling, gambling and football sites, this activity which would indicate an extraordinary interest with sports related activities.

The usage by this user increased during the lunch break (approximately 50%). The consistent nature of sites visited would indicate habitual misuse of the system by this user. A daily average of pages viewed per day peaked at 319 pages in the month of June 2002 which is considerable traffic. Very little activity by this user was actually work-related.

6.7.9 Misuser 9

This user consumed 1,088,607 kb of traffic for the period 1st January, 2002 until May 2nd, 2002 when records for this user stopped. Much of the traffic at first examination would appear to be legitimate and work related as the user accessed a wide range of hardware vendor's sites and checking prices. However, upon further investigation there was no user whose job function would fit this kind of activity.

With closer examination of the logfile entries for this user it could be reasonably concluded that the user was part of a computer sub-culture called "tweaking". Adherents of "tweaking" people push existing hardware beyond specified limits by a technique referred to as overclocking where they make CPUs perform beyond specification by altering internal or external bus speeds. "Tweaker" sites were accessed everyday by the user, subsequently undertaking checking of computer mainboard and CPU prices at various local hardware retailers websites. The mainboards examined or sought by the user were types that were capable of, or designed specifically for, overclocking.

6.7.10 Misuser 10

This user consumed 815,895 kB of traffic for the period examined. The user made extensive use of web based e-mail during this period, downloading 141,957 kB (17.3%) of material. This usage is considerable and, by the frequency of requests to the web based e-mail sites, would indicate that the web mail is checked at frequent intervals, or left on the desktop permanently polling for new mail.

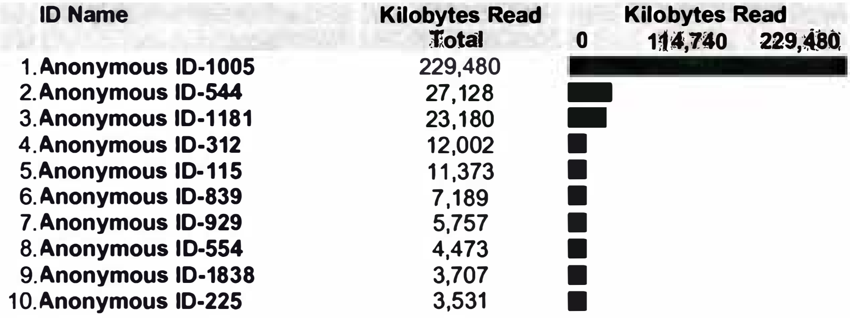
The user also downloaded 129,645 kB (15.89%) of MPEG and AVI movie files. These files were predominantly automotive displays of burnouts and drifting which are techniques that people involved in performance car culture use to display the power of their heavily modified and performance enhanced vehicles. Some of the files were requested more than once by the user and, due to their large size (over 4 Mb) they would possibly not have been cached by the proxy server caching engine. The remaining sites visited were automotive related namely: personal web sites featuring vehicles, car specific magazine sites, or automotive trading companies advertising vehicles and spare parts.

No activity examined would have been related to work by this user. The users browsing activity also peaked significantly during the 12 p.m. – 1 p.m. lunch hour.

6.7.11 Other Cases of Misuse

This section will outline other users who may not have been misusers in terms of bandwidth usage but their behaviour is abnormal or differing largely from established organisational patterns of usage.

In the category of User News Groups there was only 350,049 kB downloaded from a total of 183 users. In this category the top user downloaded 229,480 kB (65%) of material. (See Figure 6.3)

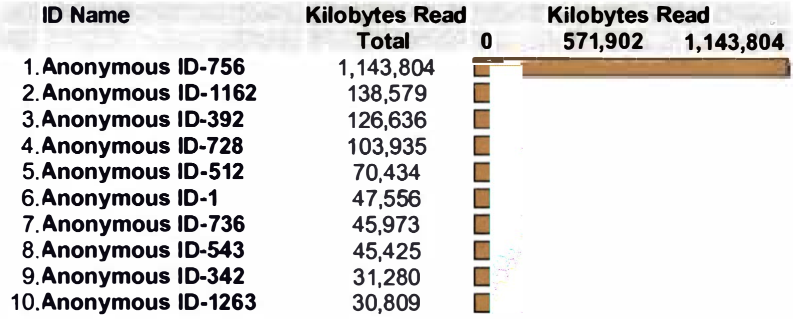


ID Name	Kilobytes Read Total	Kilobytes Read 0 114,740 229,480
1. Anonymous ID-1005	229,480	
2. Anonymous ID-544	27,128	
3. Anonymous ID-1181	23,180	
4. Anonymous ID-312	12,002	
5. Anonymous ID-115	11,373	
6. Anonymous ID-839	7,189	
7. Anonymous ID-929	5,757	
8. Anonymous ID-554	4,473	
9. Anonymous ID-1838	3,707	
10. Anonymous ID-225	3,531	

Figure 6.3: News Group Usage

This would indicate a misuse of the resource by the top user as it is against established policy. As can be clearly seen in Figure 6.3 the next nearest user is almost 1/10 of the usage of the top user.

In Figure 6.4 the Auction (Unacceptable) category had 336 users who accessed material for a total of 2,315,476 kB. The top user accessed 1,143,804 kB (49.3%) of the content downloaded and the 2nd user accessed 138,579 kB of material.

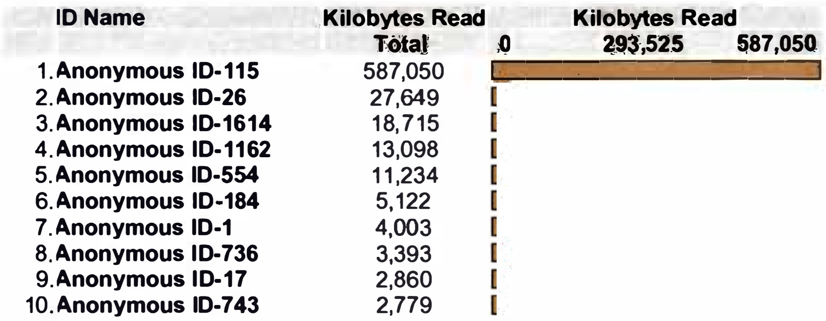


ID Name	Kilobytes Read Total	Kilobytes Read 0 571,902 1,143,804
1. Anonymous ID-756	1,143,804	
2. Anonymous ID-1162	138,579	
3. Anonymous ID-392	126,636	
4. Anonymous ID-728	103,935	
5. Anonymous ID-512	70,434	
6. Anonymous ID-1	47,556	
7. Anonymous ID-736	45,973	
8. Anonymous ID-543	45,425	
9. Anonymous ID-342	31,280	
10. Anonymous ID-1263	30,809	

Figure 6.4: Auction Category Usage

This pattern indicates significant misuse by the top user. The 2nd through to the 5th users should also be investigated while the bandwidth they have used is not excessive it is still markedly larger than that of remaining users.

In Figure 6.5, the Chat (Unacceptable) category 176 users downloaded 719,047 kB of material. The top user downloaded 587,050 kB (81.6%) of this content which is 20 times more than the next misuser.

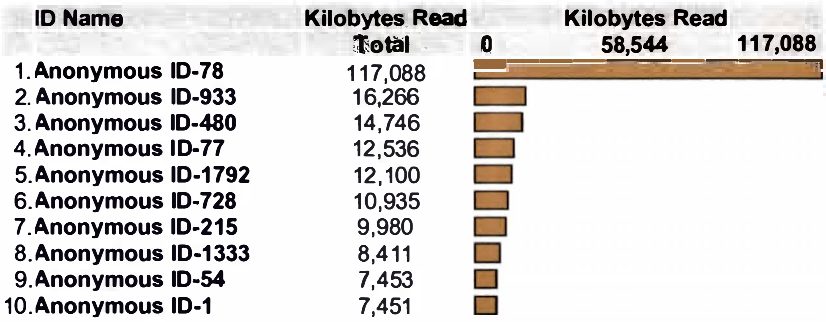


ID Name	Kilobytes Read Total
1. Anonymous ID-115	587,050
2. Anonymous ID-26	27,649
3. Anonymous ID-1614	18,715
4. Anonymous ID-1162	13,098
5. Anonymous ID-554	11,234
6. Anonymous ID-184	5,122
7. Anonymous ID-1	4,003
8. Anonymous ID-736	3,393
9. Anonymous ID-17	2,860
10. Anonymous ID-743	2,779

Figure 6.5: Chat Category Usage

This pattern indicates significant misuse by the top user in terms of bandwidth, and of time lost as a result of attending to chat based activity.

In the Personals and Dating (Unacceptable) category (see Figure 6.6) 176 users downloaded 284,257 kB of material. The top user downloaded 117,088 kB (41%) of material for this category.



ID Name	Kilobytes Read Total
1. Anonymous ID-78	117,088
2. Anonymous ID-933	16,266
3. Anonymous ID-480	14,746
4. Anonymous ID-77	12,536
5. Anonymous ID-1792	12,100
6. Anonymous ID-728	10,935
7. Anonymous ID-215	9,980
8. Anonymous ID-1333	8,411
9. Anonymous ID-54	7,453
10. Anonymous ID-1	7,451

Figure 6.6: Personals and Dating Category Usage

The statistic generated from top user's access is comparatively excessive and warrants investigation.

6.8 Post-Analysis Interview

A copy of the log file analysis was sent to the key stakeholders in the organisation 2 days prior to the scheduled presentation. The presentation and subsequent open interview was conducted with the Manager for Information Services, Information Technology Manager, Network Manager and a newly appointed Manager for Information Security. After the conclusion of the presentation, when asked about the level of misuse, all of the respondents displayed surprise at the breadth and depth of non-business usage of the World Wide Web.

During presentation of examples of users evading the current content filtering system by doing forward reconnaissance and intelligence gathering elsewhere outside of the work environment discussion ensued. The Information Technology Manager outlined a procedure that they were to put in place that would allow the IT Staff to audit server based drives for the removal of MP3, AVI and other non-essential materials. The Manager for Information Services pointed out that this would not solve the problem of misuse and could possibly be delete legitimate in-house multimedia work of users from server drives. The Network Manager pointed out that a MIME sweeper was already in use on the e-mail system to remove attachments of MP3 and other inappropriate files and extension of a similar technique onto the proxy server was possible. The Network Manager also raised the possibility of using the features of Cyber Patrol to block specific sites and admitted that they were unsure if this technique was being utilised at present and would have to confirm.

The problem of advertising banners and their possible blockage was discussed. The Manager for Information Security outlined how commonly available tools for filtering worked and that some sites would stop sending content if blocking software was detected. This response by the identified sites was not a seen as a major problem as it was agreed that these sites were of no real consequence to the core functioning of the organisation.

There was some discussion on the use of Hotmail and external mail accounts. The Manager of Information Services pointed out that he had reviewed the access policy and said that, under their existing policy it did not allow the use of Hotmail and web based e-mail. The Manager of Information Services stated that this was a problem and needed redress to secure the organisation against any possible liability. The Network Manager raised the problem of other fungible sites that could be used if Hotmail was blocked. The Manager of Information Services then said that strict enforcement of a policy that banning the use of external web based e-mail services was the only way to reduce the potential threat.

Overall the managers were pleased that pornography at least, was essentially a non-issue mainly due to the effectiveness of the content filter. The discussion came back to the high level of MP3 and other streaming content that was being accessed and, surprise was expressed at the level of such usage and at the steps that some users were using to defeat the content filtering system.

6.9 Discussion of the Case Results

The organisation has an established pattern of significant misuse of the World Wide Web during the established lunch hour, with the two allocated lunch hours seeing a peak over 50% in normal established patterns as is easily seen in (Figure 6.1). This pattern of use is an issue that needs further investigation as this behaviour is affecting the performance of the organisational WAN links to remote sites. This lack of performance not only affects the ability of these connections to serve data, but would also have a significant impact on worker productivity at the remote sites due to the reduced network response times. This issue would also extend beyond the traditional lunch hours as the network would still be servicing requests made within the lunch hour.

The incidence of pornographic material accessed within the organisation was practically non-existent at 227,965 kB or 0.15% of all downloaded material. There was a definite lack of any overt policy with regard to the access of the Internet; that is, it was not on display in any workspaces and was provided only as a static document on an Intranet server. Even though there was a general lack of policy awareness, it was well known within the organisation that the Internet connection was content filtered by CyberPatrol(SurfControl, 2002), thus leading to a low access of pornography from within the organisation. However, this did not stop determined misusers working around the content filtering restrictions that were in place. Some of the misusers employed IP numbers of sites to get around the system or used non-blocked URLs of sites that were gathered from external intelligence activities. This behaviour was demonstrated by several users who were examined in this case as their first signs of daily activity was the use of IP numbers or long URL's to access illicit material. Many of these illicit sites were unreachable when the author tried to further verify the sources. This would indicate that the sites were "zero day" sites or hacker sites that typically move around to avoid prosecution or detection by authorities or ISPs. Much of the recent literature in this area points to users leveraging forward intelligence and communities of practice to acquire URLs and content locations as becoming an emergent problem in organisations.

The proxy server that was run by the organisation logged all users' accesses of material on the World Wide Web against their network identity and not their actual IP or workstation address. This logging means that users regardless of where they login and access the corporate World Wide Web proxy server, did have their activity automatically recorded against their network identity. This technique if leveraged makes it far easier from an administrative point of view to track down the details of users who have been misusing the system. This level of logging that was being maintained on the system indicated that the organisation was not capitalising on its intelligence function by allowing the proactive monitoring or analysis of the World Wide Web activity for inappropriate usage. What further exacerbated this situation is the flagitious fact that the practice of non-systemic monitoring is embodied in the organisational usage policy in Section 2.16.3 Site Monitoring. It also indicates that the organisation was placing great faith in the content filtering software and thereby suffering from magic bullet syndrome (Markus & Benjamin, 1997) hoping that the use of this software was going to instill a magical transformation of end-user behaviour. It could be argued that software is being used as a substitute for effective management in this area.

The types and extent of behaviours uncovered in this case could be brought about by the poor execution of effective policy within the organisation. The organisational policy relating to the World Wide Web namely the IS Acceptable Usage/Conduct policy was long, verbose, ambiguous and repetitive. These attributes are not hallmarks of effective policy and not supported in much of the exemplar literature in this area(AMA, 2001; FPC, 2002; Kallman, 1993; Lichtenstein & Swatman, 1997; O'Brien, 1999; Solicitor, 2001). The policy document generally lacked cohesiveness and adequate, unambiguous definition of key terms and concepts which are essential elements in defining policies of this nature (FPC, 2002). The policy did not demarcate key terms such as breach nor outline a path of escalation, authority or responsibility for any breaches. The policy did not sufficiently empower administrators of systems to pursue proactive monitoring of the World Wide Web traffic, nor did it prescribe any audit functionality on the behalf of management which is a critical function when monitoring for compliance (FPC, 2002), Guideline 5).

The pernicious paragraph for the whole IS Acceptable Usage/Conduct policy is Section 2.16.3 Site Monitoring which reads:

> "System administrators and other authorised users with access to monitor staff Internet usage shall receive management approval prior to analysing Internet traffic of other users. If management approval is not immediately available, system administrators and other authorised users who analyse Internet usage shall document their actions and provide that documentation to management within twenty-four (24) hours." (p.5)

This section restated implies "we do not proactively monitor you for inappropriate usage". Logically, it would be reasonable to deduce that as long as users are not being overt in their misuse, they will not be detected due to a lack of system wide monitoring of World Wide Web activity. For all intents and purposes some users had already worked this out in this organisation and were enjoying extensive misuse of the World Wide Web within the organisation.

CHAPTER 7 – CASE STUDY 3

7.1 Organisation Description

This organisation is a small Western Australian state government department with 300 users statewide. The organisation works within the government's legal sphere, and handles material of a highly sensitive nature that relates to individuals, corporate bodies and the state itself. The users have access to the Internet through their desktop PC's. The system was using Microsoft Proxy 2 server to deliver World Wide Web content to users. The server was configured to use seamless user based authentication mechanisms, which transparently logged the username with all activity generated by the user whilst accessing the World Wide Web. Client computers had been configured with a Standard Operating Environment (SOE) based on the Microsoft Windows 2000 operating platform.

7.2 Pre-Analysis Interviews

An open-ended interview was conducted with the IT Services Manager and the Network Systems Manager. The organisation was in a transition from using outsourced vendors for most of its information systems support functions to an insourced approach. The IT Services Manager had started work with the organisation at the beginning of that particular calendar year. The IT Services Manager quickly implemented IT policies that he had used previously in similar organisations and he was unaware of any breaches of the newly implemented World Wide Web policy.

The IT Service Manager reviewed the log files on a monthly basis to see if there was any standout issues that warranted investigation. He has stated that he was almost dumbfounded by the fact that no suspect URLs or other errant behaviour was detected in these monthly checkups. The IT Services Manager indicated that the Internet usage compared to similar organisations in which he had worked was small overall, with five to seven gigabytes per month downloaded.

The Network Systems Manager had recently joined the organisation and was keen to find out if there were any issues that needed resolution, relating to the use of the World Wide Web in the organisation. He was also surprised by the low volume of Internet traffic coming into the organisation when compared to similar organisations in which he had previously worked.

When asked as to why they believed the Internet usage and in particular the World Wide Web usage was low, both believed it was to do with clear and enforced policy. There was no content filtering or other restrictions in place on the organisational World Wide Web proxy system.

7.3 Document Review

The organisation had extensive policy documents that related to network usage and, more specifically, Internet usage in the work place. The document relating to the area of interest to the researcher is simply titled ‘Internet Access’. The document was stored on an internal Intranet server, that used the facility of hyperlinking to connect to other relevant documents. For example, in the Internet Access policy there was a link to the organisational Discipline Policy.

The document outlined the expectations for the correct use of the Internet within the organisation. The document has two main sections on policy and procedures.

The policy section states:

“

1. Employees have access to the Internet as a business tool to assist them with the completion of their duties.
2. Internet access will be available to all ------- staff, subject to the relevant manager's approval and available capacity.
3. All authorised users and employees are responsible for proper use of the Internet and accountable for any actions taken using their login ID.

4. The information management actual manages the provision of Internet services. The service will be subject to monitoring.
5. Disciplinary or legal action will be taken if the conditions of access stated in the procedures have been infringed. Refer Discipline Policy."

These five points outline to the user/employee their obligations under the policy. Several sub-sections within these five points are hyperlinked to other documents for further clarification for the user.

The procedures section of the policy deals with Access Management, Applying for Internet Access, Transfer and Downloading of Information, Inappropriate use, Internet Security, and References and Authorities.

- The section 'Access Management' outlines the path of responsibility for the access controls relating to the use of the Internet in the organisation. Under this section it is identified that all Internet access is logged and monitored and, any breaches of policy by employees would result in their details being forwarded to the relevant business manager for resolution. If the breaches were of a serious enough nature, they would be forwarded to the relevant authorities.

- The section 'Applying for Internet Access' outlines the procedures for staff to undertake to apply for access to the Internet. This policy requires the staff member to gain their manager's approval in the form of a signature, initially. Then the staff member was to complete and sign an Internet access request form and send it to the information management branch for processing.

- The section 'Transfer and Downloading of Information' outlines that users should not download material that is copyright or prohibited. The only other point in this section directs staff to avoid network congestion and, where possible, to download large files after business hours.

- The 'Inappropriate Use' section clearly outlines what 'inappropriate use is'. This section refers to relevant legislation, where possible, and explicitly states examples of inappropriate use of the Internet such as the propagation of hate literature.

- The 'Internet Security' section outlines that staff should not attempt to subvert any restrictions that are placed upon them as a result of security counter measures. It also, stipulates that any machine connecting to Internet services must have virus scanning software installed and active.

- The final section called 'References and Authorities' contains the date of the last review and the next review, it nominates persons responsible for the policy management. It also includes hyperlinks to any relevant documents to which this particular policy refers.

The overall policy is clear, succinct and is written, where possible, in plain English. These stated attributes are those that the literature indicates as desirable for good policy formulation. The policy was overt and was displayed at several key points throughout the offices.

7.4 Method of Log File Analysis

The log files analysed in this case are Microsoft Proxy 2 proxy cache server log files (for full details of format see Appendix A – Microsoft Proxy Logfile Format) . The period examined for this case study was from the January 1^{st} to August 31^{st} 2002. During this period 309 users generated 33,445,686 kB of traffic from 11,360,801 log file entries.

The log files were analysed by 4 different log file analysis tools these were Cyfin(Wavecrest, 2002), Webalizer (Barrett, 2002), pwebstats (Gleeson, 2002) and Analog (Turner, 2003).

The process of allocating categories in the Cyfin tool was replicated from the previous two cases. Webalizer, Analog and pwebstats do not handle Microsoft Proxy Server formats natively and the log files had to be converted into Common Log file Format (Consortium, 1995) using a perl based script. These tools were to confirm general statistics generated by the other log file processors on a user and overall basis.

7.5 Overall Results

The period examined for this case study was from the 1st January to 31st August 2002. During this period 309 users generated 33,445,686 kB of traffic from 11,360,801 proxy server log file entries. Initial analysis was done with Cyfin as the author was again looking for trends and potential macro problems with usages of the World Wide Web. When the organisational policy was applied 79.8% (25.5 Gb) was Acceptable and 21.2%(6.42 Gb) was Unacceptable.

The usage per user has a broad distribution with the top 50 users (16.1%) using 58.8% or 19,698,330 kB of all downloaded material, at an average of 393,966 kB per user. The top 15 users (4.8%) used 9,254,772 kB (27.6%) of traffic at an average of 616,984 kB each over the period examined.

The top 15 categories identified by the Cyfin tool are listed in Table 7.1.

Category	%	Kbytes
Email (Unacceptable)	8%	2,863,273
Hardware and Software (Acceptable)	7%	2,419,729
News and Media (Acceptable)	5%	1,993,839
Banners/Ads (Unacceptable)	5%	1,786,960
Legal (Acceptable)	5%	1,755,153
Financial (Unacceptable)	4%	1,661,395
Government (Acceptable)	4%	1,400,353
Entertainment (Unacceptable)	3%	1,181,677
Internet Services (Acceptable)	3%	1,124,131
Search Engines (Acceptable)	3%	1,107,190
Reference (Acceptable)	2%	734,168
Education (Acceptable)	2%	718,419
Travel (Unacceptable)	2%	697,756
Shopping (Unacceptable)	2%	696,642
Sports (Unacceptable)	2%	675,243

Table 7.1: Usage by Categories by descending level of usage

From Table 7.1 it could be seen that the use of e-mail is a potential misuse candidate for this organisation. However, it should be noted that the 2,863,273 kB may seem minor but, given an examination period of 8 months it still could represent a high level of misuse. This is due to the facts that e-mails are often textual in nature comprised and, the fact that many free services such as Hotmail and Yahoo limits the mailbox size of the user to 1 Mb thereby making the use of file attachments impractical.

Size of requests	Number of requests	Percentage of the total bytes	Average size of File Kbytes
0	567938	0	0
1B- 10B	700	0	0
11B- 100B	1375127	0.38%	0.046
101B- 1kB	2937970	3.56%	0.384
1kB- 10kB	1734507	18.05%	3.30
10kB-100kB	721105	52.73%	23.2
100kB- 1MB	14198	9.40%	210
1MB- 10MB	811	6.08%	2378
10MB-100MB	74	4.91%	21048
100MB- 1GB	11	4.89%	147737

Table 7.2: Files Sizes

In Table 7.2 the files sizes of the organisation would reflect practice that would be consistent with downloading standard web pages and files that contained small documents. If the requests for files are examined, the majority of requests have occurred on objects less than 100Kb in size which would concur with the assumption that small downloads are taking place. There are 811 files in the 1Mb-10Mb file size range that accounted for 6.08% of all volume downloaded and, these files were predominately Microsoft Word or PDF documents indicative of employee's accessing work related documents.

Analysis of the hourly usage for this organisation in Figure 7.1 sees a pattern of usage that increases with the period 11 a.m. to 3 p.m. seeing the most usage. The average was 1,637,866 hits with a maximum variance of only of ± 6.7%. This small variance in traffic pattern would indicate that there is pertinent business usage within the organisation. It could also indicate well established works patterns by the employees.

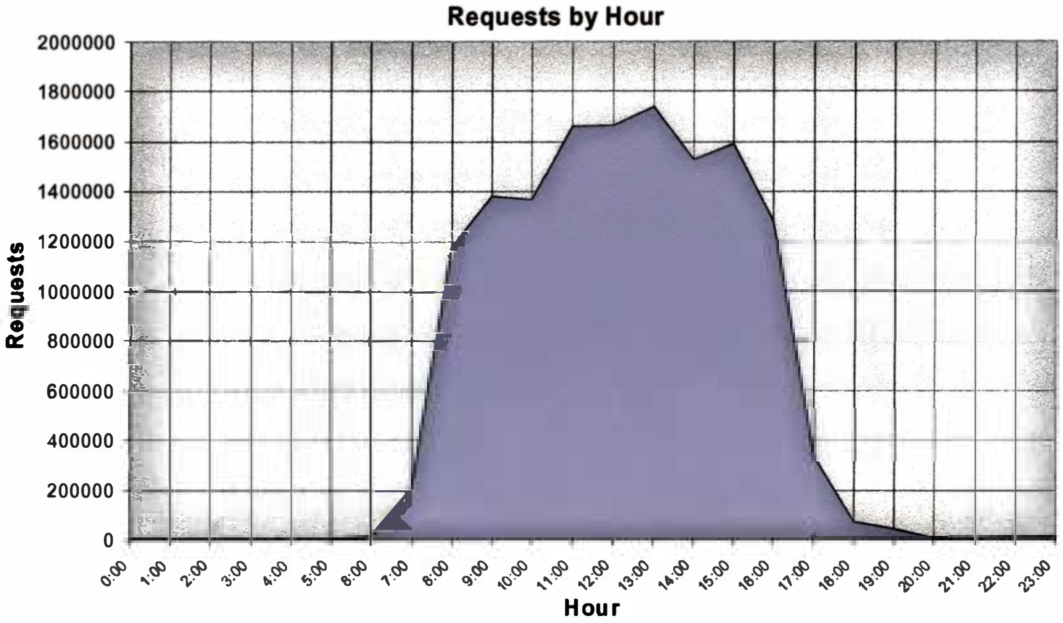


Figure 7.1: Usage by the Hour

The organisation's usage pattern displays a high level of homogeneity and has no large resourcing or spikes. The homogeneity of the pattern would indicate that the organisational World Wide Web resource was not used for recreational purposes during a lunch hour or similar break period.

In Figure 7.2 the requests per month have a large variance with the lowest month being that of March 2002 with 1,204,136 requests or 3896 requests per user and the month of July being the maximum of 2,171,142 requests of 7026 requests per user. This statistic is a large variance in the amount of usage and would warrant investigation normally. The monthly based statistics however, did not show marked changes in usage patterns, except for an increased usage of acceptable or work related sites. This type of variance could be indicative of seasonal or case based workloads.

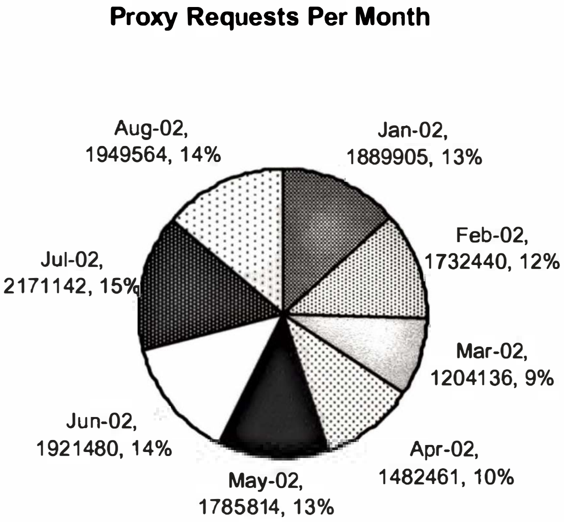


Figure 7.2: Requests per month

7.6 Analysis of Top 15 Categories

This section analyses the top 15 categories of content as calculated by the Cyfin log file analysis tool.

7.6.1 E-mail (Unacceptable)

In this category 201 users (65.0%) accessed material consuming 2,863,273 kB (8.5%) of traffic using unauthorised web based e-mail systems. The top 20 users consumed 2,129,626 kB (74.4%) of the total download. The top 2 users consumed over 22% or 648,900 kB of bandwidth which is considerable and should warrant investigation.

This usage is excessive when taken within the context that all users of the system have access to an organisationally provided e-mail account. An identifiable risk is the nature of the material, to which this organisation is privy, combined with large volumes of e-mail generated. Some employees' considerable usage of external e-mail systems would indicate productivity issues, as they would be frequently engaged in responding to non-business related e-mail.

7.6.2 Hardware and Software (Acceptable)

In this category 289 users (93.5%) accessed material and downloaded 2,419,729 kB (7.2%) of material from the World Wide Web. The top 3 users were IS/IT Support Staff who downloaded 1,480,984 kB (61.2%) of vendor patches and updates for existing systems. The remaining 286 users generated only 938,745 kB, for an average of 3,282 kB each, which taken in the context of updates and patches would be considered average traffic for the period examined.

7.6.3 News and Media (Acceptable)

This category had 213 users who generated 1,993,839 kB (6.0%) of download. The top 20 users consumed 1,515,793 kB (76%) of the total downloaded volume. It is critical that the executive and middle management are aware of issues as reported in the news and media that directly relate to the conduct of the organisation and other changes to the framework of government in which the organisation functions.

7.6.4 Banners/Ads (Unacceptable)

The total downloaded content for all users was 1,786,960 Kb (5.3%) this statistic is something that warrants possible attention. There is the obvious direct cost of bandwidth provision for the banners and subsequent storage of cacheable items. Other less tangible costs are lost opportunity or productivity costs, as downloading of web pages with banners and ads congest the system and affects performance markedly. Many of the data objects associated with advertisements are also not normally cacheable by a proxy server. If objects are not cacheable, then the page is downloaded using raw bandwidth and this action directly impacts on network performance.

7.6.5 Legal (Acceptable)

As the organisation's main function is in the legal field this category is non contentious in its usage. There were 286 users (92.5%) who accessed the materials contained within this category, with the top 20 users consuming 842,371 kB (47.9%) of the material accessed.

7.6.6 Financial (Unacceptable)

In this category 203 users or 65.6% accessed services to generate 1,661,395 kB of material. The top 20 users consumed 763,813 kB (45.9%) of material. To varying degrees the top 20 users were involved in on-line share trading during business hours. This activity would indicate a deliberate misuse of the system by the users concerned as it is against established policy.

7.6.7 Government (Acceptable)

This category is vital for this organisation to function and such activity is expected. There were 261 users (84.4%) who accessed and downloaded 1,400,353 kB (4.2%) of material from the World Wide Web. The top 20 users were responsible for accessing 598,794 kB (42.7%) of the material.

7.6.8 Entertainment (Unacceptable)

This category saw 240 users (77.6%) accessing 1,181,677 kB (3.5%) of material from the World Wide Web. The top 20 users in this category accessed 766,841 kb (64.8%) of content. The top 10 users were responsible for 579,932 kB (49.0%) of the content. This trend would indicate that the top 10 (and it could be reasonably argued that the top 20) users usage patterns are a deliberate misuse of the system: in particular, the top user consumed 131,793 kB (11.1%) of material.

7.6.9 Internet Services (Acceptable)

In this category 291 users (94.1%) downloaded a total of 1,124,131 kB of material from the proxy server. The top 20 users downloaded 494,020 kB (43.9%) of the content accessed.

The top 5 users had total usages of 166,888 kB (14.8%) which would warrant investigation due to the high comparative volumes.

7.6.10 Search Engines (Acceptable)

This category is non-contentious in usage within the organisation. In total 290 users (93.8%) accessed material within this category producing 1,107,190 kB (3.3%) of traffic.

7.6.11 Reference (Acceptable)

Due to the nature of the organisation this category is not contentious as many of the people in the organisation are dealing with different cases and need access to reference based materials. The employees access reference material to become knowledgeable within a particular area or point of law. A total of 196 users (63.4%) accessed content to download 734,168 kB (2.2%) of material. The top 20 users consumed 356,252 kB (48.5%) of the content accessed.

7.6.12 Education (Acceptable)

This category was considered acceptable as the organisations overall policy believed in extending the knowledge base of the organisations staff. There were 212 users (68.6%) who downloaded 718,419 kB (2.1%) of material from the proxy server.

7.6.13 Travel (Unacceptable)

In this category 185 users (61.1%) accessed 697,756 kB (2.1%) of material from the World Wide Web. The top 20 users used 356,284 kB (51%) and the top 10 users accessed 241,032 kB (34.5%) of content. This trend would indicate a category that is under potential non-business usage by the users. While the downloaded volume for this and similar categories is not large the downloaded material still represents activity that would be potentially counter productive to the organisations core business.

7.6.14 Shopping (Unacceptable)

A total of 194 users (62.7%) accessed material within this category to produce 696,642 kB (2.1%) of downloads. The top 20 users consumed 463,136 kB (66.4%) of the material downloaded. The top 5 users consumed 237,881 kB (34.1%) of the material which would indicate definite misuse by these users. This activity could be as a result of staff working longer hours or skipping lunch breaks and being unable to shop.

7.6.15 Sports (Unacceptable)

In this category 131 users (42.3%) accessed 675,243 kB (2%) of material. The top 20 users consumed 538,423 kB (79.7%) of the material accessed. The top 5 users accessed 319,359 Kb (47.2%) of the material which would indicate misuse. The top user accessed 109,323 kB (16.1%) alone which indicated potential misuse of the resource by this user.

7.7 Analysis of Top Users

The extended analysis of the users in this case was done with users who have the highest usage. The higher number of users examined in this case, was done primarily due to the low level of misuse detected in the initial Cyfin analysis and to maintain the rigour of the interpretation of the data. The previous two cases had top users whose activities were more extensive in total size. So to maintain the Hermeneutic Circle a higher degree of iteration was needed.

7.7.1 Misuser 1

Misuser 1 appears to be a newly appointed systems administrator who downloaded the highest volume of all users with 1,026,535 kB (3.0%). What is exceptional compared to other activity within the organisation is that this volume was downloaded in the period from the 7th to the 30th of August, 2002. Over 94% of the user's downloaded volume was files that are 10Mb or larger and of type .exe. This pattern is indicative of downloading patches and archives of update files for the various systems that the system administrator was responsible for maintaining. Upon further investigation of the log files this situation was found to be true.

7.7.2 Misuser 2

This user downloaded 875,910 kB (2.6%) of traffic over the entire period that was examined. The pattern for this user seems relatively innocuous with what would, at first appear to be typical traffic. However, upon further investigation it was found that this user was accessing and viewing e-books. There were a wide variety of e-books downloaded by this user from various sources. Predominately the books were fiction based either downloaded from author sites as sample chapters or as complete books from free or pirate sites.

7.7.3 Misuser 3

This employee consumed 776,841 kB (2.3%) for the period examined. This employee would appear to have a normal usage pattern with no excessive downloading of binaries. However, there were 18,592 requests made to the site http://www.eatinperth.com which is 9.4% of the users total requests generated in the period examined. What makes this user's requests to this particular site of interest is that he/she is accessing e-mail functionality on the web server. This access is being done via the use of the WorldClient.cgi client, which is a web based mail interface for the MDaemon e-mail system that ran on the eatinperth.com server. With this level of access to the system, it would be reasonable to assume that the user may be somehow involved with the development or support of the site. The user also made extensive use of the Hotmail web based e-mail system during the period.

7.7.4 Misuser 4

This user consumed 771,137 kB (2.3%) of material for the period and, made 11,807 requests to a share trading site which constituted 8.4% of the user's total requests generated. This activity occurs frequently, but not with any apparent regular pattern of usage. The share trading site is not accessed everyday and has had gaps in access of up to 8 days. One might draw several assumptions, either the users share portfolio is not a high risk matter and does not need constant monitoring, or the user is purposely accessing the system in a random manner to evade detection with a trade and run approach. The majority of the other activity by this user would appear to be related to processing an organisational workflow.

7.7.5 Misuser 5

This user downloaded 765,662 kB (2.3%) of material over the period. Over 66% of the user's traffic was either .exe, .zip or .cab (Microsoft Cabinet Archives) files. The potentially suspicious traffic in this 66% has the user downloading freely available tools/trial editions around a given topic (for instance, PC backup). Some of the patch files that were downloaded had no relationship to the organisations Standard Operating Environment of Microsoft Windows 2000. The downloaded materials were predominately for Windows 98, which had not been used by the organisation since December 2001.

7.7.6 Misuser 6

This user was working for the organisation until Friday 21st June 2002 during which time the user downloaded 709,944 kB (2.1%) of material by making 124,379 requests through the organisational proxy server. The user seems to have had the month of April 2002 off, or was not in a place where their accesses were registered by the proxy.

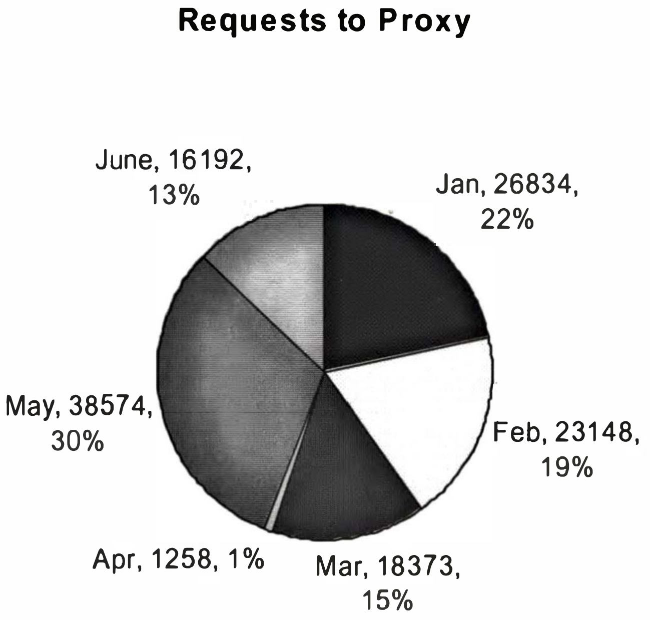


Figure 7.4: Access to the proxy server by Misuser 6

From Figure 7.4 it can be seen that in May 2002 and June 2002 this user accessed the proxy server 54,768 times or 43% of accesses which is comparatively high. This statistic is still not truly indicative of the increased usage by this user in these months. Taking into account no sick days and public holidays for the months of January 2002 to March 2002 there were 61 working days which resulted in 68355 requests or an average of 1120 requests per working day. In the period May 2002 to June 21st 2003 there was 37 working days for 54,768 requests or an average of 1480 requests per day that is 32% increase in traffic by this user. This trend possibly indicates misuse as a result of leaving the organisation's employ and the inability of the organisation to impose effective sanctions.

7.7.7 Misuser 7

This user downloaded 669,033 kB (2.0%) of content with .exe files accounting for 81.25% of his/her usage. The majority of downloads themselves were updates to the Desktop SOE of the organisation via the Windows 2000 update mechanism. However, amongst the legitimate downloads were 2 files that by their nature are suspicious and potentially illegal if used. The files are both from the same site. The first file is called "Enabler" and its use is described as follows:

> "Enabler - I wrote this program because I was curious about manipulating object properties in running applications. Well that's what the program does. You are given a list of all top level windows, you can open the windows and browse the child controls and change their state. I have found many applications have hidden controls. You can save the changes you have made, monitor for the app and have Enabler load the settings automatically or even create an icon to load the manipulated app and adjust the settings. Also, the program allows you to see text obscured by ***** in some password fields. This option may grow into a full blown object editor if time and interest warrants." (Anonymous, 2002a)

The second file is called "Look 4" and its use is described as follows:

> "Look4 - This one is NT/W2K only for now. I found myself wasting time at work looking for people who were not in the office. (Calling or walking to their desk...) I also sometimes missed people I wanted to see before they left. Look4 can watch for people to arrive or leave so you can catch them before they get started or before they walk out. You can also use the program to send them network messages. (-- For fun when you send a message you can forge the sender.--) You can search by Computer Name, User Name or IP. It resolves host names as well." (Anonymous, 2002b)

The download and subsequent use of these tools on a network in Australia would be against several Federal and State Criminal Codes and Laws. Possession of tools similar to this could at least see a conviction secured on the grounds of intent. At a malicious level, these programs could be used to target a staff member who is absent from the PC via the use of Look4. The targeter could use the Enabler program to reveal secrets from the targeted PC avoiding detection and compromising the security of the system or data files contained thereon.

7.7.8 Misuser 8

During the period examined, the user downloaded 515,180 kB (1.5%) making 350,279 requests overall. The user extensively accessed a particular share trading site over the period examined with 49,947 requests which resulted in 105,512 kB of downloads from the share site. The misuse by the user was routine in that they accessed the site everyday and at random times throughout the working day. The remainder of the activity for this user was within policy guidelines for the organisation.

7.7.9 Misuser 9

This user accessed 496,253 kB (1.5%) of material from the proxy server during the period examined. The user accessed material relating to a Palm OS handheld computing device and downloaded electronic versions of manuals and shareware applications for same. The user was a car enthusiast and regularly accessed high performance car sites. The misuse, either by accident or design, saw much of his/her inappropriate activity perpetrated between large bursts of regular use. As an example, the tracking down of a specialist high flow rate oil filter normally found on high perfiormance car engines was researched over a period of 7 hours and interspersed with normal traffic. The same task if done without interspersing would have taken 5 to 10 minutes of time.

7.7.10 Misuser 10

This user accessed 466,797 kB (1.4%) of material as a result of 145,089 requests to the proxy server: 81,237 requests were sports related. The user was covert in his/her misuse. At a cursory glance, it would seem that the user was accessing a particular news service, which is a legitimate activity under organisational policy. However, the user frequently accessed the same site URL, which would indicate the use of bookmark.

Delving further it is clear that the user has accessed streaming pages that update scores or give the world wide latest sports news to the viewer. Some of these streamed text sessions were in excess of 100 kB. This metric means 100,000 characters or 50 typed pages of text characters have passed the screen during the viewing session, which on many occasions amounted to the full day at work.

This type of text traffic is typically very wasteful of network bandwidth as, often, a whole network packet which could hold up to 64kB of data is in fact limited to several bytes of data. This produces network overhead for the handling of these small network packets at certain layers in the TCP/IP protocol. It is equally arduous to process a small data packet, as it is large data packet due to the packet switched nature of the TCP/IP protocol.

7.7.11 Misuser 11

The user downloaded 454,287 kB (1.3%) of material from the proxy server. This user followed a normal usage pattern apart from an extensive and persistent use of the Hotmail web based e-mail system. Out of the total 98,003 requests made to the server, 38,575 requests (39.3%) were made to Hotmail. The Hotmail content was responsible for 389,934 kB (85.8%) of this users' accessed material which given organisational policy is considered unacceptable.

7.7.12 Misuser 12

During the period examined this user accessed 444,050 kB (1.3%) of material through the proxy server. The user made extensive use of the on-line radio features of Windows Media Player to listen to radio on-line. The user frequently accessed local radio stations websites and most URLs indicated the accessing of competitions related to the particular radio stations.

7.7.13 Misuser 13

This user downloaded 438,459 kB (1.3%) of material from the proxy server. The usage patterns were varied and could be classified as opportunistic misuse of the World Wide Web with no set pattern of access. They user accessed various computer games related sites to download game cheats, demonstration files or reviews of new games. The user did a spot of holiday planning and reconnaissance which involved browsing of airline sites, money exchange sites and hotel related sites. The user regularly accessed auction sites and shopping sites, always looking for the same genre of article. Over 90% of downloads by this user were not work related and constituted a misuse under policy.

7.7.14 Misuser 14

This user made 124,863 requests which produced 432,568 kB (1.3%) of material through the organisational proxy. The use of Hotmail was extensive, resulting in 11,480 requests that produced 67,828 kB (15.6%) of traffic. This user made use of the National Channel 9 Television's 'ninemsn' web portal with 25,601 requests producing 96,981 kB (22.4%) of traffic. The user was also accessing various real estate and bridal sites as well during the period.

7.7.15 Misuser 15

This user accessed 412,116 kB (1.2%) of material during the period. The user made extensive use of a web based e-mail system from a local ISP. These accesses would at first glance seem like accesses of a local web site in the proxy log files. However, upon closer scrutiny many of these accesses terminated at the ISP's web mail system. The user was actively using a technique to mask his/her intentions by layering access to the site. It would be more efficient to use a bookmark to access the e-mail system but by not doing so the user is hiding the misuse in amongst seemingly legitimate traffic.

The user heavily participated in recreational 'surfing' of movie and theatre related sites, body building sites and real estate sites. Over 80% of the material accessed by this user was considered misuse under the policy.

7.8 Post-Analysis Interviews

After the presentations, post-analysis interviews were conducted with the IT Services Manager and the IT Support Manager of the organisation. Both parties were sent a thorough analysis of the traffic usage prior to the meeting. The analysis confirmed what both the managers had expected in terms of large issues to be dealt with.

When asked as to why they believed the level of access was so low, The IT Services Manager could only put this level down to the fact that they had a highly visible and publicised access policy for the network. They were both relieved to hear that there were no large issues created as a result of traditional internal threats, such as pornography.

When asked about the on-line banking activity, the IT Services Manager stated that while this was not officially sanctioned, it was not officially persecuted, by management as they understood that people worked long hours and allowed staff to bank on-line as a consequence. When asked about the top misusers in the financial section who were regularly accessing share trading sites, the IT Services Manager said that this activity was not acceptable and should be stopped.

When asked if they foresaw any problems with current misuse and its potential effect on the impending planned move to a higher capacity network connection in the near future. The IT Services Manager stated that the existing connection was running at less than 10% of total bandwidth capacity and did not see the shift causing any foreseeable problems as a pattern of good behaviour was already established.

7.9 Discussion of the Case Results

On a per capita basis, this organisation has a low usage of the World Wide Web which would typically indicate a low level of Internet misuse. However, some of the misuse uncovered could possibly indicate otherwise. The users in the organisation are aware of, and have access to well documented, highly available and regularly updated policy regarding the use of the World Wide Web. This awareness has possibly contributed to the lower level of usage, however, it has revealed some interesting behaviours on the part of misusers.

Several users in the organisation had shown deliberate and covert misuse of the World Wide Web. This type of user attempted to mask their activities by hiding suspect traffic amongst large volumes of legitimate traffic. Several users employed a technique of hiding inappropriate use by accessing what would appear at first glance to be legitimate traffic. A misuse example is the accessing of sites such as web based e-mail sites through several layers. This activity executed in a more efficient and direct manner through the use of a bookmark. A user by, utilising bookmarked navigation, would instantly cause a log processing tool to highlight the fact that the he/she was accessing just the web based e-mail service. This *modus operandi* requires a single URL to be accessed each visit, instead of the masking or burying the target URL in amongst near target URLs. By adopting this method of access the user leveraged the hierarchical nature of web pages to defeat, or at least lessen, the effect of log analysis tools at tracking down his/her misuse.

There was also the case of the user accessing a single URL that would, at first glance appear, to be a legitimate use of a news service. However, with closer scrutiny and testing by the author it would appear that the user had captured a sports news ticker that would scroll game scores and latest sporting news throughout each working day.

Due to the restrictive nature of the policy in the organisation some users had adopted 'surf and run' tactics in the use of the Internet. Instead of accessing a share trading site during the whole working day they appeared to be trading only on a 'needs to' basis. This activity made their misuse relatively hard to detect as it was not habitual or apparently constant in nature. The policy, while not completely eradicating misuse, possibly had an undesired effect by causing deceptive behaviours in misusers: for example policy, made them sensitive constantly and frequently accessing sites which then would be highlighted and leave an audit trail.

This case also saw minimal downloading of traditionally problematic material such as pornography, copyright works such as MP3 song files or large binary downloads. This trend was not related to the size of the Internet connection the total bandwidth consumed was 1.1% of total volume capacity of the 1.5 Mbits per second connection. This outcome could be related to several factors other than policy, such as actual workload within the organisation, layout of the office which was predominately open plan, or the nature of the organisation being in the legal field.

CHAPTER 8 – DISCUSSION OF CASE IMPLICATIONS

This chapter examines trends and patterns that have emerged from the cases studied. The cases have had the same logic of investigation applied to each of them. Using this interpretive method, the author can draw inferences from the case finding but can make no generalisations beyond the cases examined, although speculation for further research is always possible.

This chapter addresses implications that relate to the research questions. These were:

1. To what extent is the World Wide Web service used by employees for non-business related activities in selected organisations?
2. What is the perceived versus measured reality of non-business use in an organisation?
3. Does the enforcement of countermeasures such as policy affect the behaviours of users within an organisation?

8.1 Non-Business Related Activities

The level of non-business usage is not a hard and fast metric but one that is developed within each organisation through social interactions that determine the management and use of the World Wide Web system. The concept of non-business usage is specified within the organisational usage policy for the World Wide Web.

When applying prescribed policy all of the cases had significant levels of non-business usage. Table 8.1 quantifies the level of non-business usage for each case.

Case	No. of Users	Level of Unacceptable Usage	Volume (Gb)	Cost of Volume @ $200 AU per Gb	Cost Per Month	Cost Per Year	Cost Per User
1	846	75%	90.4	$18,080	$4,254.12	$51,049	$21.37
2	1995	56%	80.1	$16,020	$2,912.73	$34,953	$8.03
3	309	21%	6.4	$1,284	$160.50	$1,926	$4.16

Table 8.1: Levels of Non-Business Use

The level of unacceptable usage was determined from the final optimised output from the Cyfin log file analysis tool. This metric measured the level of usage and was based on each Cyfin category allocation of Acceptable, Neutral or Unacceptable as a result of applying that case organisation policy. The allocation of a category status, as Acceptable, Unacceptable or Neutral, had been confirmed with the organisational stakeholders before any log file analysis began.

This allocation allowed for the calculation of the unacceptable volume usage by each case organisation based on the total downloaded volume for each case. The base rate of $200 AU per Gigabyte was calculated, based on pricing from ISPs at the time of the access by each organisation.

During the period examined the average weekly wage for a person working in Australia was $853.40 per week or $44,376 per annum (ABS, 2002). Based on this average wage Case 1 in total and Case 2 could justify a 78% salary to reduce the non-business usage on direct costs alone. This justifiable salary in Case 1 and Case 2 based on the evidence is contrary to findings of Mirchandani & Motwani (2002, p.55) "Some practitioners we interviewed estimated the average cost of Internet abuse to their company (in terms of lost productivity) did not exceed the annual salary of one employee."

The costs in the table above do not attempt to quantify intangibles such as lost productivity. A costing based on wages alone would be a complex measure and would need to take into account information that organisations were unwilling to impart.

It is reasonable, however, to assume that the costs from non-business usage and productivity losses would be significant. A simple example can be given to highlight this problem based on evidence from Case 2, which had the organisational WAN links becoming congested with non-business related material in the afternoon. This congestion of the WAN links for transportation of World Wide Web traffic was affecting the ability of other workers to complete business related tasks in a timely fashion, thereby causing the productivity losses cited by management.

If we use a simple measure that each legitimate business activity a database transaction takes three seconds more to complete as a result of the congestion caused by non-business usage, then it appears comparatively small as a singular micro instance. However, extrapolating this scenario to 120 employees, who perform an average of 20 database transactions per hour as cited by management which equals 1 minute lost per employee, the net cost in time is 120 minutes per hour for the organisation. Taken at an hourly rate of pay of $18 for this class of employee this equates to $36 loss per hour. With this loss of 4 hours per day this costs $144 per day or $3168 per month (based on 22 working days a month) in lost productivity. This simple example highlights the magnitude of the problem, when for Case 2 this three second per transaction loss is outstripping the cost of downloaded Unacceptable bandwidth by 9%.

Not all non-business usage was a result of user action via direct request and download of content. In all of the case organisations examined the Banners/Ads, category consumed significant bandwidth. Much of this material serves no purpose other than to advertise products and, they typically are graphical in nature and are non-cacheable objects. In some pages, banners and ads can account for up to 80-90% of downloaded page size (Valli, 2001). This type of content has significant impacts on performance on both the server and client sides.

Misusers in all case organisations investigated displayed behaviours or web activity patterns that would indicate they were deliberately avoiding detection. In all case organisations some frequent misusers had conducted forward intelligence and environment scanning and had gained target websites links in the form of URLs or IP addresses via other channels; possibly via home computers or targeted e-mails.

Evidence for this was indicated by users who upon accessing the World Wide Web for the first time on a give day enter selected links into their browsers. During the studied period there was no attempt made to search for this content via search engines to find the website address or content. These links were unique; they were not recorded in any previous activity by the user on the World Wide Web and, were highly focused often down to specific files. In addition, the URLs were often complex and long in structure making short term memorisation by users a very difficult if not impossible task.

Many of the pre-researched sites accessed by these users were no longer accessible by the author when examining and verifying the data contained in the log files. Some of the accessed sites had notices from website providers saying that the site had been removed because it contained inappropriate content. This situation would indicate that these sites that are rotated through in order to avoid detection or prosecution (McCandless, 1997). By these traits alone these sites and file dumps are highly transitory and have short persistence on the Internet.

This activity is a deliberate and premeditated misuse of the World Wide Web system where users are using the organisation's superior bandwidth capabilities to access materials for personal and targeting specific files. The motivations for such behaviour could be that of a download speed advantage. It is unlikely that this type of activity could be classified as "constructive recreation" (Oravec, 2002) due to its premeditated nature. It could be argued that it may be as a result of addressing perceived injustice in terms of conditions of work on the part of the user (Greenberg, 1990; Lim, 2002).

The behaviour aids in masking activities as there is no extensive evidence trail provided as would be the case with conventional methods of locating content used by other potential misusers. For example, lets say a misuser wants to find a site to hire a function room at a hotel for a party they are planning. It would be normal practice for the user to engage in targeted searching and browsing of sites trying to find a web site that is suitable for their party planning purpose. This activity would leave a sizeable audit trail which can allow an administrator to reconstruct the user session. By performing a keyword search on the phrase "function room" the administrator would receive multiple hits from the log file relating to the user's search based activities. By using the pre-researched URL the habitual misuser would be denying this tangible audit trail to the administrator.

This type of misuse activity is hard to combat with conventional log file analysis tools and often needs manual hand tracing and investigation to garner results. The resulting benefits are often small when compared with effort: that is, the time input cost alone on behalf of the administrator often outweighs the cost of the bandwidth needed to access the offending material in the first place.

In Case 1 users were trying to avoid detection by accessing the material early in the morning for example, between 1 a.m. to 4 a.m., where the probability of detection through natural surveillance such as the use of active audit trail monitoring or a "walk-in", would be low. This behaviour was as a result of the users having intimate knowledge of the workings of the organisation, and socially engineering their activities to reduce the probability of detection. This type of simple masking can have a reverse effect and make it easier for an administrator to track misuse as by focusing the examination of the log files to the periods of low natural surveillance.

In Case 3 users were trying to mask non-business use by hiding their activity in amongst large volumes of legitimate traffic which could be called the 'needle in a haystack' approach. Unlike the previous cases of detection avoidance, the users actively researched the sites they were visiting but did so in a stilted and controlled manner. The threads of evidence were there to be found but they were buried under considerable amounts of legitimate traffic generated by the misuser. This behaviour is hard to counter without extensive, highly granular examination of log files as occurred in all case studies for this thesis.

This 'needle in a haystack' approach is simple and exploits a weakness in the *modus operandi* of a web log file analysis tools. The role of a web log file analyser is to reduce the thousands, sometimes millions, of log file entries into a neat, reduced executive report that a system administrator analyses to interpret trends and patterns in traffic consumption. Typically, most log file analysers will list the most frequently accessed pages from the log file or they will employ some pre-determined limit like the Top 100 URLs, set typically by the administrator. Similarly, log file analysers will often list the heaviest consumers of web traffic. Although, this form of attack may highlight a user as a high volume user of a system, it will typically not flag them as a misuser because their top 50 sites are legitimate. So, this form of avoidance is effective as the pages accessed by the misuser will have a low frequency count on the abusive activity and, subsequently will not be output into the log file analyser's report.

The significance of these behaviours is that traditional log file analysis tools are merely statistical tools and they have weaknesses when analysing misuse or non-business usage within organisations. A singular reliance on such analysis tools to detect misuse by users is questionable. The incorporation into an analyser of blacklisted or approved site lists at the analysis phase goes some way to countering this problem for organisations. The content categorisation ability of Cyfin using policy guidelines as the arbiter for Unacceptable or Acceptable is a more effective means of measuring non-business usage. Using the Cyfin tool content rather than being processed and analysed on a user or site based context which is typical for conventional log file analysis tools is processed and analysed from a content based context with a policy focus. By using this content based context with organisational the policy oversight this tool allows the investigator to concentrate on the content rather than on individuals.

8.2 What is the perceived versus measured reality of non-business usage?

Each case organisation's key stakeholders held views and opinions that were not consistent with activity that was discovered in the log files. In all cases no organisation demonstrated any awareness of levels of usage of the World Wide Web. When asked as to what levels of non-business usage they believed was occurring, staff did not know or they speculated at best. This situation further confirms suspicion about some management surveys that are conducted on non-business usage or misuse of the World Wide Web. The three cases are by no means provide a conclusive finding but they do concur with reservations pointed out by Holtz (2001).

All cases cited the cost of monitoring software and its deployment as an impediment to the use of ongoing systemic monitoring. This observation concurs with research conducted by Mirchandani & Motwani (2002, p.26) where "Most companies already have the means to track the Internet usage of their employees but choose not to do so because of the effort involved." It is apparent in all cases that the initial cost of purchasing monitoring software could see a return on investment realised within 2 – 12 months based on direct cost of the Internet bandwidth arising from non-business usage. This return on investment is predicated on the organisation effectively using the software to enforce policy and, hence, reducing inappropriate usage. This situation indicated a large gap in management perception and reality of the log file data.

The reasons for this perception gap could be seen as the management being unaware of the true monetary loss that non-business usage represented: but the reasons may go deeper into the organisational psyche than this. The implementation of countermeasures may simply be perceived as "yet another job" that the for which the management becomes responsible for. The reluctance to deploy these measures may be a fear of unleashing or uncovering a greater evil to cope with, or it may be that the management is uncomfortable about monitoring employees' activities.

A major suspicion the author had was that much of the literature regarding the accessing of pornographic content in the workplace suffered from hyperbole and exaggeration. In each case organisation's there was a belief that pornography was the organisations biggest risk exposure and most likely source of non-business usage. The three case studies showed minimal or nil use of pornography within the organisations. The first case study had 20 users (2.3%) regularly accessing pornographic material and they were responsible for 72.7% of such material and, saw the largest downloaded content of 3,725,700 kB (3.1%) of pornography. The second case study saw only 0.15% of total downloaded content being classified as pornographic. Similarly, the third case study had only 0.13% of total downloaded material classified as pornographic during the period examined. In the latter two cases this usage could easily be explained by unsolicited pornographic e-mails.

The level of access of pornographic material in the three cases examined in was minor to virtually non-existent. If we were to believe some of the existing body of literature such as articles by Greengard (2000) or Hickins (1999), it would have been reasonable to find significantly higher download volumes and levels of participation in accessing and viewing of pornography. In the cases examined this expected level of access did not occur. A reason for this could be the increasing level of end-user awareness about accessing inappropriate material in the modern workplace, or that pornography has lost is novelty value for a large percentage of users.

The first and second case study demonstrated a high download of MP3 and streaming type files such as AVI, MOV and other movie or audio formats. In most cases, the audio based content downloaded by the misusers, was in breach of copyright and organisational policy. All the case organisations operated firewalls and countermeasures that made it difficult for users to access streaming media via the standard file sharing tools facilitating such purposes like Napster, Gnutella and Kazaa. The only real alternative for users was to acquire this content through the web browser via the organisational http proxy server from either http or ftp sites.

This situation has significant implications for the use of World Wide Web in organisations. The main perceived threat of pornography was largely not realised in two out of the three cases. Also, it could be argued that due to the low level of users accessing material found in Case 1 it was also relatively low in potential impact to the organisation. The illegal accessing of streaming content such as MP3, AVI and WMA and other file types that contain copyright material has been shown to be a greater risk exposure in all cases.

A reason for the high exposure to risk from MP3 is the ubiquity and inclusion of technology to support MP3 and similar multimedia file formats in most modern operating systems. This ubiquitous bundling of technology provides ample opportunity for users to utilise this technology and access related materials whether they are legal or not. The penalties for copyright infringements in most developed nations are becoming increasingly higher, ostensibly, to protect digital rights of content producers. The risk of prosecution is now becoming a reality, as a result of the activism of the Recording Industry Association of America (RIAA) and similar bodies in pursuing transgressors of copyright law.

The RIAA is heavily pursuing MP3 based infringement of copyright recordings and is now starting to investigate large institutions and organisations (Cox, 2003; Naraine, 2003). RIAA has successfully prosecuted MP3.com, a site that permitted the download of copyright material from their servers with damages paid in the order of US$100 million. RIAA has recently filed against Morpheus who produce a freeware application that facilitates the sharing of files between users and is commonly used to share MP3 files (Reuters, 2003). In some recent cases that the RIAA is prosecuting it is seeking damages of $US125,000 per breach of copyright, significant and sizable damages. The RIAA has even gone to the point of advocating for the right to attack systems that hold copyright material wherever they are located on the Internet (McCullagh, 2001).

All examined organisations that were could possibly be susceptible to pursuit by an aggressive RIAA or an Australian equivalent and potential damages could be significant. There was a perception that pornography was the paramount problem and the one that could cause the most damage, but the difference in perceived use of the World Wide Web and the reality has brought about this erroneous perception by management. As a result of this gap between perception and reality many organisations are lacking any effective security and acceptable usage policy that places the legal onus back with the end user. By not having effective and enforced policies in place, ultimately, any organisation will become responsible for any liability as result of their staff accessing copyright materials. This problem is a serious issue and needs redress by many organisations if they are to lower risk exposure as a result of increased accessing of MP3 and streaming media within organisations.

The cases have found nexus points where the perception of management versus the reality of usage of the World Wide Web did not intersect and there were credible and substantial. The first nexus point is that of actual usage, humans are poor estimators and all 3 cases have been shown to be no exception. In all, cases management were surprised by the level of non-business usage uncovered by the study of their log files. The second nexus was the reality of costs associated with non-business usage. On the simple measure of raw bandwidth cost alone, Case 1 and Case 2, could justify a salary for reduction of cost in Internet bandwidth, yet all organisations were reluctant to measure usage or expend effort in doing so. The third nexus was that due to a lack of diligence or business intelligence many of the risk exposures for the organisations were badly aligned with the reality displayed in the log files. All organisations saw pornography as the paramount threat yet the more current and realisable threat was that of copyright MP3, streaming media and other files. This problem has follow on effects in enforcement of policy and remediation of risk. If one holds a preconceived notion of misuse then one might focus one's energies at finding that risk/behaviour and, consequently, may miss other risks/behaviours entirely.

8.3 Are countermeasures effective at reducing non-business activity?

Each case used policy with varying levels of awareness and transparency to achieve communication of the organisation's wishes with regards to employee access of the World Wide Web. Case 1 had the most pornography even though it had a policy that clearly stated penalties for accessing such materials. Furthermore, Case 2 applied policy and a content filtered approach for their World Wide Web traffic. However, Case 3 used no content filtering and achieved similar levels of end user compliance regarding the downloaded pornographic content.

The main difference between Case 3 and Case 1 was the former actually had cases where policy and sanctions had been fully enforced as per the published policy with transgressors having punitive sanctions applied against them. This would indicate that there is at least a possible causal link between policy effectiveness and enforcement based upon the existing policy. In Case 1, there were instances of users having been caught grossly transgressing the acceptable usage policy with no subsequent exercise of punitive threats eventuating. This inaction would have sent a signal to existing users that transgression of the policy resulted in no penalty being applied.

No case organisation was actively monitoring World Wide Web usage with sufficient windows of opportunity for detection and remediation of end-user activity via organisational policy. Case 3 was the only organisation that regularly analysed log files and then was monthly and with minimal analysis on the part of IT Service Manager. The other two cases had no active monitoring or reporting of activity. All cases lacked basic business intelligence resulting from the simple and timely analysis of proxy server log files. There was no formal reporting of usage or subsequent follow up based on any anomalies discovered. This is a serious deficiency in all cases.

No case organisation in their policies explicitly stated what rights an end user had to privacy when using the World Wide Web. Furthermore, there was no specific statement about redress or due process should a user be cited for misuse when using the World Wide Web or a prescribed level of penalty.

All cases had instances of users who left the particular organisation within the time frames examined by the author. In the cases investigated, a marked and large increase in baseline usage by the identified users was noted during the notice period.

This issue is significant and potentially shows that policy will only work when the threat of a credible and enforceable sanction is realisable (Harrington, 1996). For the departing employee, some of the most powerful methods of sanction had become greatly reduced in potency. The threat of dismissal and any apparent financial penalty for such action is a paramount threat used by many organisations to maintain control. The power of this threat is largely removed as the departing employees had another job/position to take up and any potential loss of income is at most four weeks pay.

An additional sanction would be the threat of losing a favourable reference. This threat is of limited potency as the user has in all probability secured employment elsewhere and consequences from misuse are lessened due to the employee's impending departure. As a result, the user would not have to suffer any long term social or group effect for being "discovered" accessing the World Wide

Web in an inappropriate or unauthorised manner. Dependent on how long organisational processes take to react or detect misuses, a user may be able to avoid detection until he/she has left the organisation.

This situation, evident in all 3 case organisations, would indicate that organisations need to address exit issues in policy. The employee is normally behaviourally conformant due to operant conditioning, through the use of a penalty or disincentive enacted in policy, however, when he/she resigns having attained employment elsewhere these penalties or disincentives become ineffective. The logical conclusion would be to counter the lack of control that the policy has over the resigned employee and use other measures of sanction.

The problem is a balance between maintaining control and human relation issues such as trust, perceived justice and job satisfaction. If the organisation adopts countermeasures that are perceived by staff to be draconian in approach, then employee distrust of the organisation could occur. An extreme, although not entirely implausible scenario, could see the consequence of tendering a resignation by an employee result in automatic loss or severe restriction of Internet based access. This action could be perceived as a lack of management trust by the employee and, could result in end-user resentment and users seeking to rectify the injustice by taking more liberties with the system. These behaviours are already apparent in other studies dealing with perceived injustice in the workplace (Greenberg, 1990; Lim, 2002).

Another possibility would be to increase the level of surveillance of the departing employee's Internet activity to monitor it for anomalous behaviour. This action would engender a lack of trust in the employee and could cause them undue stress which in turn could affect performance and job satisfaction. A study by Chalykoff & Kochan (1989) found that computer-based monitoring had a direct and significant influence on the overall job satisfaction of employees. In article, (George, 1996, p.463) outlines several cases and cites studies where employee health had been severely affected by such overt monitoring.

Many World Wide Web filtering products are often advertised as a panacea for organisational control of incoming content. It has been found in previous studies that the product rhetoric often does not meet reality when it comes to management of content (George, 1996; Hunter, 2000; Nunberg, 2001).

The tool used to conduct preliminary investigation into all of the case organisations was Cyfin. It uses a list of categories to classify content into 55 different pre-determined groups. It could be reasonably argued that this is attempting to do a static examination of content that an active content filtering tool such as Cyber Patrol would perform. In Case 1, the organisation did not have content filtering installed but the initial analysis had a very high percentage of sites in the Unknown/Unclassified category.

The percentage of unclassified sites was greatly reduced when it was realised that many of these sites were foreign language sites that were being accessed. Herein lies another potential failing of content filtering management systems namely a cultural or ethnic bias in the production of the tool. The classification tool used in thesis was developed in the USA, as are many available content filtering tools. It is reasonable to assume that many of the customers of these products would have English as a first or second language. Therefore, the developers would spend much of the time developing the detection of English based content, as opposed to other languages based on Mandarin Chinese or Hindi content for example.

To complicate matters further the foreign language sites often used colloquial terms for the content they were supplying to users. Unless a person has lived in a particular culture, they often do not pick up many of the subtle colloquial nuances. This lack of enculturation is possibly as a result of developers having learnt or acquired their language skills by conversing with other second language speakers or from textbooks. This weakness in linguistic diversity could easily translate to content filtering systems development, if companies employ second language speakers to aid in judging content.

Cultural bias is a factor that is rarely mentioned, if at all, in any of the literature relating to content filtering. There can be great bias or difference even within the same cultural framework, for example in the United States of America a liberal Californian will have a different view of obscenity compared to a strict fundamentalist Christian from the Southern Bible belt. Further extending this examination let us consider India's strict approach to censorship. Recent Indian law reform specifically the Indian Information Technology Act 2000, Chapter XI Para 67, shows that the Government of India clearly considers online pornography as a punishable offence (NIC, 2003). Cultural sensitivity to pornography in India has seen providers of search engines prosecuted because their search engines listed pornographic sites. The prosecution occurred even though the search engine providers' websites did not contain actual physical pornographic material.

Many of sites accessed by the users in Case Organisation 1 were oriented toward Indian nationals and ex-patriates, and under Indian law were plainly illegal. The same sites under Australian Law and American would hardly rate a mention as pornographic sites due to cultural bias. The Indian sites often only displayed partial nudity and no graphic depictions of sex acts. The sites would have mostly received lower than restricted ratings in Australia.

If companies are selling these content filtering products, how do they cope with these extremes of cultural bias? The simple answer is they cannot cope, and herein lays a dangerous trap for organisations utilising filtering, if they have employees from diverse cultural backgrounds. What may seem inoffensive and acceptable to a content filtering system may offend and discriminate against an employee due to cultural dichotomous views of acceptability making for some potentially interesting industrial arbitration or court cases. Similar cases have been already heard in the US legal system with Microsoft and Chevron being defendants and involving $US 2.2 million in payouts to the plaintiffs in the cases heard (Greengard, 2000).

Filtering proved ineffective against determined non-business users in Case 2. The users employed pre-researched or known URL's to access information as a method of avoiding detection by the content filtering system. By using URLs gleaned from research conducted externally to the organisation or received in e-mail from communities of practice no extended audit trail is provided on the organisations log files as many of these are unique and single instance. By obtaining URLs in this manner users can keep ahead of a manufacturer of content filters ability to research, classify and distribute a block for an inappropriate site. A single instance URL under most statistical analysis will not appear in the report unless the download itself was comparatively large in size, Case 1 and Case 3 also showed users that were taking the same approach even though their connections were not filtered.

Many of the sites that contain illegal content such as “warez” or MP3 files have short life spans due to detection by the ISP or authorities. Determined habitual users are known to move content around sites to avoid detection or prosecution, often referred to as dump sites or zero day sites (McCandless, 1997). As a result content filtering systems will miss these sites due to production cycles of their banned sites databases. These databases are like virus signature databases in that they contain lists of sites that are considered to fit a particular category which the content filtering system then uses to block or allow access dependent upon settings by the administrator. These databases are distributed in a similar manner as virus signature updates and, hence, there may be a seven day or worse lead time before sites are added to the content blocking database.

In all cases examined, where there was significant misuse of the system a small number of users were responsible for downloading large percentages of the total download. In some extreme cases, 1 or 2 users were consuming 80 - 90% of bandwidth for a particular category. By targeting and remediating these users, bandwidth usage can be reduced as can the organisation’s risk exposure. The problem users may have some addictions that could ultimately impact on their working life for example, gambling. Organisations have a duty of care to ensure that these types of users are counselled and helped with their addictions.

Countermeasures such as policy and filtering do have some effect on end user behaviours. At the lowest level, it would be reasonable to say that countermeasures had some effect as identified problem users were finding mechanisms to defeat or counteract them. Policy sees, at best, end user compliance and, at worst, users changing or masking behaviours to avoid detection. Filtering in the one case encountered, Case 2 had users employing covert channels and zero day sites to acquire content in an attempt to bypass restrictions as a result of content filtering. These bypassing activities were cited in the other cases, but the level and repetition of these incidences was higher in Case 2.

Monitoring was not done by two case organisations and Case 3 only did a monthly processing of log files and quick simplistic analysis, hardly comprehensive. This activity is rudimentary business intelligence and indicates that the organisations have failed to, or are unwilling to, engage with policy compliance issues relating to the World Wide Web.

8.4 Summary

A significant level of non-business usage was encountered in the organisations examined. Case 1 had 74.6%, and Case 2 had 56.4%, of bandwidth consumed by non-business usage if policy was strictly applied. Case 3 while not having high non-business bandwidth consumption, still had 20.6% of non-business related use occurring. Even Case 3 with the lowest of non-business usage volume is seeing one in five kilobytes being wasted by non-business users this is a significant level of misuse. Case 1 and 2 could be argued to have high to very high levels of non-business usage by the simple metric of bandwidth consumed.

All cases highlighted the issues of non-business use and its true cost goes beyond the technical metrics of bandwidth measurement, reaching into a range of organisational areas and issues. For example, Case 2 had issues with non-business usage of the World Wide Web impacting on the organisation's ability to deliver information in a timely manner to stakeholders due to congested intra-organisational network links. Case 1 had IT Staff reluctant to perform any measurement of misuse due to past inactions on the part of management when offenders were identified. All cases had staff becoming deliberately deceptive and covert in their actions to access non-business related material. These instances all generated hidden costs and losses, many of which are hard to quantify accurately.

Ascribing a true monetary value to each of the levels of non-business usage in each case presented is difficult and complex. Measuring the cost of providing a network and, subsequent, delivery of non-business content and converting this to a monetary figure based on an ISP's usage charges is not truly indicative of the overall problem. Case 3 taken, on cost measures based on bandwidth alone would see the non-business usage of the World Wide Web as non-problematic. However, based on case context and the type of information the organisation accesses this non-business usage is significant with 20.6% or one in five pieces of content a misuse or threat.

Policy was used in all cases with varying degrees of success as a countermeasure to non-business activity. Case 1 had problems with effective enforcement of policy which was largely ignored by stakeholder groups. Case 2 had policy but it was not overt and it relied on a content filtering approach as a substitute for effective management. Case 3 appeared to be an exemplar in the use of policy to reduce non-business activity. The cases demonstrated that policy or the lack of policy has some effect on end-user behaviour. Cases 2 and 3 highlighted policy that requires consideration and that is what to do with staff members who have resigned. Evidence in these cases would suggest that as a countermeasure policy is greatly reduced in efficacy once a person has resigned.

All cases had users attempting to mask behaviour as a result of the countermeasures in place. Many of the identified high level non-business users of the World Wide Web in the organisations had adapted and developed mechanisms that would defeat countermeasures such as log file analysis or content filtering. These cases apart it would appear that even the existence of policy had a positive effect on user behaviours.

There was a lack of alignment between the perceived non-business usage occurring and management perceptions of the level of non-business usage. Each of the examined cases had significant non-business usage of the World Wide Web by end-users. In the initial interview sessions, the key stakeholders within the organisations at appeared to speculate about the level of non-business usage occurring. Case 3 was the exception where the IT Services Manager reviewed a monthly report on World Wide Web usage but when asked about bandwidth issues he said he was unsure of the exact amount of bandwidth consumed and had little or no idea of any misuse.

In the post-analysis presentations, the reaction from management was typically surprise and amazement at the depth and severity of the level of actual non-business usage discovered. This disparity indicates a large gap in the cases between actual activity and perceived activity. This perception gap highlights several important key issues for consideration.

The perception gap by the key stakeholders would hamper effective decisions on policy making with regards to the use of the World Wide Web. It could be said that the key stakeholders were basically unaware of the World Wide Web *modus operandi* within their respective organisations. This lack of awareness could ultimately translate to poor policy or management of the resource.

The perception gap significantly increases the risk exposure for the organisations. Management perceiving no real problem with World Wide Web usage or being unwilling to address the issue, did not encourage nor develop appropriate feedback and reporting mechanisms to deal with non-business use. Even the simplistic use of log file analysis tools to provide basic day to day monitoring was not occurring within the organisations, a dangerous practice.

Management's incorrect perception of inappropriate usage and it's lack of intervention both gave rise to a resultant, dangerous risk identification lag in their understanding of inappropriate usage. All cases had key stakeholders concerned primarily with access of pornographic material which was low to non-existent. Each organisation policy stated pornography as an inappropriate use of the World Wide Web: however, due to the perception gap in the organisations there were new emergent risks such as downloading of copyright MP3 files. These would be identified as inappropriate or problem behaviours if there was appropriate monitoring or measurement taking place. These issues are were introduced into the system and were not incorporated into organisational countermeasures, thereby setting up a potentially destructive positive feedback loop with respect to the management perception and the true reality of World Wide Web usage.

CHAPTER 9 – CONCLUSION

9.1 Lessons learned from research approach and conduct

Using an interpretive approach limits the ability of the researcher to apply the findings to a population. Aspects of this type of research may benefit from a more positivist approach to analysis in particular of the log file data. Nonetheless, the author still believes that the interpretive approach employed allowed for a richer examination of the case situations.

Openness of organisations to provide data of this nature proved problematic. This problem was due mainly to privacy and discovery concerns and the ability of the data if not sufficiently anonymised to uncover individuals or the organisation as a centre of illicit World Wide Web activity. The ability to extract any financial figures for provision of the World Wide Web to the desktop also proved problematic and calculation of financial metrics was based upon statistical information.

One of the limiting factors for increasing the richness of the study was the inability of the author to interview identified high level misusers of the World Wide Web. Interview data from these subjects would have allowed the author to examine motivations and reasons for their behaviours. Interview of these types of misusers in future studies could be constructive and informative.

The research was limited in scope due to time and resource constraints for processing and analysing the content obtained from the case organisations. The author had intended initially to analyse data from more than three case organisations. However, the large volume of log file data collected from the case organisations made time-efficient analysis difficult. Manipulation and investigation of such large data sets was problematic even for high end server based hardware on which some of these tests were run, taking these machines up to 12 hours to complete a single log file analysis run. Due to the iterative nature of the interpretive case methodology and the application of the principle of the hermeneutic circle, this process was repeated at least twenty times on the large concatenated data sets. Analysis of some data sets taken from some the top identified misusers was also problematic due to their overall size. To garner computational advantage the author used Beowulf compute clustering to process some of the largest log files which were over 4 Gb in size. Further research needs to be conducted into the manipulation of large data sets based on log files so that effective extended analysis can be undertaken within acceptable timeframes.

9.2 Assessing Level of Non-Business Usage

The measured level of non-business usage in each organisation was significant. The in-depth analysis of log files uncovered anomalous behaviour on the part of users. The use of several log file analysis tools aided in the investigation of non-business usage. The content categorization ability of the log file processing tool aided greatly in pinpointing areas and individuals for further attention by the researcher. This scenario has implications for how organisations view data for compliance or non-business usage.

The other log file analysis tools used to verify and triangulate results from Cyfin did not have the categorization functionality which limited their ability to target problem behaviours other than simple over use of bandwidth. Cyfin allowed the author to classify readily these content related categories as Acceptable or Unacceptable. It also had the ability to customize or add unclassified sites to specific categories, or create custom categories as well. The implication from this research is that some form of content classification system is critical in determining non-business or misuse behaviours. By engaging with a content classification system, stakeholders are better able to target and identify what they consider to be non-business usage. Research into the effectiveness of using a content categorisation approach for World Wide Web traffic analysis could be undertaken.

Relying on simple monetary value calculations of "lost bandwidth" as a result of non-business use, has been shown to be simplistic and not reflective of the overall true cost to the organisation. With the reluctance of organisations to even provide basic monitoring metrics, it would prove a difficult task to provide accurate measures. The complexity of computing a finite non-business measure would be hard to effect within an organisation without some Hawthornian effect (Mayo, 1933) occurring in the research. Based on limited the calculation performed in the thesis the tangible cost of the bandwidth a small fraction of the total problem. Measuring productivity per se is a difficult phenomenon, but developing better metrics for tracking end-user use of the World Wide Web is needed.

The established metrics of download cost and times are becoming a decreasingly ineffective measure of non-business usage, as we progressively move to higher speed broadband connections. A measure for total cost of lost revenue/productivity as a result of non-business usage ignoring opportunity costs is :

Cost of Non-Business Use = (Cost of download) +
(Time taken to Transmit) +
(Time taken to Interpret)

As we move to higher bandwidth connections the time taken to receive and transmit data will continue to decrease and currently the cost of download is also reducing. The near constant in this equation is the ability of human to process or interpret the download information. This constant of human interpretation is where the greater proportion of the cost of misuse or non-business usage will lie in the future.

Current countermeasures rely heavily on server side logging of end user activity which is only indicative of end-user content access behaviours. Research and investigation into some form of combined client and server side schema to try and measure time lost for non-business usage would help in measuring the levels of such activities.

9.3 Actual versus Perceived Reality of the Non-Business Usage

This thesis highlighted that there is a significant and credible gap between management perception of non-business usage and the measured reality of World Wide Web traffic. The gap in perception was not aided by the fact that all case organisations were reluctant to engage any form of consistent, timely monitoring and subsequent reporting of the use of the World Wide Web. This gap does not allow the organisation to generate the necessary business intelligence to make informed decisions about policy and appropriate countermeasures. The reluctance by organisations to engage in monitoring in this case is an issue that needs exploration. The main justification used by management for not engaging in such activity the issue of cost. This management belief has been exposed as a relatively shallow excuse and, in at least two of the cases studied, their view would not be an accurate assessment. The reasons for this issue possibly lie within the socio-political constructs of the organisations.

This management perception gap is an area on where there is a need for further research as this perception gap could become problematic and hamper effective management of potential problem areas within an organisation. This gap raises questions on the validity or accuracy of many of the management based surveys of organisations that relate to the misuse of the World Wide Web or the Internet.

9.4 Countermeasures against Non-Business Usage

Policy effectiveness and enforcement is an area that needs continual research if we are to achieve best practice. The case organisations deployed policy with varying levels of success. An investigation into the perception gap between management and the reality of end user activities and whether this affects policy is a area of potential study. The perception gap in the cases examined highlighted that lack of rigorous management, due to inaction or lack of understanding of the current risks and threats, was possibly affecting progressive countermeasure and policy development.

Each case had users' undertaking diversionary or substitute behaviours to attain inappropriate material. Further study of these particular users' habits, patterns and motivations could aid in finding effective countermeasures or risk reduction techniques. The content filtering system encountered in Case 2 also proved ineffective against dedicated non-business users who used covert channels or communities of practice to obtain forward intelligence which defeated the system. An investigation into whether this particular system or all content filtering systems are susceptible to this form of attack could be conducted.

Possible sharing of monitoring information at an inter-organisational level may help identify misuse patterns or suspect website URLs. In this way the organisations themselves establish a counterbalancing community of practice to combat some of the activity by end-users who use similar tactics to defeat the countermeasures put in place.

The ability to classify content into categories was a key tool used in this study to aid in the measurement of non-business usage. The growing multi-cultural nature of the web presents problems for content filtering and content classification systems as effective stand alone countermeasures. The ability of these systems to identify content and accurately classify content from a non-English speaking origin is questionable and needs further exploration. This issue could see research into the development of profiled content filtering systems suited to a particular ethnicity, religious or political belief. A cross-cultural research project could be conducted into the issue of cultural, political and religious bias in the design of these filtering systems and how it affects usage and management of these systems.

This thesis has presented a multiple interpretive case based study of non-business usage of the World Wide Web. Unlike many of the management reports and studies conducted before, it has heavy empirical underpinnings through the extensive analysis of organisation World Wide Web proxy log files. It gathered data from stakeholders and organisational documents to support the conclusions made. The issue of non-business related usage of the World Wide Web is complex and varies from organisation to organisation and warrants continual research.

Hopefully, this study is one small part of the jigsaw that forms our understanding of non-business usage of the World Wide Web. It is only through continued objective study and examination of complex phenomena that we will we then begin to understand the whole jigsaw.

REFERENCES

ABS. (2002). *Measuring Australia's Economy - Section 6. Prices and Income - Average Weekly Earnings.* Retrieved 30 June, 2003, from http://www.abs.gov.au/Ausstats/abs%40.nsf/94713ad445ff1425ca25682000192af2/277de45665c11900ca256cbf00017219d!OpenDocument#Links

Akdeniz, Y. (1998). Who watches the watchmen?: Internet content rating systems and privatised censorship. *Australian Library Journal, v.47, no.1, Feb 1998 : 28-42., 47*(1), 28-42.

AMA. (2001). *Workplace Monitoring & Surveillance:Policies and Practices.* New York: American Management Association.

Anandarajan, M. (2002). Internet abuse in the workplace. *Association for Computing Machinery. Communications of the ACM, 45*(1), 53.

Anonymous. (2001). *Human Resources Management - the Missing Link in Controlling Internet Access.* Retrieved 20th January, 2001, from http://www.strategichr.com/shrsweb2/InternetAccess.shtml

Anonymous. (2002a). *Enabler.* Retrieved 14th April, 2003, from http://www.securitysoftware.cc/apps.html

Anonymous. (2002b). *Look4.* Retrieved 14th April, 2003, from http://www.securitysoftware.cc/apps.html

Ballard, K. (2000). *Online trading and entertainment sites to impact workplace productivity.* Retrieved Jan 8th, 2001, from http://www.itweb.co.za/sections/internet/2000/0011220831.asp

Barrett, B. L. (2002). Webalizer: Barrett, Bradford L.

Benbasat, I., & Weber, R. (1996). Research commentary: Rethinking "diversity" in information systems research. *Information Systems Research, 7*(4), 389-399.

Block, W. (2001). Cyberslacking, business ethics and managerial economics. *Journal of Business Ethics, 33*(3), 225-231.

Brown, A. (1999, Jul 19). Controlling interests. *New Statesman (London, England: 1996) [H.W. Wilson - SSA], 128,* 42-43.

Cavaye, A. L. M. (1996). Case study research: a multi-faceted research approach for IS. *Information Systems Journal, 6*(3), 227-242.

Chalykoff, J., & Kochan, T. (1989). Computer-aided monitoring: Its influence on employee job satisfaction and turnover. *Personnel Psychology, 42*(4), 807-834.

Cooper, J., & Martin, E. (2000). Employment Law: Misuse of Email and the Internet in the Workplace. *Australian Company Secretary, 52*(5), 299-300.

Cox, B. (2003). *RIAA Trains Anti-Piracy Guns on Universities.* Retrieved June 6th, 2003, from http://www.internetnews.com/bus-news/article.php/1577101

Davis, G. B., Lee, A.S., Nickles, K.R., Chatterjee, S., Hartung, R. and Wu, Y. (1992). Diagnosis of an Information System Failure: A Framework and Interpretive Process. *Information & Management, 23*(5), 293-318.

Foster, M. (2001, Feb). Snuffing out Internet abusers at work. *Incentive, 175,* 67-69.

FPC. (2002). *Guidelines on Workplace E-mail, Web Browsing and Privacy (30/3/2000).* Retrieved 8th October, 2002, from http://www.privacy.gov.au/internet/email/index.html

George, J. F. (1996). Computer-based monitoring: common perceptions and empirical results. *MIS Quarterly, 20*(4), 459-481.

Gleeson, M. (2002). pwebstats (Version 1.38). Melbourne, Australia: Gleeson, Martin.

Greenberg, J. (1990). Employee theft as a reaction to underpayment inequity: the hidden costs of pay cuts. *Journal of Applied Psychology, 75*, 561-568.

Greengard, S. (2000). The High Cost of Cyberslacking. *Workforce, 79,* 22-24.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In *Handbook of Qualitative Research* (pp. 105-117).

Hamilton, M. (2002). *E mail/internet abuse at work outstripping other forms of misconduct.* Retrieved 24th September, 2002, from http://www.kpmg.co.uk/kpmg/uk/press/detail.cfm?PR=1500

Harrington, S. J. (1996). The Effect of Codes of Ethics and Personal Denial of Responsibility on Computer Abuse Judgments and Intentions. *MIS Quarterly, 20*(3), 257-278.

Hickins, M. (1999). Fighting surf abuse. *Management Review, 88*(6), 8.

Holtz, S. (2001). Employees online: The productivity issue. *Communication World, 18*(2), 17-23.

Hunter, C. D. (2000). Social impacts: Internet filter effectiveness--testing over-and underinclusive blocking decisions of four popular Web filters. *Social Science Computer Review, 18*(2), 214-222.

Kallman, E. (1993). Electronic monitoring of employees: Issues and guidelines. *Journal of Systems Management, 44*(Jun), 17.

Klein, H. K., & Myers, M. D. (1999). A Set Of Principles For Conducting And Evaluating Interpretive Field Studies In Information Systems. *MIS Quarterly, 23*(1), 67-94.

Lawrence, S., & Giles, L. (1999). Accessibility of information on the web. *Nature, 400*(6740), 107-109.

Lee, A. S. (1989). A Scientific Methodology for MIS Case Studies. *MIS Quarterly, 13*(1), 33-52.

Lichtenstein, S., & Swatman, P., M.C. (1997). Internet acceptable usage policy for organizations. *Information Management & Computer Security, 5*(5), 182-190.

Lim, V. K. G. (2002). The IT way of loafing on the job: Cyberloafing, neutralizing and organizational justice. *Journal of Organizational Behavior, 23*(5), 675.

Lim, V. K. G., Thompson, T. S. H., & Loo, G. L. (2002). How do I loaf here? Let me count the ways. *Association for Computing Machinery. Communications of the ACM, 45*(1), 66-70.

Luotonen, A. (1995). *Logging Control In W3C httpd.* Retrieved 1st May, 2002, from http://www.w3.org/Daemon/User/Config/Logging.html#common-logfile-format

Markus, M. L., & Benjamin, R. I. (1997). The magic bullet theory in IT-enabled transformation. *Sloan Management Review, 38*(2), 55-68.

Mayo, E. (1933). The human problems of an industrial civilization. New York: Macmillian.

McAuliffe, M. (2001). *Child porn legislation questioned.* Retrieved 2nd October, 2002, from http://www.zdnet.com.au/newstech/news/story/0,2000025345,20216806-1,00.htm

McCandless, D. (1997). *Warez Wars.* Retrieved 10 May, 2003, from http://www.wired.com/wired/archive/5.04/ff_warez.html?pg=1&topic=&topic_set=

McCullagh, D. (2001). *RIAA Wants to Hack Your PC.* Retrieved June 6th, 2003, from http://www.wired.com/news/conflict/0,2100,47552,00.html

Mills, J. E. (2001). Cyberslacking! A liability issue for wired workplaces. *Cornell Hotel and Restaurant Administration Quarterly, 42*(5), 34.

Mirchandani, D., & Motwani, J. (2002). Reducing Internet abuse in the workplace. *S.A.M. Advanced Management Journal, 68*(1), 22-26, 55.

Naraine, R. (2003). *RIAA Targets File-Sharing in the Workplace.* Retrieved June 6th, 2003, from www.atnewyork.com/news/article.php/2112521

Neumann, P. G., & Weinstein, L. (1999). Risks of content filtering. *Association for Computing Machinery. Communications of the ACM, 42*(11), 152.

NIC. (2003). *Indias fight against Online Pornography.* Retrieved 12th May, 2003, from http://netsafety.nic.in/cyberlaws.htm

Nunberg, G. (2001). The Internet filter farce: why blocking software doesn't--and can't--work as promised. *American Prospect [H.W. Wilson - SSA], 12,* 28-33.

O'Brien, P. R. (1999). Understanding copyright risks. *Security Management, 43*(4), 68-73.

Ohlson, K. (1998). *Recreational Web Surfing at Work Is on the Rise.* Retrieved 21st February, 2001, from http://www.cnn.com/TECH/computing/9808/13/surfing.idg/

Oravec, J. A. (2002). Constructive approaches to Internet recreation in the workplace. *Association for Computing Machinery. Communications of the ACM, 45*(1), 60.

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying Information Technology in Organizations: Research Approaches and Assumptions. *Information Systems Research, 2*(1), 1-28.

Orso, P. (2001). Squid Analysis Report Generator (Version 1.21): Orso, Pedro.

Phipps, P. A. (1996). Electronic monitoring in the workplace. *Monthly Labor Review, 119*(3), 33-35.

Pinto, J. K., & Slevin, D. P. (1989). The Project Champion: Key to Implementation Success. *Project Management Journal, 20*(4), 15-20.

Reuters. (2003). *Play It Again, RIAA: Sue Morpheus.* Retrieved June 6th, 2003, from http://www.wired.com/news/digiwood/0,1412,59097,00.html

Roberts, B. (2000). When cyber slacking crosses the line. *Electronic Business, 26*(13), 56-58.

Robey, D. (1979). User attitudes and management information system use. *Academy of Management Journal, 22,* 527-538.

SIA. (2000). *WebSense Survey on Internet Misuse in the Workplace.* San Jose, California: Saratoga Institute.

Solicitor, A. G. (2001). Misconduct in E-mail and Internet Use at Work (Vol. 58): Australian Government Solicitor.

Straub, D. W., Carlson, P. J., & Jones, E. H. (1993). Deterring Cheating by Student Programmers: A Field Experiment in Computer Security. *Journal of Management Systems, 5*(1), 33-48.

Straub, D. W., & Nance, W. D. (1990). Discovering and Disciplining Computer Abuse in Organizations: A Field Study. *MIS Quarterly, 14*(1 (March)), 45-62.

SurfControl. (2002). Cyber Patrol. Westborough, MA: SurfControl.

Taylor, G. (2001). *Regulatory Failure: Australia's Internet Censorship Regime.* Retrieved 2nd October, 2002, from http://www.efa.org.au/Analysis/aba_analysis.html

Turner, S. (2003). Analog (Version 5.24): Clicktracks.

Urbaczewski, A., & Jessup, L. (2002). Does electronic monitoring of employee Internet usage work? *Association for Computing Machinery. Communications of the ACM, 45*(1), 80.

Valli, C. (2001). *A Byte of Prevention is worth a Terabyte of Remedy.* Paper presented at the We-B Conference 2001, Scarborough, Western Australia.

Walsham, G. (1993). *Interpreting Information Systems in Organizations.* Chichester: Wiley.

Walsham, G. (1995). Interpretive Case-Studies in Is Research - Nature and Method. *European Journal of Information Systems, 4*(2), 74-81.

Wavecrest. (2002). Cyfin Reporter (Version 4.01). Melbourne, Florida: WaveCrest Computing.

Weare, C., & Lin, W.-Y. (2000). Content analysis of the World Wide Web: Opportunities and challenges. *Social Science Computer Review, 18*(3), 272-292.

Web Improprieties Moved Xerox to Fire 40 in U.S. This Year. (1999, Oct 7). *Wall Street Journal,* p. A15.

Wood, T., & Lo, M. (2001). Net and email abuse is no cyber myth: empolyment law. *Keeping Good Companies, 53*(2), 94-97.

Zaiton, H. (2000, Mar). Insider cyber-threats: Problems and perspectives. *International Review of Law, Computers & Technology, 14,* 105-113.

APPENDIX A

W3C Common Log File (CLF) format

remotehost rfc931 authuser [date] "request" status bytes

Field	Description
remotehost	The private IP address used by the client accessing the proxy server (for example, 10.1.1.2)
rfc931	The remote logname of the user
authuser	The username as which the user has authenticated themself
Date	Date and time of the request. This takes the form **DD/MM/YEAR:HH:MM:SS:GMT ZONE** **DD/MM/YEAR** - The date on which the request was made (for example, 03/Dec/2001) **HH:MM:SS** - time of day at which the request was made (for example, 12:07:41) **GMT ZONE** - time zone on the server at which the request was made, as an offset from Greenwich Mean Time (for example, +8000)
"request "	The requests made to the server and stored in logs as HTTP_REQUEST_TYPE URL HTTP_VERSION HTTP_REQUEST_TYPE GET (read a page or entity within a page) HEAD (obtain the header information for a page) POST (submit the results of a form) URL The URL of the site being accessed, including the domain name and the full path within the site (e.g. http://www.ecu.edu.au/index.htm) HTTP_VERSION The version of HTTP used by the requesting entity
Status Code	A number which indicates diagnostics such as successful download of file or blocking of access

File Size	The size of the log file in bytes (for example, 3264)

Some example CLF lines follow

```
10.101.45.177 – jerry [02/Jan/2002:09:26:23 +0800] "GET
http://musiccity.streamcastnetworks.com/desktop.htm HTTP/1.0" 200 916

10.101.45.177 – jerry  [02/Jan/2002:09:26:26 +0800] "GET
http://musiccity.streamcastnetworks.com/headertop.htm HTTP/1.0" 304 59

10.101.45.177 – jerry  [02/Jan/2002:09:26:35 +0800] "GET http://ads.musiccity.com/
HTTP/1.0" 200 2931

10.101.45.177 – jerry  [02/Jan/2002:09:26:35 +0800] "GET
http://musiccity.streamcastnetworks.com/images/header_lines.gif HTTP/1.0" 200 389

10.101.45.177 – jerry  [02/Jan/2002:09:40:21 +0800] "GET http://www.mp3netseek.com/
HTTP/1.0" 200 584

10.101.45.177 – jerry  [02/Jan/2002:09:40:23 +0800] "GET
http://www.mp3netseek.com/index2.html HTTP/1.0" 200 30133
```

Squid Proxy Cache Logfile Format

Squid by default stores all access attempts to the proxy server in a file called access.log using Squid logfile format . The format of the logfile entry is

```
time elapsed remotehost code/status bytes method URL rfc931 peerstatus/peerhost type
```

Time	A Unix timestamp as UTC seconds with a millisecond resolution. This is the time since the 1/1/1970.
Duration	The elapsed time considers how many milliseconds the transaction busied the cache. It differs in interpretation between TCP and UDP protocols used by Squid. For HTTP/1.0, this is basically the time between accept() and close(). For persistent connections, this ought to be the time between scheduling the reply and finishing sending it. For ICP, this is the time between scheduling a reply and actually sending it
Remotehost	The IP address of the requesting instance, the client IP address. The client_netmask configuration option can distort the clients for data protection reasons, but it makes analysis more difficult.
Code/status	Code is specific to Squid cache responses and contains information on the kind of request, how it was satisfied, or in what way it failed. Status contains the HTTP result codes with some Squid specific extensions

Bytes	The amount of data transmitted to the client
Method	The request method to obtain an object.
URL	The URL requested
Rfc931	The eigth column may contain the ident lookups for the requesting client.
peerstatus/peerhost	A code that explains how the request was handled, e.g. by forwarding it to a peer, or going straight to the source.
type	The content type of the object as seen in the HTTP reply header

Microsoft Proxy 2 Log file Format

Field Position	**Description**	**Field Value**
1	ClientIP	The requesting client's IP Address
2	ClientUserName	Windows NT Account of the user making the request. "Anonymous" will be recorded if Proxy Access Control is not being used.
3	ClientAgent	Contains the browser type sent by the client in the HTTP header. When Proxy Server is actively caching, the client agent is "MSProxy/2.0"
4	ClientAuthenticate	Indicates whether or not client has authenticated with the Proxy Server (Y/N)
5	LogDate	Timestamp indicating date when client request was made.
6	LogTime	Timestamp indicating time when client request was made. The LogTime is in local system time.
7	Service	Indicates which Proxy Server service the client is using.
8	ServerName	The local Proxy Server's Host name.
9	ReferredServer	If Proxy Server is used upstream in a chained

		configuration, this indicates the ServerName of the downstream server that sent the request.
10	DestHost	If Proxy Server is being downstream used in a chained configuration, this indicates the ServerName of the upstream server.
11	DestHostIP	The IP address of the destination. A hyphen indicates that the request was serviced from the Proxy Server's cache.
12	DestHostPort	The TCP or UDP destination port.
13	ProcessingTime	Connection time (in milliseconds). Begins when Proxy Server receives client request and ends when client closes the connection.
14	BytesSent	Number of bytes sent to remote host by Proxy Server.
15	BytesRecvd	Number of bytes received by Proxy from the remote host.
16	Protocol	Specifies the Application protocol used for the connection (HTTP, FTP, Gopher, Secure)
17	Transport	Specifies the Transport protocol used for the connection (TCP, UDP)
18	Operation	Specifies the Application method being used (GET, PUT, POST, and so on)
19	URL	Destination URL
20	MimeType	The Multipurpose Internet Mail Extension of the requested object. (Text/Plain, Image/GIF, and so on)

21	ObjectSource	Indicates final source of request object. • 0 = No information available • Cache = Source is the Proxy cache • Inet = Source is the Internet • VCache = Source is cache. Object was verified against original source • NVCache = Source is cache but original source could not be verified • VFInet = Source is Internet. Object was in cache, but original source was newer. • NotModified = Object retrieved from cache. Client performed an If-Modified-Since and object was found to be current • MemberObject = Source is the cache of another Proxy array member server • Upstream = Source is an upstream Proxy Server's cache
22	ResultCode	Status code of transaction • A value less than 100 is a Win32 error code. • A value between 100 and 1000 is an HTTP status code • A value greater than 10000 is a Winsock status code.
23.	CacheInfo	the Combined Hotfix or SP1 has not been stalled, this number is always zero (0). After the roxy 2.0 Combined Hotfix or SP1 has been stalled, this number reflects the cache status of e object, which indicates why the object was or

		'as not cached. The number is the bitmask result f the values described in CACHE_INFO_* efined in iaslog.h

APPENDIX B

Cyfin Categories

1. Agriculture

Description: Sites related to the Agriculture industry. Includes farming, ranching and related subjects.

Examples.

http://www.narqs.org

http://www.garden.com

http://www.admworld.com

http://www.usda.gov

http://www.farms.com

2. Auction/Classified

Description: Auctions and general classified related sites. Job related sites fall into the Employment category.

Examples:

http://www.klik-klok.com

http://ehammer.com

http://auctions.msn.com

http://dealdeal.com

3. Banners/Ads

Description: Sites containing electronic banner advertisements referenced from other Web sites.

Examples:

http://ad.doubleclick.net

http://ad.linkexchange.com

http://ads.lycos.com

4. Business Services

Description: Any type of service for businesses or corporations such as publishing, printing and general business information. Also includes office supplies, furniture, and equipment that is not IT related.

Examples:

http://www.cordis.lu

http://www.forrester.com

http://www.mainspring.com

http://www.ey.com

5. Chat

Description: Web-based chat groups pertaining to any subject.

Examples:

http://chatropolis.com

http://www.bridalchat.com

http://www.chatalyst.com

6. Construction

Description: Sites related to the construction industry.

Examples:

http://www.yops-wilkie.com

http://www.graybar.com

http://www.halliburton.com

http://www.mtarch.com

7. Cults

Description: Web sites with references to cults, paganism, satanism or other alternative religions.

Examples:

http://www.adf.org

http://www.thedojo.com

8. Dotted Decimal

This category has been depreciated. It does not provide any information about the content of a Web site.

9. Download Sites

Description: Web sites that specialize in a broad area of software downloads such as games, shareware, etc. MP3 sites are in the Entertainment category.

Examples:

http://pigeons.net

http://www.jumbo.com

http://www.topdownloads.net

http://www.oncemore.com

10. Drugs

Description: Information on the purchase and use of illegal or recreational drugs. Prescription drugs are in the Health and Medicine category.

Examples:

http://www.cleartest.com

http://www.blocpot.qc.ca

http://www.norml.org.nz

http://www.splif.com

11. Education

Description: K-12, colleges, universities, training, history and other education-related information.

Examples:

http://www.harcourt.com

http://www.marquette.edu

http://delta.ds2.pg.gda.pl

http://www.ufl.edu

http://www.ucf.edu

12. Email

Description: Free on-line mail and forwarding services.

Examples:

http://www.mail.at

http://www.mailcity.com

http://mail.excite.com

13. Employment

Description: Employment, job placement services, resumes, recruiting and employment resources.

Examples:

http://elance.com

http://www.ajb.dni.us

http://www.intercristo.com

http://www.monster.com

14. Entertainment

Description: Pertains to movies, television, music, theater, theme parks, fan clubs and other entertainment related information.

Examples:

http://www.subpop.com

http://www.Capitalfm.co.uk

http://www.virgin.net

http://www.film.com

15. Environment

Description: Environment and nature sites including air quality, natural resources, waste management, water quality and pollution. Also includes animal protection information.

Examples:

http://www.mountwashington.org

http://www.ndbc.noaa.gov

http://www.globe.gov

http://www.oil-spill-web.com

16. Financial

Description: Sites dealing with investments, discount brokers, banking, stocks, mortgages, bonds and other financial matters.

Examples:

http://www.ditech.com

http://www.sixer.com

http://401k.com

http://www.gmacfs.com

17. Food and Drink

Description: Restaurants, bars, pubs, recipes, grocery products, etc.

Examples:

http://www.tavolo.com

http://www.winetoday.com

http://www.publix.com

http://www.drpepper.com

18. Gambling

Description: Casinos, gambling, lottery, and sports advisory services.

Examples:

http://www.mariowins.com

http://www.flalottery.com

http://www.betandwin.com

19. Games

Description: On-line games, computer games, video games and other related information, and download sites that specialize in games only.

Examples:

http://www.pogo.com

http://www.chess.com

http://www.pcgamer.com

http://www.candystand.com

http://www.casesladder.com

20. Government

Description: Any government Web site that does not fall into a more specific category.

Examples:

http://www.gsa.gov

http://www.eren.doe.gov

http://www.dss.gov.uk

http://www.treasurer.state.ks.us

21. Hardware/Software

Description: Any Web site selling computer hardware or software, or computer related information. Also includes related on-line magazines.

Examples:

http://www.microsoft.com

http://www.hp.com

http://www.sunsoft.com

http://www.javasoft.com

http://www.gateway.com

http://www.insight.com

http://www.cisco.com

http://www.wavecrest.net

22. Hate and Crime

Description: Sites involving hate speech, terrorists, criminal activity, etc.

Examples:

http://www.privacyworld.com

http://www.digicrime.com

http://www.peacefire.org

http://www.2600.com

23. Health and Medicine

Description: Hospitals, self-help, pharmaceuticals, fitness, vitamins and general medical information sites.

Examples:

http://www.allegiance.net

http://www.allergan.com

http://www.baxter.com

http://www.pharmacia.se

http://www.thebody.com

24. Insurance

Description: Any site related to the insurance industry.

Examples:

http://www.statefarm.com

http://www.travelers.com

http://www.usfg.com

http://www.bestquote.com

http://www.datalife.com

http://www.compdent.com

25. Internet Services

Description: Sites related to provision of Internet or Intranet Services.

Examples:

http://www.via-net.net

http://www.tcinternet.net

http://www.iswest.net

http://www.giant.net

http://www.quest-net.com

26. Legal

Description: Courts, law enforcement, lawyers, prisons and other government and commercial organisations related to the law.

Examples:

http://www.copnet.com

http://www.townpolice.com

http://www.pursuittechnology.com

http://www.policeforce.org

http://campussafety.org

27. Manufacturing

Description: Sites related to traditional mass production of products not related to other categories herein. Also includes mining, petroleum, and gas-energy sites.

Examples:

http://active.boeing.com

http://www.exxon.com

http://www.ab.com

http://www.fele.com

http://www.motorola.com

28. Marketing

Description: Sites related to marketing and advertising and companies that provide these services.

Examples:

http://www.go-atomic.com

http://www.ama.org

http://www.glreach.com

http://www.mediametrix.com

http://www.admedia.org

29. Military

Description: Any official military site.

Examples:

http://www.defenselink.mil

http://www.navy.com

http://www.marines.com

http://www.militarycity.com

http://www.dtic.mil

30. News and Media

Description: Major news organisations and magazines that cover multiple topics.

Examples:

http://www.spectator.org

http://www.usatoday.com

http://www.cnn.com

http://www.pioneerplanet.com

http://www.freep.com

31. Non-Profit Organisations

Description: Any non-profit organisation such as the United Nations, AARP, Red Cross, etc.

Examples:

http://www.sparky.org

http://www.audubon.org

http://www.redcross.org

http://www.unitedway.org

http://www.wfnet.org

32. Personals/Dating

Description: People meeting other people, personal ads, single groups and mail order brides.

Examples:

http://couplesmagazine.com

http://www.singles.com

http://www.datingfaces.com

http://www.theromantic.com

http://www.blinddate.com

33. Politics

Description: Political advocacy of any type. Any site promoting or containing information on any political party, pro or con. Official government sites are in the Government category.

Examples:

http://www.thenation.com

http://www.anc.org.za

http://www.suck.com

http://www.rollcall.com

http://www.democracy.org.hk

34. Pornography

Description: Sexually explicit pictures, adult video sales, sex toys, prostitution, sex discussions, nudity and other sexually oriented sites. Non-pornographic but sex-related sites are included under health, medicine, society, and culture categories, depending on content.

Examples:

http://www.playboy.com

http://www.adultdreams.com

35. Public Proxy

Description: Any type of public proxy that allows an individual to "hide" the actual Web activity, such as anonymizer.com.

Examples:

http://www.anonymizer.com

http://anonymizer.egroups.com

36. Real Estate

Description: Housing, property management, rentals and other real estate industry related topics.

Examples:

http://www.apartments.com

http://realtor.com

http://www.eraonline.com

http://moving.com

37. Reference

Description: Catalogs, dictionaries, libraries, maps and other reference publications.

Examples:

http://www.mapquest.com

http://yp.bellsouth.com

http://www.pcwebopedia.com

http://www.elibrary.com

38. Religion

Description: Religious advocacy, pro or con. Cultist religions are in the Cults category.

Examples:

http://www.catholic.org

http://www.tricycle.com

http://www.jcn18.com

http://www.zchurch.com

39. Science

Description: Companies, services and information related to science.

Examples:

http://theory.uwinnipeg.ca

http://www.ksc.nasa.gov

http://www.sciam.com

http://www.chemsoc.org

http://www.kennedyspacecenter.com

http://cimewww.epfl.ch

40. Search Engines

Description: Internet search engines, search directories, and portals.

Examples:

http://www.yahoo.com

http://de.yahoo.com

http://www.google.com

http://search.about.com

http://www.linkguru.com

41. Shipping

Description: Sites related to the shipment of documents or merchandise.

Examples:

http://www.usps.gov

http://www.stickle.com

http://www.fedex.com

http://www.iacsf.com

http://www.ups.com

42. Shopping

Description: Retail, E-tail, E-commerce sites, flowers, books and other shopping related sites.

Examples:

http://www.oakley.com

http://www.guess.com

http://www.nordstrom.com

http://www.vans.com

http://www.jcpenney.com

43. Society and Culture

Description: Sites related to society, arts, culture, lifestyles and hobbies other than sports.

Examples:

http://www.oxygen.com

http://www.zonezero.com

http://www.20thcenturymasters.com

http://www.realart.com

44. Sports

Description: Sites containing information related to sports.

Examples:

http://www.espn.go.com

http://www.fieldandstream.com

http://www.sportsillustrated.com

http://www.golfonline.com

45. Tasteless

Description: "Hard-to-stomach" sites, that present offensive, worthless or useless information.

Examples:

http://www.celebritymorgue.com

46. Telecommunications

Description: Phone Companies, Wireless, Cable TV Systems, etc.

Examples:

http://www.1800collect.com

http://www.nextel.com

http://www.cwusa.com

http://www.aethersystems.com

http://www.attws.com

47. Transportation

Description: Mass transportation (including buses and trains), taxi services, bus schedules. Rental cars are in the Travel category.

Examples:

http://www.amtrak.com

http://www.gmemd.com

http://www.railtrack.co.uk

http://www.aar.org

http://www.csx.com

48. Travel

Description: Rental cars, airlines, cruises, travel agents, hotels and other travel industry related sites.

Examples:

http://www.hotels.fr

http://www.aircanada.ca

http://www.swissair.com

http://www.brusselsairport.be

http://www.hertz.com

49. Unsolicited or Push

Description: Infogate (formerly Pointcast), Microsoft Channels and Backweb technology, including vendors or this technology.

50. User Groups

Description: Internet User Groups.

Examples:

http://forums.delphi.com

http://messages.yahoo.com

http://www.supernews.com

http://www.wugnet.com

51. Utility

Description: Utility companies including petroleum products.

Examples:

http://www.cinergy.com

http://www.magin.com

http://www.scfm.com

http://www.cim.org

http://www.plexco.com

52. Vehicles

Description: Includes vehicle manufacturers, dealerships, suppliers, services and magazines, and any general information about this topic. Vehicles include automobiles, motorcycles and boats.

Examples:

http://www.ferrari.it

http://www.autotrader.com

http://www.carsdirect.com

http://www.leasesource.com

http://www.turbonium.com

53. Weather

Description: Information on weather and meteorology.

Examples:

http://www.meteo.fr

http://www.weather.com

http://weather.noaa.gov

54. Worthless

Description: Completely worthless sites that do not fall into any of the previous categories.

Examples:

http://www.stephen.com

http://www.thesquat.com

http://www.rocketcharged.com

55. Youth

Description: Sites targeting children. Sites selling children's toys are in the Shopping category.

Examples:

http://www.crayola.com/

http://nick.com/

http://www.nickuk.com/

http://www.spacecoastteens.com/

http://www.hotwheels.com/

56. Local

Description: Sites focused on geographically local events and topics.

Examples:

http://www.lvrj.com

http://kozmo.yakima.net

http://www.staugustine.com

http://www.virtualvenice.com

http://www.peru.com

57. Multimedia

Description: Any URL that is not a Web page but an image (.jpg, .gif, etc.), video or audio.

58. Downloads

Description: Any URL that is not viewed in the browser, but is actually downloaded to the user's PC or workstation. Downloads could be programs (.exe), zip files (.zip), Unix tar files (.tar), Unix compressed files (.gz or .Z) or movie peg files (.mpg).