Securing Web Access

By

Matt Shea

Submitted to
the Faculty of the Information Technology Program
in Partial Fulfillment of the Requirements for
the Degree of Bachelor of Science
in Information Technology

University of Cincinnati
College of Applied Science

May 2009

Securing Web Access

by

Matt Shea

Submitted to
the Faculty of the Information Technology Program
in Partial Fulfillment of the Requirements for
the Degree of Bachelor of Science
in Information Technology

_____	_____
Matt Shea	Date
_____	_____
Mark Stockman, Professor	Date
_____	_____
Hazem Said, Ph.D., Department Head	Date

Acknowledgements

I would like to thank all of the professors from the IT department for their wisdom and guidance throughout the Senior Design process. A special thanks goes to Professor Stockman for being such a great technical advisor. Special thanks goes out as well to the previous co-op employer that has allowed me to work with them on this project.

Table of Contents

List of Figures

Abstract

Securing Web Access is a networking project that will allow a small-to-medium sized business setup a cost effective Web proxy solution. The project makes use of entirely open-source software packages designed to run on the Linux operating system. These packages include Squid Web Proxy Server, SquidGuard content filter, and MySQL Squid Access Reports. These tools will allow a company to monitor users' Internet access as well as restrict certain Web sites and file types from being accessed and downloaded that are non-business related. The system will be designed to run in a transparent operation mode meaning that no configurations will need to be made on the client machines. This will make for minimal administration required by the IT staff, but at the same time will enhance the security of the overall network.

Securing Web Access

1. Problem Description

The Information Technology Department at a local accounting firm employing approximately 60 workers currently does not have an efficient way to restrict Internet access to its users. The IT department currently uses an 800 series Cisco router as the main gateway and firewall to connect to the Internet. This device has the capability to block user-defined ports, but more granular control is needed. The ability to block non-business related Web content, in addition to certain file types, is needed to better secure the company's network.

The problem at hand is two-fold. First, there is no current way for the IT staff at the firm to monitor Internet bandwidth usage (1). In an accounting and auditing firm, working remotely from clients' offices plays a vital role in the way the company operates on a daily basis. For this to occur, employees need to have the ability to login via Terminal Server to access internal files. This process requires a substantial amount of bandwidth to work efficiently with minimal lag-time; therefore, proper utilization of the company's Internet resources by employees is a must to keep costs down.

With a system in place that allows the IT department to monitor Internet usage on a per-user basis, the firm will be able to develop policies to help them better utilize their current bandwidth. This system will also allow them to make better decisions when they are considering upgrading their Internet connection. As of now, the IT department does not have an intelligent way to make this decision, so they could potentially make a purchasing decision based on false information.

The second problem the IT department at the firm faces is they do not have a way to block certain types of Web content (1). They would like to have the ability to block content that is not related to the operation of the business, in addition to blocking certain file types for security reasons.

First, blocking content that is not business related will help the firm better utilize its Internet connection as mentioned above. The IT department is currently aware of Web radio and streaming video as some of the content types that are being abused. Analyzing the reports generated from monitoring Web traffic will allow them to identify additional content types that are being misused by employees, and potentially restrict access in the future.

Secondly, blocking certain file types from being downloaded will greatly increase the security of the firm's network. Currently, any file can be downloaded and opened from a user's local machine. Having the ability to block files ending in .exe, ActiveX controls, and Java Applets, in addition to others, will greatly enhance the security of its network.

Web mail is another security liability the firm faces, as it must adhere to various government privacy guidelines and regulations. Currently, employees have access to their personal e-mail accounts from their work computers, therefore allowing them to send and receive confidential information that cannot be traced. Having the ability to restrict Web mail access will allow the company to better adhere to privacy regulations, and ensure that confidential information is kept safe.

2. Product Description and Intended Use

This project provides the firm a solution to these current issues. The system that has been developed is made up of three main software components including the open source Squid proxy server, SquidGuard content filter plug-in, and a reporting tool utilizing a MySQL database backend.

Proprietary alternatives have been researched, including Microsoft's Internet Acceleration Server and Websense content filtering software (5, 11). However, due to the company's relatively small size and limited budget, the open source projects mentioned above will be used to meet its needs.

2.1 Design Protocols

As mentioned before, the system that has been developed makes heavy use of open source software applications and an open source operating system. The main server utilizes the CentOS operating system, which is based on the Red Hat Enterprise Linux distribution. This operating system was chosen as it is the current standard which the accounting firm uses for the majority of their Linux servers. It is also becoming much more popular among Linux enthusiasts with its relatively small form factor and availability for commercial support.

Squid proxy server has been installed to run on top of CentOS and serves as the core of the system. Additional plug-ins that have been implemented include the SquidGuard content filtering software and MySQL Squid Access Report. SquidGuard content filtering software is used to set up complex access control lists, utilizing white and black lists to determine which Web traffic to allow or block. MySQL Squid Access

Report is a reporting tool that makes use of the popular open source database MySQL. It comes with a fully functional Web front-end to display the reports in a user-friendly manner.

In order for these software applications to work, there were a few prerequisite packages that needed to be installed on the server including Apache Web server, the PHP scripting language, and the MySQL database engine. The reporting tool relies heavily on all three of these packages in order to function properly.

2.1.1 Network Diagram

The overall network architecture for the system can be found in Figure 1 shown below.

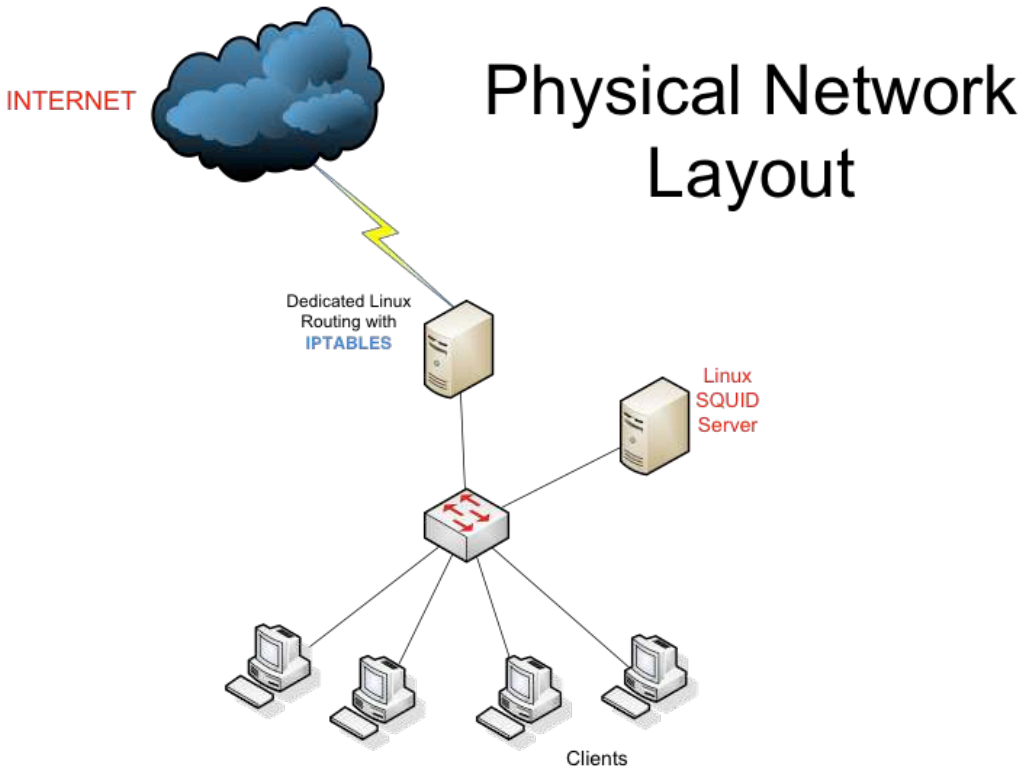

Figure 1: *Network Architecture*

As seen from the network diagram in Figure 1, all Web requests from client machines are directed to the existing gateway router as normal. At this point, the router will examine the port of the outgoing request, and if it is on port 80 or 443, will redirect the request to the proxy server. The proxy server will then process the request, log it, and if not explicitly defined to block the request, will forward the request back to the gateway router. From here, the router will allow the traffic to leave the network, only if it has first gone to the proxy server. The routing of Web traffic has been configured this way so that no configuration changes need to be made on the client machines. Otherwise, the gateway device on each client would need to be manually set to the proxy server. Requests on ports other than port 80 or 443, such as FTP requests on port 21, will leave the network directly by the gateway router as normal, unless they are set to be blocked by the built-in firewall on the router.

Figure 2 below shows the order in which a Web request is made. The origin of the request comes from a client machine with its destination on the Internet.

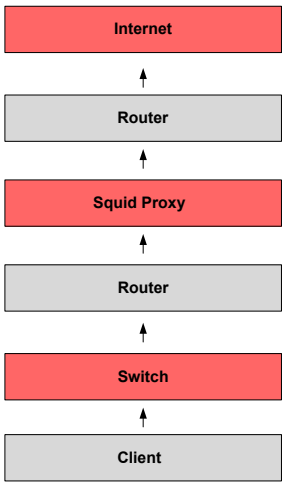

Figure 2: *Web Request Data Flow*

2.2 User Profiles

There are two main user groups of this system, primarily the IT staff at the accounting firm. The accounting and administrative staff will be indirect users of the system as well. The descriptions below describe each user profile in more detail.

2.2.1 Administrators

The primary users of this system are the IT staff members at the accounting firm. Once the system is moved into the production environment, they will be responsible for monitoring the daily reports that are generated by the reporting tool and up-keeping the Website white and blacklists. Since the customizations for these tools are done through Linux configuration files, initial training has been done to ensure familiarity with the tools.

2.2.2 Accounting Staff

Employees of the firm will not be directly using the proposed system, but they will be impacted by it. Documentation will be developed by the IT department to inform employees of new policies and procedures that will be in place as a result of the new system.

3. Deliverables

Listed below are the deliverables that have been completed for this project. The list fulfills both the requirements for the Senior Design course and for the project advisor at the accounting firm.

1. Install/Configure CentOS v 4.7 Virtual Machine
2. Set up 2 – Windows XP Client Virtual Machines
3. Install/Configure Squid Proxy Server
4. Install/Configure MySQL Squid Access Report
5. Install/Configure SquidGuard Content Filter
6. Write Script to translate IP Address to MAC Address based on external data sheet
7. Set up Linux machine as a router to forward all Web requests from clients to the proxy server

The deliverables listed above have been completed in entirety throughout the three quarters of the Senior Design sequence.

4. Development

4.1 Timeline

Figure 3 displays the timeline of tasks that will be completed by the end of Senior Design III. The dates represented start at the beginning of Senior Design II through the final presentation for Senior Design III.

Starting Week of	Task
1/18/09	Install/Configure Server OS on VM
1/25/09	Set up 2 XP Client VM
2/1/09	Install Squid Proxy Server
2/15/09	Install SquidGuard Plug-in
2/15/09	Install MySQL Reporting Plug-in
2/15/09	Design Freeze - First Draft
2/22/09	Write MAC address To IP address Script
3/8/09	Give Prototype Presentation
3/8/09	Set up Linux Router
3/29/09	Testing
4/5/09	Write Final Report
5/3/09	Present at Tech Expo
5/17/09	Give Final Presentation

Figure 3 *Timeline*

4.2 Budget

The budget for this project is listed in Figure 4 below. Using mainly open-source tools and software products provided through UC licensing agreements, there has been no actual cost for this project.

Item	Description	Retail Cost	Incurred Cost
Server	Will be using virtual machines.	$3, 149.00	$0.00
VMware Server	Free	0.00	0.00
CentOS v. 4.7	Open Source	0.00	0.00
Squid Proxy Server	Open Source	0.00	0.00
SquidGuard	Open Source	0.00	0.00
MySQL Squid Access Report	Open Source	0.00	0.00
Windows XP	Provided by MSDN Academic Alliance	199.99	0.00
120 GB External HD	Already Have	129.99	0.00
	Retail Total:	**$3,478.98**	
	Incurred Total:		**$0.00**

Figure 4 *Budget*
Sources: (4,6,7,9,10)

4.3 Hardware and Software

Listed below are the hardware and software used to complete the project. All development of the system has been done in a virtual environment. After Senior Design III is completed, the virtual machines will be moved to the production environment at the accounting firm.

4.3.1 Hardware

- **External Hard Drive**: Used to back-up virtual machines.
- **Personal Laptop**: Used to build and test virtual machines.

4.3.2 Software

- **VMware Server**: Virtual hypervisor used to host virtual machines.
- **CentOS v. 4.7**: Base operating system.
- **Windows XP**: Client operating system used for testing proxy server.
- **Squid Proxy Server**: Open source proxy server used for the core of the project.
- **SquidGuard**: Content filtering plug-in for Squid proxy server used to restrict Web access.
- **MySQL Squid Access Report**: Reporting tool used to parse the Squid log file and neatly display information in Web browser.

5. Testing Plan

The majority of the testing for this system has taken place continually throughout the design and development process. After each piece of software was installed, testing was performed to ensure it was functioning properly. This included analyzing Linux configuration files and making changes to these as needed. All of the elements of this system have been built in a virtual environment, allowing for isolation from outside networks. Since this is a networking project, this is important to properly ensure outside environments are not affected by the system. The virtual environment can also easily be moved to production servers for implementation at the end of the project.

Before moving to the production environment though, load testing has been done on the server to ensure it will be able to handle the traffic load of an average business day at the accounting firm. This has been done by estimating how much traffic the firm currently uses, and then this traffic was simulated on the network with a tool called *Web Application Stress Tool* made by Microsoft. The tool was designed to test Web applications, but since it can simulate Web requests, it worked well in testing this project.

Security testing has also been conducted to ensure the server is well protected from outside threats. This includes ensuring all unnecessary services are turned off, as well as making sure best practices have been followed for each software package that is installed and running. Also, since Linux relies heavily on file and folder permissions for granting access to specific users, groups and services, extra precaution has been taken in setting up these permissions to ensure unauthorized access is not allowed.

6. Risk Management Plan

The risk associated with this project is relatively low since it is using all open source products. Since all of these software packages can be obtained for free, having the budget pulled or decreased is not an issue. The main threat associated with this project is the potential for the virtual machine images to become corrupt. To help mitigate this threat, multiple backups have been made on different storage mediums to ensure integrity of the virtual images.

Figure 5 shown below describes each threat for this project in detail and associates a risk level with each potential item. It also provides a risk mitigation plan to help minimize the level of each risk.

Risk Description	**Risk Level**	**Risk Mitigation Description**
Virtual machine images become corrupt.	Medium	Make at least one backup of each virtual machine onto a different storage medium. Preferably have at least two backups on separate storage devices for each machine.
Script to cross-reference IP Addresses with MAC addresses cannot be integrated with the MySQL Squid Access Report tool that is used to generate reports and display them in a Web front-end.	Medium	Identify the current way the reporting tool is gathering IP addresses and develop a new script that pulls information from this report.
Company that I'm doing the project for no longer has a need for it.	Low	Communicate status updates on the project with the company and make sure they are still on board with it.
Am unable to develop a working prototype to fulfill the requirements for Senior Design.	Low	Work with technical advisor and discuss any issues that arise as they happen.

Figure 5 *Risk Analysis*

7. Proof of Design

7.1 Squid Proxy Server

Installing and configuring Squid proxy server is a multi-step process. First, to install the Web proxy server on CentOS, the Yellow Dog Updater Modified (YUM) command can be used to install it on the system. This command will search the YUM configured libraries and download and install the package. YUM takes care of installing any dependencies that may be needed, which usually makes for a smooth install.

Once installed, the Squid engine may be started and stopped easily by using the Red Hat "service" command. This will also show the current running status of the service. Figure 6 below shows a screen shot from a terminal window of the Squid service running on the server.

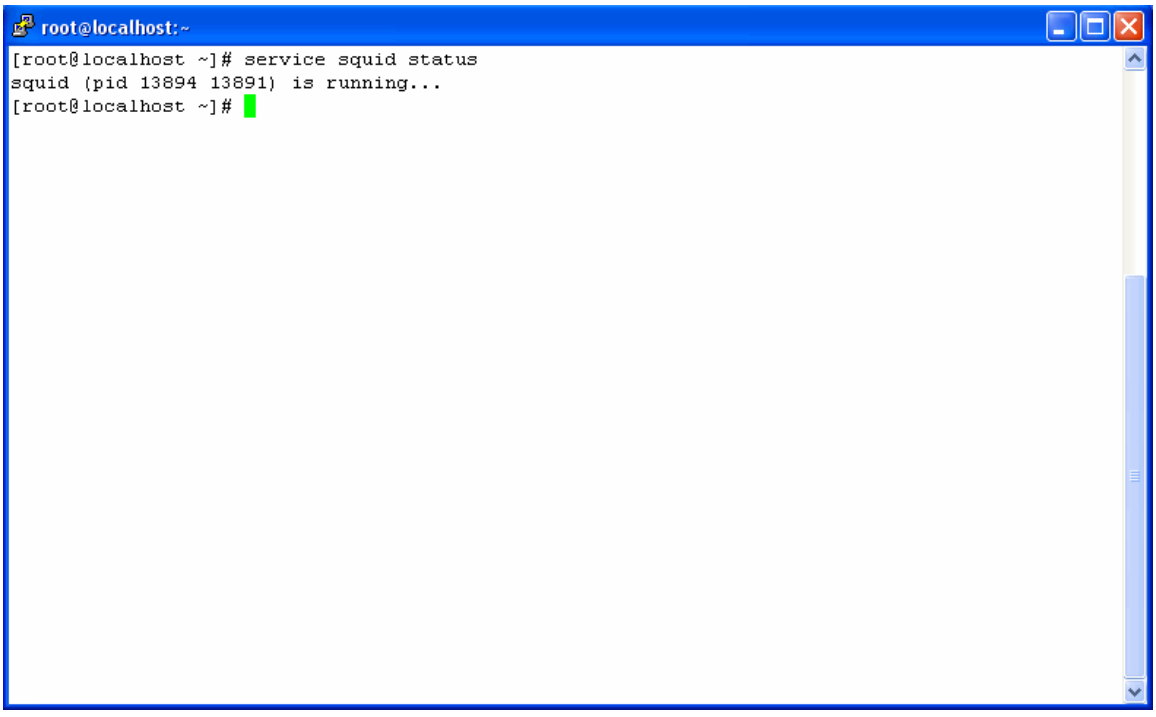

Figure 6 *Squid service running*

Once installed, Squid Web proxy server must be configured with its configuration file. This file named "squid.conf" can be found in the '/etc/squid/' directory on CentOS. This file holds information regarding hundreds of possible different configurations for the piece of software, but for this project many of the default settings will be left intact. One item that did need to be added to the file was a couple of lines to insure proper interfacing with the content filter plug-in. Figure 7 below shows a screenshot of part of the Squid configuration file.

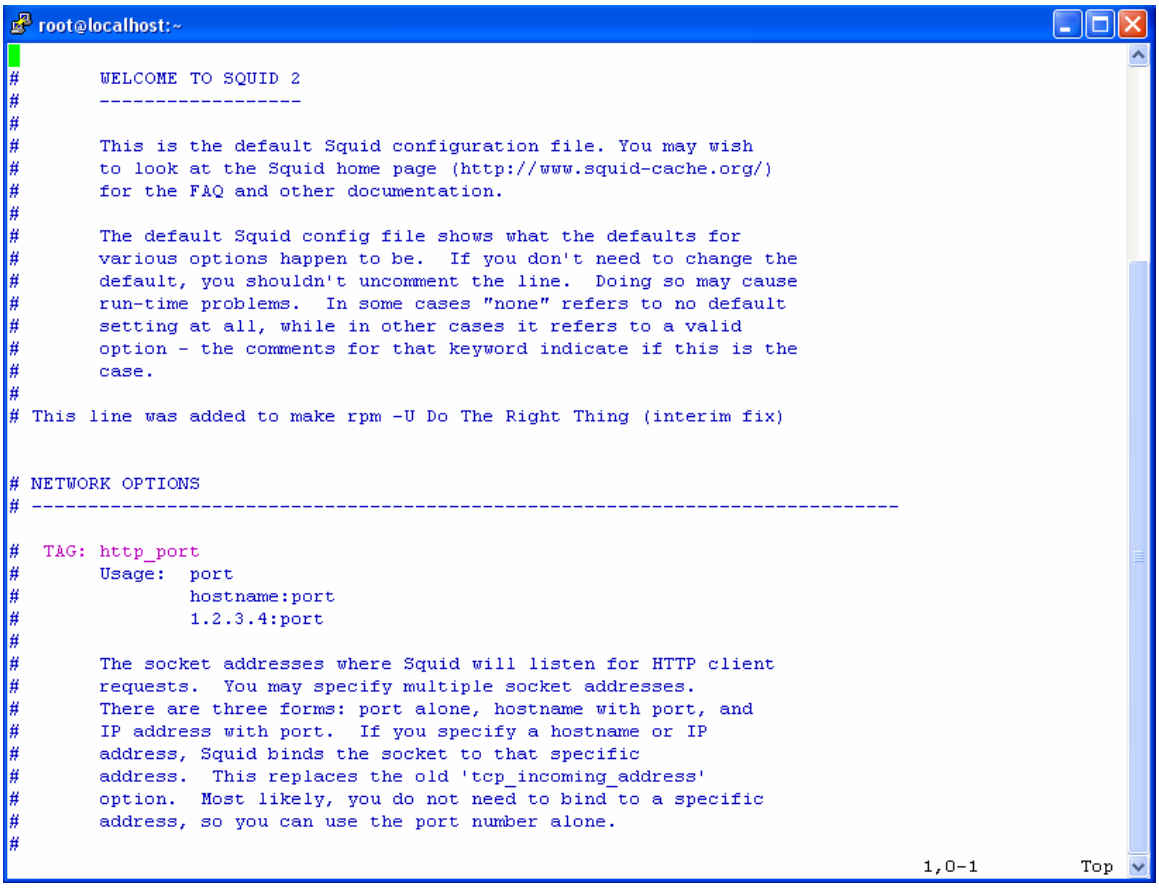

Figure 7 *Squid configuration file*

7.2 SquidGuard

The SquidGuard content filter works similarly to Squid itself in the sense that it utilizes a configuration file named 'squidGuard.conf' located in the '/etc/squid/' directory. This package was installed a little differently as it did not use the 'YUM' command but required the 'rpm' command. The source RPM file for the software was downloaded from SquidGuard's website and then installed using 'rpm –ivh'. This way of installing software packages is not as easy as using 'YUM' as it does not install all required dependencies. This was the case with installing SquidGuard and several of its dependencies had to be installed first, including the Berkeley Database libraries. Shown below in Figure 8 is a screenshot of 'squidGuard.conf'.

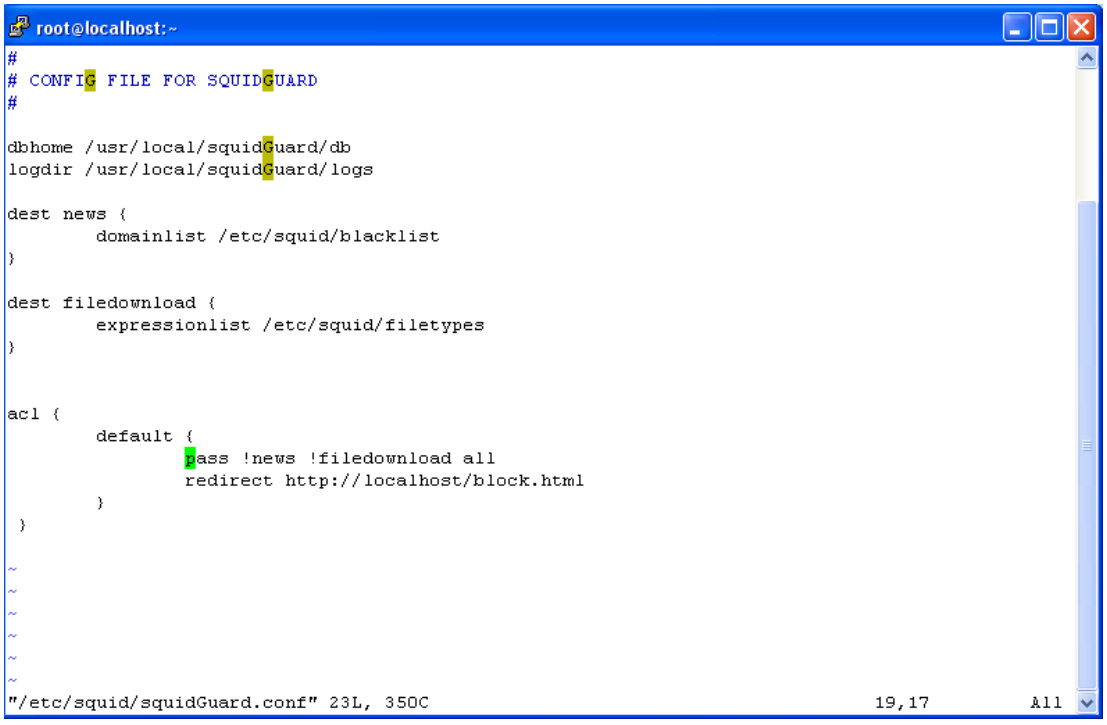

Figure 8 *SquidGuard configuration file*

As seen from the configuration file, there are two destination categories that are being blocked. The first group is *news* where all domains listed in the

'/etc/squid/blacklist' file will be blocked. The second group is *filedownload* where all file types listed in the '/etc/squid/filetypes' file will be blocked. Finally, the access control list is set up to pass all Web traffic except explicitly blocked destinations. If a site is blocked, the user will be redirected to the Web page 'block.html'. Figure 9 below shows an example of the page users will see when trying to access a restricted site.

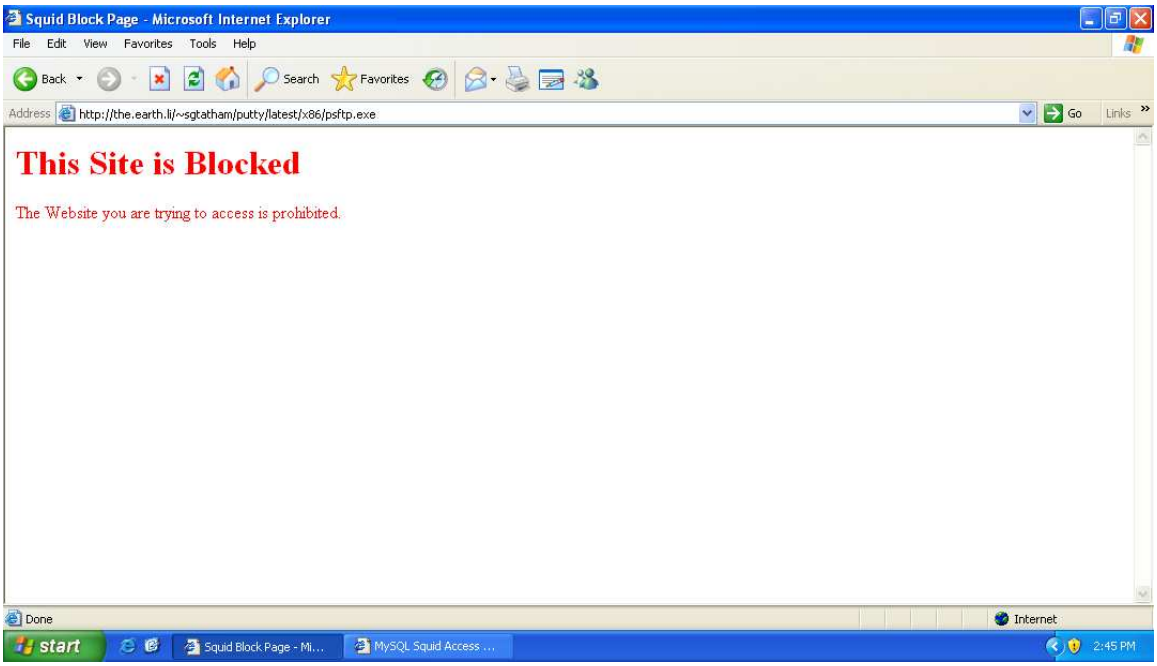

Figure 9 *Blocked Web page*

7.3 MySQL Squid Access Report

MySQL Squid Access Report is the reporting tool that will be used for this system to allow reports to be generated based on Web activity by end users. In order for this tool to work, three pre-requisites needed to be installed on the server. These include Apache Web Server, the PHP scripting language, and the MySQL database engine. These were

all installed using 'YUM' and the default configurations for each package are being used. Figure 10 below shows that these services are running on the server.

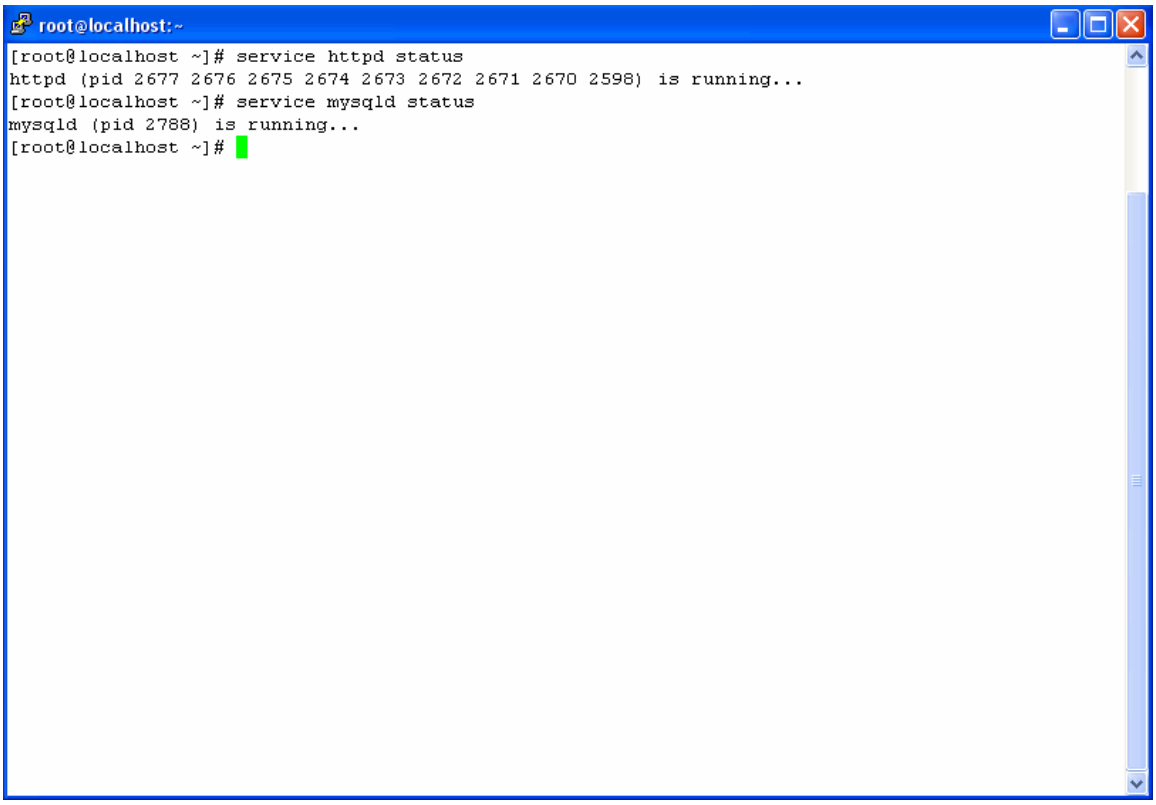

Figure 10 *Apache and MySQL Services Running*

To install MySQL Squid Access Report, the package was downloaded from the project's Website. From here, it needed to be de-compressed and placed in the '/usr/local/mysar/' directory. A symbolic link was formed from this directory to Apache's 'www root' directory. Next, the Web front-end of the tool could be reached by pointing to this directory from a Web browser. Once accessed, a wizard walks the user through the install process. This involves setting up the database and setting certain file

and folder permissions. Once the wizard is complete, the homepage for the tool can be reached through a Web browser. Figure 11 shows this below.

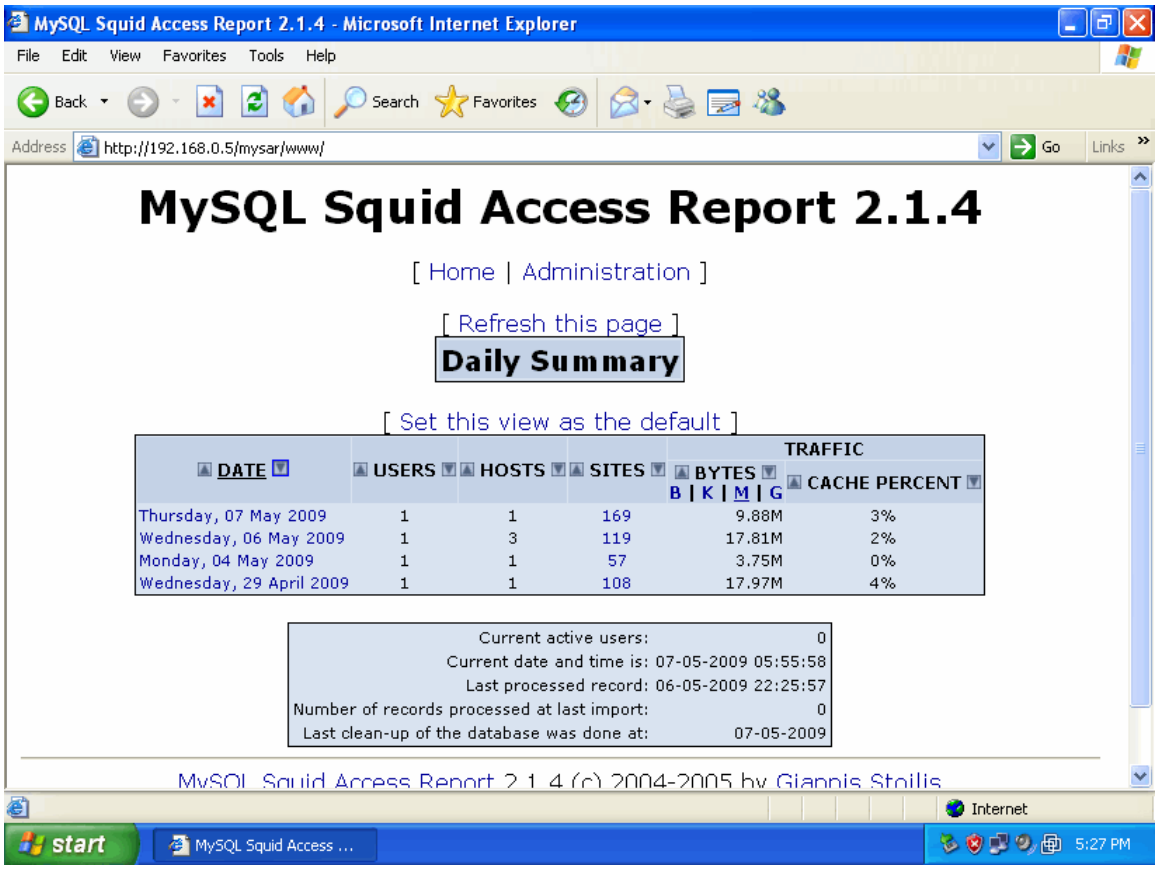

Figure 11 *MySQL Squid Access Report Homepage*

Once setup, MySQL Squid Access Report can be used to monitor Web usage broken up by several different criteria. From the homepage, the user is presented with a list of days to click on. Once clicked, a list of hosts by IP address is presented. By selecting an IP address from this list, all Web activity for that user can be seen. Figure 12 below shows a report broken down for a single host machine.

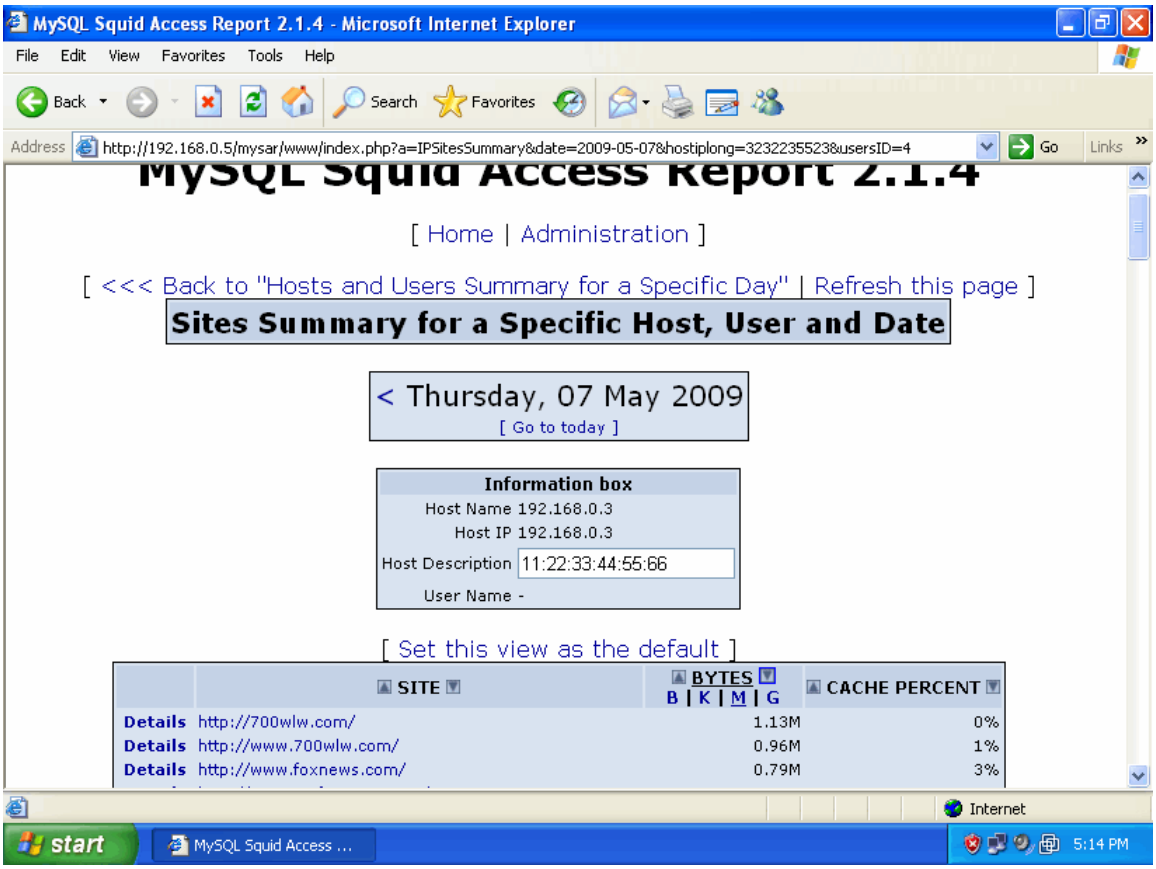

Figure 12 *MySQL Squid Access Report for Single Host*

As seen above, the host IP address is 192.168.0.3. In the 'Host Description' field, a MAC address has been populated. This is not done by default by the reporting tool, but a script has been written to automatically populate this field. The script queries the tool's database and does a mapping based on an external text document. This was done to help the IT staff better identify host machines within the reporting tool. Figure 13 below is a screenshot of the Perl script that was written to accomplish this mapping.

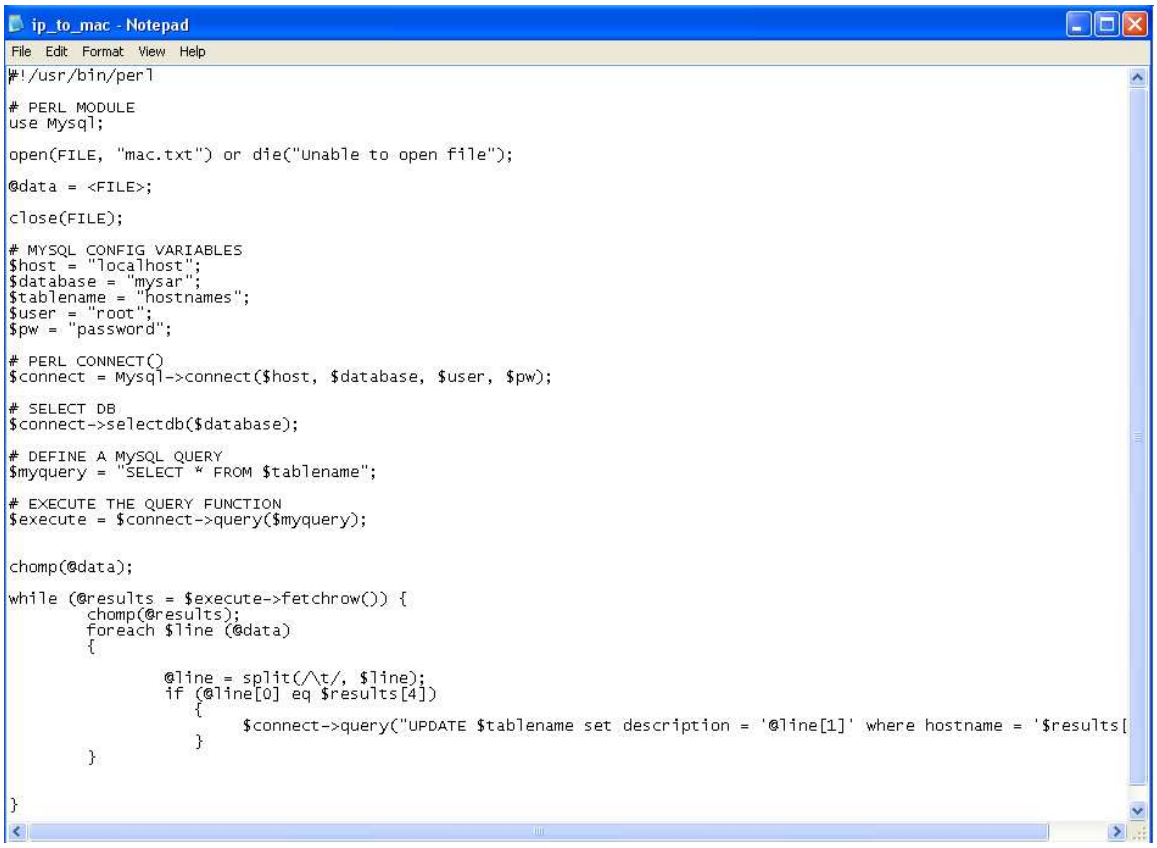

Figure 13 *IP to MAC Address Script*

7.4 Transparent Operation

The system has been designed to work in a transparent manner, meaning that no settings need to be added on the client machines. This has been accomplished by setting up a separate Linux server running iptables to act as a router. By setting up routing rules within iptables, all internal Web requests are forwarded to the Squid proxy server for processing. This ensures that all Internet traffic is being logged by the Squid server, and does not require editing settings on all 50 client machines. Figure 14 below displays the routing rules that have been setup on the Linux router.

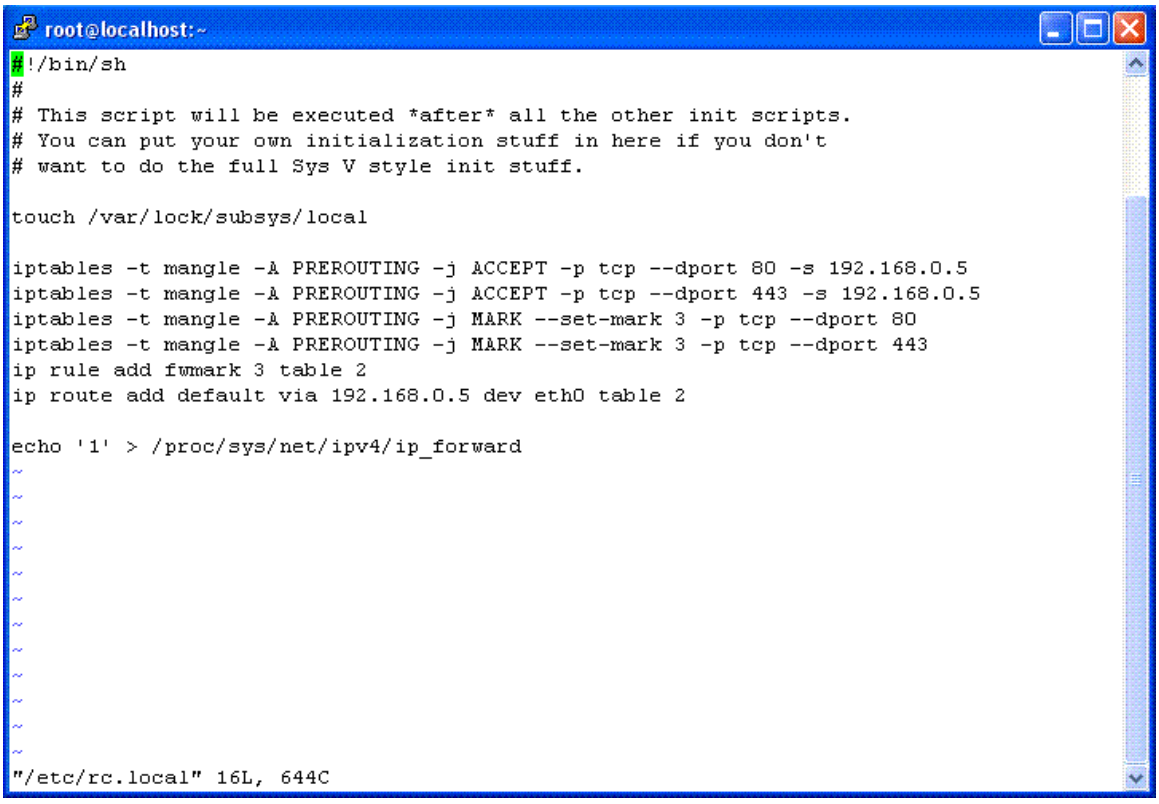

Figure 14 *Routing Rules*

8. Conclusion

The system built for this project will make it much easier for the IT staff at the accounting firm to monitor Web access. It will allow the company to better utilize its current Internet connection, in addition to putting controls in place to better secure its network. Administrators will be able to monitor employee Web access, and make more intelligent business decisions related to their Internet bandwidth usage with the system that has been developed.

References

1. Arnold, David. System Administrator, Cincinnati Based Accounting and Auditing Firm. Personal Interview. 29 Oct. 2008.

2. Fox, Tammy. "Squid Proxy Server Configuration". <http://www.linuxheadquarters.com/howto/networking/squid.shtml >. 3 Nov. 2008.

3. Gite, Vivek. "Linux: Setup a transparent proxy with Squid in three easy steps". <http://www.cyberciti.biz/tips/linux-setup-transparent-proxy-squid-howto.html>. 18 Nov. 2008.

4. HP. "Hewlett-Packard Website". <http://www.hp.com>. 18 Nov. 2008.

5. ISA. "Microsoft Internet Security and Acceleration Server Webpage". <http://www.microsoft.com/forefront/edgesecurity/isaserver/en/us/default.aspx>. 3 Nov. 2008.

6. MSAR. "MySQL Squid Access Report Webpage". <http://giannis.stoilis.gr/software/mysar/>. 16 Nov. 2008.

7. Microsoft. "Microsoft Corporation Website". <http://www.microsoft.com>. 18 Nov. 2008.

8. SARG. "Squid Analysis Report Generator". 16 Nov. 2008. <http://sarg.sourceforge.net>.

9. Squid. "Squid Web Proxy Server Project Webpage". <http://www.squid-cache-org>. 2 Nov. 2008.

10. SquidGuard. "SquidGuard Project Webpage". <http://www.squidguard.org/index.html>. 2 Nov. 2008.

11. Websense. "Websense Software Webpage". <http://www.websense.com/content/home.aspx>. 3 Nov. 2008.