Selection of an Internet Content Filtering Solution using the Analytic Hierarchy Process

Tim D. Hunt
Heather Wickham
Kathy Murphy
Michael Elrick
The Waikato Institute of Technology,
Hamilton, NZ
tim.hunt@wintec.ac.nz

Proceedings of the 16th Annual NACCQ, Palmerston North New Zealand July, 2003 (eds) Mann, S. and Williamson, A. www.naccq.ac.nz

ABSTRACT

This talk describes the selection of an Internet Content Filtering solution suitable for the specific requirements of the Waikato Institute of Technology (Wintec). Product data was collected from a variety of sources including: vendor product datasheets, industry benchmark tests, the experiences of other institutions and the academic literature. The available solutions were compared using the Analytic Hierarchy Process (Saaty 1980), a multicriteria decision support tool, using the above data and priority weightings determined for each criterion.

Keywords

Analytic Hierarchy Process, SPAM, virus, email, Internet content filtering.

1. INTRODUCTION

The issues of Internet Content Filtering (ICF) are an increasing concern for industrial and educational organisations that wish to limit their volume of non-work related Internet traffic (Burke 2002). Although a large source of this traffic is due to third parties such as SPAM and virus creators, the tendency of both employees and students to use the Internet for non-work related activities exacerbates the situation. The Internet traffic generated can be a real cost to organisations in terms of paying for the traffic, but it can also cause issues such as: virus attacks, time wasted dealing with SPAM email, lost productivity due to non work related browsing and increased down load times for legitimate work. There is also the concern for organisations to provide a safe working / study environment that does not subject staff or students to 'passive' viewing of objectionable material: This portrays the organisation as being unprofessional and potentially gives grounds for grievance actions by staff or students. There are numerous potential solutions for these issues, each with different methods of implementation and cost/maintenance tradeoffs.

2. WINTEC ICF REQUIREMENTS

Peak Internet traffic levels (Kb/s) at the Waikato Institute of Technology (Wintec) are routinely in excess of the 1 Mbps level guaranteed by our network provider under the current contract (Figure 1a). It is not expected that Wintec will fund an increase in guaranteed traffic and so options for reducing the traffic levels are being assessed to avoid a possible loss of service quality due to provider enforced maximum traffic. Figure 1b shows the average hourly International traffic rates and it is clear that on a per hour basis, the traffic is mostly less than that guaranteed. Analysis of Internet traffic indicated that reductions could be achieved by reducing non – work based activities and unwanted emails (SPAM and emails caused by or containing viruses). Although some manual administration of content currently occurs, it is a time consuming process and cannot be performed at a level that achieves significant Internet traffic reduction.

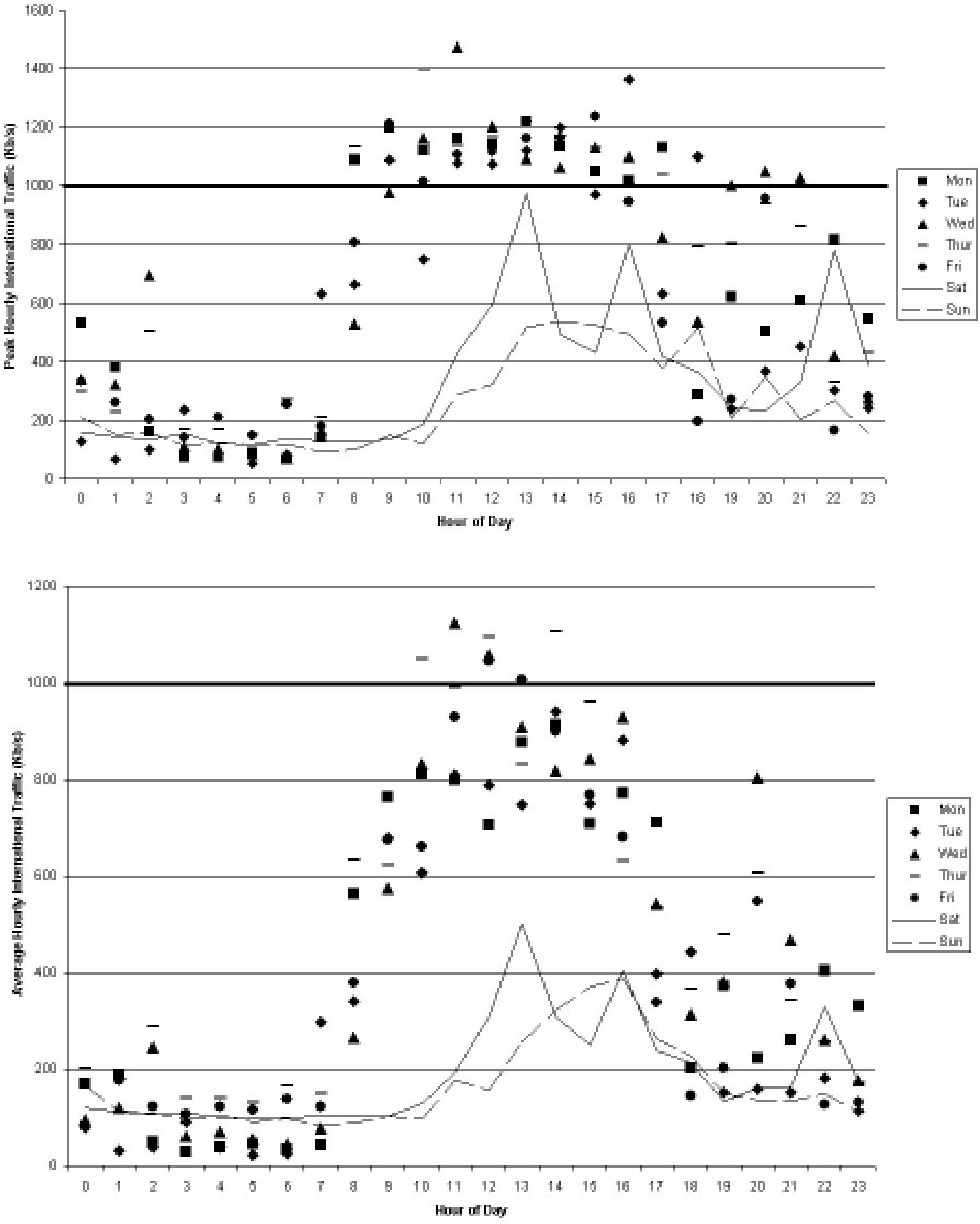

Figure 1a shows the peak International traffic (during each hour) versus hour of day, for a randomly chosen week in May 2003. The solid horizontal line indicates the purchased level (1000 Kb/s) and it can be clearly seen that for the majority of the working week day, the traffic peaks are greater than that guaranteed by our network provider. Figure 1b shows the same data as Figure 1a but averaged on a per hour basis.

2.1 INITIAL CONDITIONS AT WINTEC

At the start of this work Wintec already had in place, a number of ICF strategies, which are described in the following sections. Figures 2a and 2b give simplified views of the Wintec Network. When considering the overall strategy of ICF these diagrams provide a means of planning a system that considers all points of access, infection and control.

2.1.1 ANTI VIRUS PROTECTION

Command Anti-Virus for Windows (Command Software Systems 2003) is installed on individual workstations and the virus definitions are automatically updated when users log on to the Novell Network, by accessing an MSWindows Server that in turn obtains new definitions from the Command Software FTP Server. Server versions of Command Anti-Virus are installed on Wintec servers and are also updated on a regular basis. Wintec has a small number of Apple Mac computers and they do not have anti virus software installed.

2.1.2 NON WORK RELATED HTTP TRAFFIC

Figure 2a shows the location of the two (Staff and Student) Novell BorderManager (NBM) 3.6 firewalls (Novell 2003) which provide firewall, proxy and cache services. NBM also included the CyberPatrol content database, for url blocking but this was subsequently replaced (by Novell) with the SurfControl Content database (SurfControl 2003) during the course of this work.

2.1.3 SPAM - UNSOLICITED EMAIL

There are currently ongoing discussions at a political level to try to reduce the effects of SPAM (see Metz 2003 for a USA perspective), but regulators have not yet agreed on a definition of what SPAM is and is not.

At Wintec SPAM is managed using a black list that filters emails based on the senders address. This process is time consuming as the email addresses of SPAM distributors is continually changing in an attempt to keep ahead of the system administrators black lists. No email filtering based on email subject or body text is being performed.

2.1.4 WEB PAGE POPUPS

Web page Popups are becoming an increasing nuisance when viewing web pages. Wintec has no data on the quantity of traffic that they create and no control of web page Popups is being carried out. However, popups may become a significant issue and products do exist that aim to reduce their prevalence.

2.1.5 COMPUTER USE POLICY AND ETHICAL ISSUES OF ICF

Wintec's computer use policy covers issues such as deliberate introduction of software and hardware and deliberate harmful behaviour. Intentionally accessing objectionable material and deliberately wasting computing resources is not acceptable use of computer facilities.

Deciding what is and what is not objectionable material is a subject of much debate in the literature. Wintec uses the definition given in Section 3 of the Films, Videos and Publication Classification Act (1993). This definition does not cover what many staff may consider material that would be objectionable when viewed in a place of work.

3 INITIAL THOUGHTS FOR OVERALL STRATEGY

There are a number of possible implementations of ICF that might be applied. This section gives the overview of what they might include.

3.1 ANTI VIRUS PROTECTION

The current level of anti virus software is not considered suitable, although at the time of implementation it was seen as a sensible choice with the limited budget that was available and viruses were not propagating at the rate that they currently do. In theory installing anti virus products at the following physical/logical locations should provide a high level of protection:

♦ SendMail –All Internet email passes through this email product. As the Internet is seen to be the most likely source of viruses, an anti virus product here should stop the majority of email viruses entering the Wintec network.

♦ GroupWise – This is used as the staff email system. Email that contains a virus can be passed between staff without being intercepted by a SendMail anti virus product.

♦ Novell Border Manager – Prevent HTTP and FTP based viruses reaching workstations.

♦ Servers – All servers are liable to virus infection and so all servers (including the SendMail, GroupWise

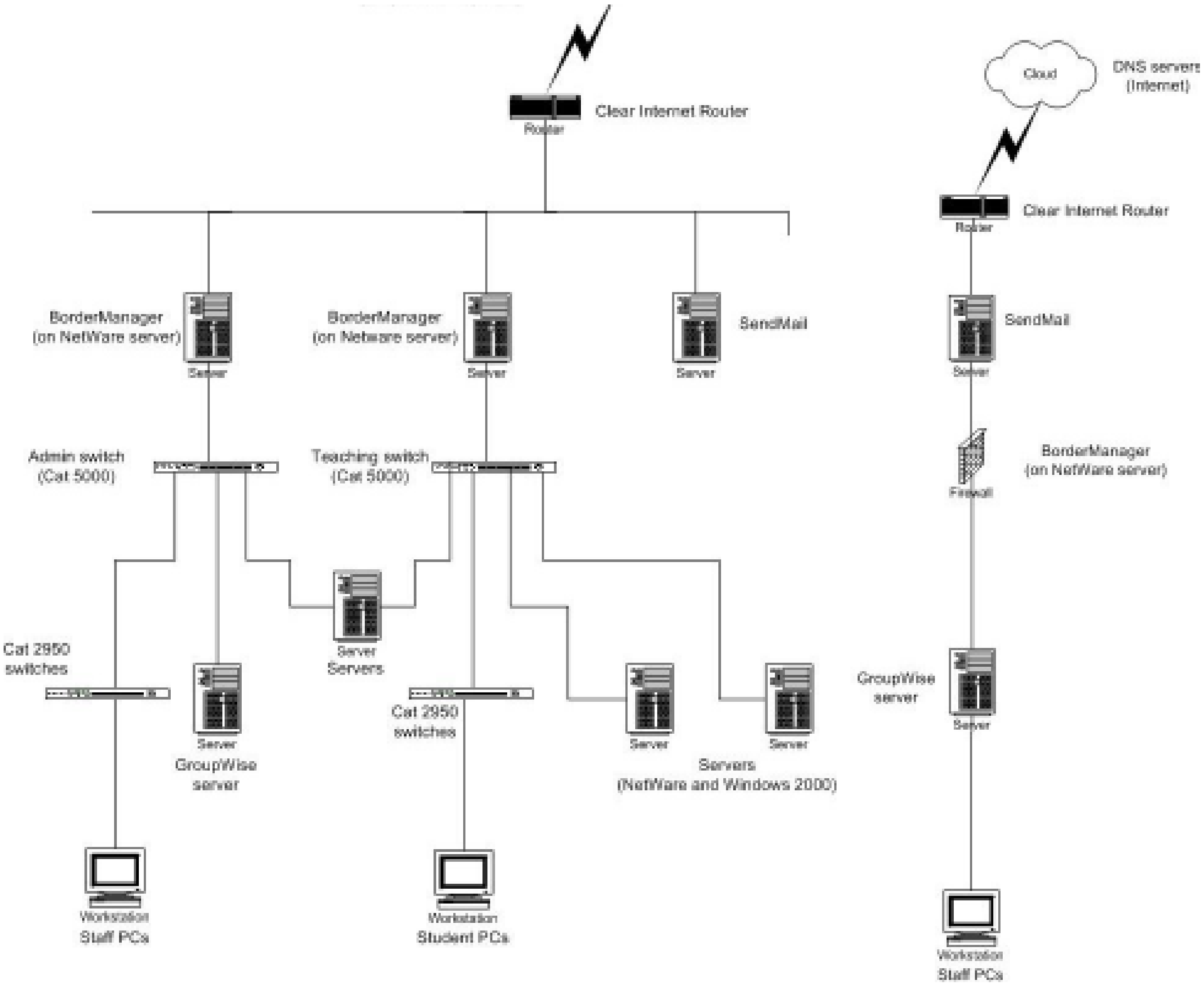

Figure2. A simplified view of the Wintec network a) Physical and b) Logical. Knowledge of the network assists with the ICF design by identifying points of access, infection and control.

and NBM servers) should have an anti virus product installed. Server platforms include: Novell, Unix and Windows2000.

Workstations – These should continue to have virus protection software to intercept viruses introduced via floppy disks, CDROMs etc.

Currently students use the Pegasus Mail (2003) email application hosted on a NetWare server that could be protected by a NetWare anti virus product. However, it is planned to migrate this service to Novell NetMail XE (Novell 2003b) that includes anti virus support.

3.2 NON WORK RELATED HTTP TRAFFIC

There are numerous products available for blocking access to urls and in particular the latest version of NBM now includes the N2H2 Internet Filtering solution (N2H2 2003) that uses the SurfControl Content database. Implementing this software fully may well provide sufficient control of HTTP traffic.

Charging traffic back to the end user (or their department) has been shown as an effective measure of reducing HTTP traffic for at least one other institution [Travaglia, personal communication]. An intermediate step to charging is to make line managers aware of the traffic volume and sites visited by staff. Charging students for Internet traffic (above a sufficient quota deemed suitable for completing studies) may also be a suitable option for reducing traffic volume.

3.3 SPAM

There are two main strategies for reducing SPAM. 1) Refuse email from 'known' SPAM sources using third party lists, 2) Use software that looks at each email and using various rules makes a judgement about whether it is SPAM or not. Once this judgement has been made, the email can either be blocked from going to the users account, or it can be tagged with a value describing the likely hound of it being SPAM. Users can then set up a rule in their email software that delivers all email over a certain SPAM value to their SPAM email box.

Wintec would prefer to give staff and students some control over what they regard as SPAM as this will hopefully avoid the problems of blocking non SPAM (HAM) email by mistake. There are also numerous anti SPAM solutions available and the latest version of Novell GroupWise has introduced limited anti SPAM protection and so this will be considered as a possible solution.

3.4 NETWORK APPLIANCES

Hardware products are available that perform a number of ICF functions and can be installed between the Internet Router and the internal network. These products promise the simplicity of just plugging in a single device that can perform a high level of ICF. They will not be able to deal with internal issues such as viruses being introduced via Workstations, or will they have knowledge of user preferences/privileges that products that can access Wintec's Novell Directory Service (NDS) might utilise.

3.5 OUTSOURCED SOLUTION

It is possible to have all Internet traffic analysed by a third party, who provide the range of ICF services. This considerably reduces the support required to provide an ICF solution, however, like the Network Appliance solution, it does not deal with internal network sources of viruses and cannot be configured with individual preferences.

4 METHOD

4.1 DATA COLLECTION

Data has been collected from 1) Commercial and non-commercial product information and web sites, 2) Vendor presentations, informal telephone interviews of key staff at a number of New Zealand educational institutions, library searches of academic and non-academic materials.

The problem of finding the optimum ICF solution is easily split into sub-problems as outlined in section 2 above. Students taking the 3rd year paper 'Introduction to Research Methods' on the Bachelor of Information Technology Degree at Wintec were each asked to choose one of these sub-problems as a topic for their major assignment. The authors were then able to extend this work and bring each of the sub-problems back together.

Two of the authors are Wintec network engineers and so have a detailed knowledge of the Wintec network and the requirements of a suitable ICF solution: This knowledge is a critical component of the data analysis process described in the next section.

4.2 DATA ANALYSIS

The task of choosing the appropriate ICF solution is complex due to the large number of potential candidates and is essentially a decision support process. The Analytic Hierarchy Process, AHP, (T.

L. Saaty 1980) is a popular method for making complex decisions and was chosen for this study.

The AHP has been embedded into the decision support software called 'Expert Choice' and the book 'Decision by Objectives' (Forman 2003) on the Expert Choice web site gives a good introduction into the decision making process using AHP and Expert Choice. The AHP is a multicriteria decision making technique that allows you to incorporate both objective and subjective factors. For example, suppose a decision was made to purchase a car; the criteria for comparing are first chosen e.g. price, engine size and interior design. Two cars at a time are then compared on each of the identified criteria and the criteria are also ranked in order of importance by comparing two criteria at a time. The end result is a ranked order of cars that reflects how individual criteria compare and how important each criterion is to the decision. It is then possible to 'defend' the decision and if required revisit the weightings that are placed on each factor.

4.2.1 WEIGHTING OF CRITERIA

At the start of the project, broad ICF requirements were proposed and discussed and this was used to guide the search for suitable software and solutions and enable 'intelligent' questioning of vendors at a number of vendor presentations.

5. RESULTS

Results of the data analysis will be presented at the conference and will also be published in full at a later date.

ACKNOWLEDGEMENTS

The authors would like to thank the following students who through their 'Introduction to Research Methods' course assignments have made valuable contributions to this paper: David Dutiyabodhi, Richard Haines, Asher Lucus and David Speake.

REFERENCES

Burke, B., Christiansen, C. and Kolodgy, C. (2002), "Worldwide Secure Content Management Software Market Forecast and Analysis, 2002-2006: Vendor Views" IDC Bulletin 27299: http://www.idc.com

Command Software Systems (2003), Accessed May 2003: http://www.commandsoftware.com/index.cfm

CyberPatrol (2003), Accessed May 2003: http://www.cyberpatrol.com/how_to_buy/business.aspx

Films, Videos and Publication Classification Act (1993), Accessed May 2003: http://rangi.knowledge-basket.co.nz/gpacts/public/text/1993/se/094se3.html

Forman, E., (2003), "Decision By Objectives (How to convince others that you are right)", Accessed April 2003: http://www.expertchoice.com/dbo/

Metz, C (2003), "SPAM is getting worse", Accessed May 2003: http://www.pcmag.com/print_article/0,3048,a=41379,00.asp

N2H2 (2003), Accessed May 2003: http://www.n2h2.com/index.php

Novell (2003), "Novell BorderManager 3.7", Accessed April 2003: http:// http://www.novell.com/products/bordermanager/

Novell (2003b), "Novell NetMail XE", Accessed May 2003: http://www.novell.com/products/netmailxe/

Pegasus Mail (2003), Accessed May 2003: http://www.pmail.com/

Saaty, T.L. (1980), "The Analytic Hierarchy Process", McGraw-Hill, New York.

SurfControl (2003), SurfControl Web Filter for Novell BorderManager" Accessed April 2003: http://www.surfcontrol.com/products/web/novell/