

**Project Report on**

**Netgear Privacy and Advertising Platform**

**BY**

**Santhosh Kumar Pattabhiraman**

**008638110**

**Bhavin Niranjan Mehta**

**008267025**

**Submitted to the Office of Graduate Studies of**

**San Jose State University**

**In partial fulfillment of the requirements for the degree of**

**MASTER OF SCIENCE**

Department of Electrical Engineering

San Jose State University

San Jose, CA -95112

**Netgear Privacy and Advertising Platform**

**A Project By**

**Santhosh Kumar Pattabhiraman**

**Bhavin Niranjan Mehta**

**Submitted to the Office of Graduate Studies of**

**San Jose State University**

**In partial fulfillment of the requirements for the degree of**

**MASTER OF SCIENCE**

**Department of Electrical Engineering**

Approved as to style and content by

-------------------------------------

**Prof. Richard Sinn**

**(Graduate Advisor)**

-------------------------------------

**Dr. Nader. F. Mir**

**(Graduate Co-Advisor)**

----------------------------------------

**Dr. Tri Caohuu**

**(Graduate Coordinator)**

Santhosh Kumar Pattabhiraman Bhavin Niranjan Mehta

201 S 4th Street Apt # 634 3566, Dickenson Common

San Jose, CA 95112 Fremont, CA 94538

Ph: (408)-571-8906 Ph: (408)-796-8996

Email:santhoshmsedu@gmail.com Email:bhavin.m20@gmail.com

SJSU ID: 008638110 SJSU ID: 008267025

**Acknowledgements**

We would like to thank our advisor Professor Richard Sinn for guiding us through the course of the project. His continuous encouragement throughout the project gave us the impetus to complete the project. We would also like to thank Dr. Ahmed Hambaba, Associate Dean of Graduate and Extended Studies at San Jose State University for giving us the opportunity to work with Netgear Inc. and being supportive over the course of the project. Our acknowledgements to Dr. Jalel Rejeb for mentoring us and guiding us through the initial stages of our project.

We would also like to thank Cedar Milazzo, VP of Engineering at Netgear Inc. for guiding us and helping us to accomplish this project. We would also like to thank Dennis Aldover, Senior Software Engineer at Netgear Inc. for giving us valuable advice at the required time. They created a very open learning environment and guided us when we were stuck during the course of the project. Regular meetings between the students, Professors and the Faculty of Netgear Inc. helped us in successfully completing this project.

Also we would like to thank Dr. Tri Caohuu, Department Graduate advisor of Electrical Engineering Department and Professor Morris Jones for guiding us and providing a systematic approach and the resources necessary for the completion of the project. We would like to thank Dr. Nader Mir for taking time from his busy schedule and mentoring us as the Co-Adviser of this project.

And finally, we would like to acknowledge the work of Gino Paglia, student in the Software Engineering department, who was a member of our team at the Netgear Inc. for his contribution to this project.

**Abstract**

The project is a prototype of a model that enhances a user's internet experience whilst also ensuring the security and privacy of the user. With the introduction of cookies to the internet world, most of the websites in the world track user data and profile each user as per their requirements. The user usually does not know he is being tracked. The motivation of this project is to ensure the privacy of the user and also not be a deterrent to his online experience. The task in this project involves, eliminating the cookies by preventing them reaching the browser. A profile is then created for the user that is stored locally. Most of the incoming advertisements are blocked, so that the user is not burdened with the irrelevant advertisements. An advertising platform is created so that the user can be fed relevant ads based on his profile. In the prototype, the advertisements have been removed, and a few of them have been replaced by a Netgear Logo to show the advertising platform.

Contents

[INTRODUCTION 9](#_Toc384946338)

[Background 10](#_Toc384946339)

[Privacy and Advertising platform 11](#_Toc384946340)

[HYPOTHESIS 12](#_Toc384946341)

[STATE OF THE ART 13](#_Toc384946342)

[Cookies 13](#_Toc384946343)

[Online Advertising 19](#_Toc384946344)

[Ad serving and Ad Blocking 25](#_Toc384946345)

[IMPLEMENTATION 27](#_Toc384946346)

[Implementation Tools and Requirements 27](#_Toc384946347)

[Architecture 31](#_Toc384946348)

[Browser 31](#_Toc384946349)

[Router Firewall 32](#_Toc384946350)

[Analyzer/Proxy Server 33](#_Toc384946351)

[Cookie Component/Tracking Cookie 33](#_Toc384946352)

[Advertising Blocking/Ad Request Component 34](#_Toc384946353)

[URL Component 35](#_Toc384946354)

[PROJECT DESIGN 37](#_Toc384946355)

[Proxy Server Component 40](#_Toc384946356)

[Tracking Cookie Component 42](#_Toc384946357)

[Ad Server Component 51](#_Toc384946358)

[Category Component 53](#_Toc384946359)

[TEST CASES 57](#_Toc384946360)

[Test Case for Cookie Component 57](#_Toc384946361)

[Test Case for Ad-Server Component 58](#_Toc384946362)

[Test Case for URL Component 59](#_Toc384946363)

[DEPLOYMENT 60](#_Toc384946364)

[Browser Settings 60](#_Toc384946365)

[Deployment on Apache Tomcat Server - Java Application 65](#_Toc384946366)

[Deployment using Boost Libraries - C++ application 65](#_Toc384946367)

[SUMMARY 74](#_Toc384946368)

[RESULTS AND CONCLUSIONS 75](#_Toc384946369)

[RECOMMENDATIONS FOR FUTURE RESEARCH 78](#_Toc384946370)

[DIVISION OF WORK 80](#_Toc384946371)

[REFERENCES 80](#_Toc384946372)

**List Of Figures**

[Figure 1 - Hyper Text Transfer Protocol Request[12] 14](#_Toc384946373)

[Figure 2 - Hyper Text Transfer Protocol Response[12] 14](#_Toc384946374)

[Figure 3 HTTP Request and Response headers from the chrome browser 16](#_Toc384946375)

[Figure 4 - Percentage using Cookies and Session cookies out of the top 10 million websites[16] 18](#_Toc384946376)

[Figure 5 - Percentage using Cookies and Persistent cookies out of the top 10 million websites[17] 18](#_Toc384946377)

[Figure 6 - Based on collecting information from the user, this is an example on how Online advertisers look at them 22](#_Toc384946378)

[Figure 7 - Online Advertising - Different types of advertisements possible in a web page[27] 23](#_Toc384946379)

[Figure 8 - State of the art tools and their functionality 26](#_Toc384946380)

[Figure 9 - Functional Requirements set forth by Netgear 28](#_Toc384946381)

[Figure 10 - Non Functional requirements set forth by Netgear 29](#_Toc384946382)

[Figure 11 - Tools used in the project 30](#_Toc384946383)

[Figure 12 - Netgear Privacy and advertising platform Overview 31](#_Toc384946384)

[Figure 13 - Proxy setting changes for the Browsers 32](#_Toc384946385)

[Figure 14 - Proxy Server Overview 33](#_Toc384946386)

[Figure 15 - Cookie Component Diagram 34](#_Toc384946387)

[Figure 16 - Advertising Server Component 35](#_Toc384946388)

[Figure 17 - Advertisement Request Direction 35](#_Toc384946389)

[Figure 18 - URL category component 36](#_Toc384946390)

[Figure 19 - Netgear Privacy and Advertising Platform class diagram 38](#_Toc384946391)

[Figure 20 - Advertisements blocked on nba.com and replaced by Netgear Logo 39](#_Toc384946392)

[Figure 21 - Advertisements on timesofindia.com replaced by Netgear Logo 39](#_Toc384946393)

[Figure 22 - Sequence Diagram for the Netgear Privacy and Advertising platform project 41](#_Toc384946394)

[Figure 23 - Code snippet showing Thread Pool Creation 42](#_Toc384946395)

[Figure 24 - Chrome Developer Tools showing Cookie Name, Values and Expiration Dates 44](#_Toc384946396)

[Figure 25 - Adserver packet showing Tracking Cookie in the Response 45](#_Toc384946397)

[Figure 26 - Sequence Diagram showing Tracking Cookie Component 46](#_Toc384946398)

[Figure 27 - Set-Cookie shows expiry date and hence tracking Cookie 47](#_Toc384946399)

[Figure 28 - Tracking Cookie Removed 48](#_Toc384946400)

[Figure 29 - Session Cookie 49](#_Toc384946401)

[Figure 30 - Session Cookie Highlighted 50](#_Toc384946402)

[Figure 31 - Sequence Diagram for the Advertising Component 51](#_Toc384946403)

[Figure 32 - CNN ad server is blocked while Turner ad server is not blocked 52](#_Toc384946404)

[Figure 33 - Category Component Sequence Diagram 53](#_Toc384946405)

[Figure 34 - USER\_PROFILE\_CATLIST.TXT example 55](#_Toc384946406)

[Figure 35 - catmap.txt 56](#_Toc384946407)

[Figure 36 - catcount.txt 56](#_Toc384946408)

[Figure 37 - Settings tab on Google Chrome 60](#_Toc384946409)

[Figure 38 - Setting Drop down Menu 61](#_Toc384946410)

[Figure 39 - Change proxy settings option 61](#_Toc384946411)

[Figure 40 - select your proxy server 62](#_Toc384946412)

[Figure 41 - HTTP packets to go through your proxy server running on port 80 63](#_Toc384946413)

[Figure 42 - Settings in Internet Explorer 63](#_Toc384946414)

[Figure 43 - Choose Internet options 64](#_Toc384946415)

[Figure 44 - Connections tab 64](#_Toc384946416)

[Figure 45 - Setting Environment Variables 67](#_Toc384946417)

[Figure 46 - Bootstrap command with mginw toolset 68](#_Toc384946418)

[Figure 47 - Run bjam with the required options only 69](#_Toc384946419)

[Figure 48 - Properties in Visual Studio C++ 70](#_Toc384946420)

[Figure 49 - Additional Include directories option 71](#_Toc384946421)

[Figure 50 - Setting no Precompiled Headers 72](#_Toc384946422)

[Figure 51 - Build and Run options under the Debug tab 73](#_Toc384946423)

[Figure 52 - Advertisements blocked on nba.com 75](#_Toc384946424)

[Figure 53 - Cookies before running the application 76](#_Toc384946425)

[Figure 54 - Cookies blocked after running through the application 77](#_Toc384946426)

[Figure 55 - Classification types[37] 79](#_Toc384946427)

[Figure 56 - Flat and Hierarchical classification[37] 79](#_Toc384946428)

# INTRODUCTION

In this technology era, the internet has by far become the most important resource to mankind. Over billion users worldwide use the internet. It has given a whole new dimension for one to communicate with each other. With the evolution of software and different products invented and created, the internet has become the largest advertising field for ones products. From the small scale industries across the world in Asia to the Heavy weight software companies in America, all of them want to market themselves and have to market themselves if they are to expand their industry. This gave rise to a whole new field on Online Advertising, where the advertisers target a particular type of audience to sell a particular product.

These days, with internet playing a key role in ones daily life, most of the users sensitive information is stored online. This is kept online assuming that, no one is monitoring him and his information is safe. The user wants his privacy in the internet. In the current internet world, getting any sort of information on a particular user is very helpful in advertising and marketing. For example, if an advertiser knows that the person is a male in his mid twenties, he will display him advertisements on cars, home theatres, etc. If the person is found to be female, then the advertiser will show advertisements of jewelry, perfumes, etc. Basically, any information available about a user is useful to the advertisers. If the advertiser is given a complete profile of the user and his traits and habits and preferences, it is a gold mine to them.

How are these profiles created? The users online activity is constantly monitored without his knowledge. The websites he visits, the time he spends on each of them is monitored. His preferences and other information about him are stored on advertising servers. The tracking cookies make this monitoring and profiling of the user possible[1]. Tracking cookies are set by the online servers and stored on the browser. These remain on the browser from a single day to several years unless removed by the user. These cookies monitor the traffic of the user and send back information to the server. The server in turn decodes the data and creates a demographic profile of the user[2].

There are several tools being created these days with an interest to help the user and assist in this matter. Microsoft's Internet Explorer allows the user to adjust his privacy setting to suit the requirement[5]. For various browser's there a tools such as Adblock plus created by Wladimir Palant, which help in blocking certain information and advertisements on the browser. But even though the information is blocked, the cookies are still set by the third party servers, thus monitoring the user. The Google Chrome browser by itself monitors the user adn sends the analytics data to the Google Servers[3].

The user has every right to feel that his privacy is at stake. The reason is, even though the websites are considered safe, his information is given to third party ad servers. These servers are unknown and their job is to have a demographic profile of the user. It is these third party cookies which is the major reason of concern.

## Background

The fundamental idea in this project is to remove advertisement, prevent the monitoring of the users and create a profile for the user. The profiling of the user is done with the help of cookies. The user may or may not be aware of being monitored. The user is not only being monitored by the high profile websites he may be aware of - such as Amazon, Facebook, etc. But the issue is that the user is being tracked by unreliable and unknown third party advertisement servers as well. At present, there are tools to remove advertisements on the browsers and manipulate required stuff of the browsers. But, much more research is still needed to enhance the privacy of the user.

## Privacy and Advertising platform

In this project, we ensure the user privacy by eliminating the tracking cookies in the application. To provide a good user experience with ads, we create a user profile that is stored on the router itself. In this prototype, we have created a text file with a set of categories, that keep count of each category. Based on the category count we could provide a relevant advertisements. An option for further enhancement on this project is to data mine the websites used by the user and to create a demographic profile for the user. Even though the user profile is being track all the day will be kept on the router or a known secure server.

# HYPOTHESIS

In this project, a proxy server is created. This acts as a middle man between the remote web servers and the users surfing through them on the browser. The ongoing web traffic can be monitored using this proxy server. This helps us view the Http packets that are part of the online traffic. We can read the headers for the HTTP Requests and also read the HTTP Responses. This gives an insight and all the required information that needs to be read and manipulated. In the HTTP Requests, the important fields for interest are the URL names which is required to create the profile of the user. Moreover, the URL names are matched with a given blacklist of advertisement servers. If the name matches with those in the black list, then the request is blocked thus preventing the advertisement. Instead of these blacklisted request, occasionally a Netgear Logo is placed in its stead. In the HTTP Responses, the important fields for us are the Cookie headers. The entire Cookie contents are parsed looking for the tracking cookies. If a tracking cookies is found, it is removed from the HTTP response itself, thus preventing it from reaching the browser and being set on it.

# STATE OF THE ART

The router is a device that is used to tie multiple businesses together. It is able to underst

and the different networks present and find the best route to each one of them. The router analyzes the data in the traffic flow, packages the packet in the format for the other network and sends the packet to that network[6]. The router is a layer 3 device in the OSI model. It works in the Network layer and looks at the IP packets[7]. Router use some routing protocol for its functioning[8].

There are several companies that manufacture the Router hardware devices. To name a few - Cisco Systems, Juniper Networks, Netgear, Belkin, etc[9]. Any additional feature the router can perform, other than its routing functionality is a huge plus point in this industry to market their router. In the Netgear routers, this privacy and advertising project platform can be deployed and run along with the other routing functionalities without affecting the performance of the router. The following sections describe cookies and online advertising and their current state of art in the industry.

## Cookies

The Hyper Text Transfer protocol is a stateless protocol[10]. This means that it does not retain any information from the previous requests or responses. To make the HTTP protocol seem like a stateful protocol, Cookies were created. The origin server can maintain the state information of the user agent and the user agent can read that state information and in return send the state information back to the origin server[11]. In the absence of Cookies, each time a new browser is opened or as and when we browse through the web pages, each of it will be considered as a new connection with a new HTTP Request and a new HTTP Response. Though this makes it easy to create new browsers and user agents, it makes it very complex to design and create a practical web application. Cookies make it possible to go through the user page and makes browsing user friendly and convenient. Else, the user would have to give the same information over and over as he browses through the web site. An example HTTP Request and Response -



Figure 1 - Hyper Text Transfer Protocol Request[12]



Figure 2 - Hyper Text Transfer Protocol Response[12]

When the user accesses a website, a HTTP connection is setup to the web server. A HTTP Request is sent to that server. The server accepts the connection and along with connection sends a HTTP Response. In this response, it sets the Set Cookie header field. This is a piece of text or information, created, understood and usually encrypted by the server. Once this response reaches the user agent or the browser, the Cookie is stored on the browser. This Cookie remains on the browser, collecting information about the user from the user agent or browser. When the user accesses the same web server after a while, this Cookie is attached in the HTTP Request and sent to web server. The server collects the cookie, decrypts it and reads the information about the user from it and sends it back with the Response. The Cookie is exchanged this way, thus relaying information about the user to the web server. The web server while returning the Cookie can also add additional Cookies or remove Cookies as and when required[12].

The following figure shows the HTTP request and Response headers taken from the Chrome browser. Open the developer tools in the chrome browser. Either open it in setting or hit ctrl+shift+i keys. Open a web page. Hit the Network tab to see all the packets that have been transmitted and received for that web page. Click on a packet to see the HTTP Request and HTTP Response headers.



Figure 3 HTTP Request and Response headers from the chrome browser

Cookies were created to make web browsing easier and faster. Currently, online shopping and online industry have seen a huge growth and it looks likes it is likely to grow further beyond our imaginations. Cookies play a very important role in online shopping. They gave rise to features like the shopping cart and helped the web servers remember the user registration details and items stored in the shopping cart. The cookies remain on the browser monitoring the users interest and is exchanged between the web server and browser as explained earlier. It helps to build a demographic profile of the user. This enables the website to provide better and relevant web service to the user[13]. A good or apt example to illustrate the use of Cookies is to learn a few things about online shopping giants, Amazon. The privacy notice of Amazon, clearly states to the user about the Cookies Amazon use and its usefulness. It is clearly stated in their privacy notice that, unique identifiers or Cookies are transmitted by the server to identify the type of device and more information. The Cookies have enabled them to provide online services such as 1 Click purchasing. They also create a personalized recommendation page for each user to enhance his web experience. Moreover, shopping carts seem to have the memory to store the items as it may appear, due to the Cookies being used to track items in them.

Cookies can be classified broadly into Session Cookies and Persistent Cookies. The websites on their own have no memory. The Accept a Request, process it and send the Response before closing the connection. This way each request is processed. Through the website, the user will be treated as a new user on each page. This is where, Session Cookies come into play. They are stored on the browser and exchanged to and fro so that the website knows the user. Due to this, it does not see the user as a new user and allows the user to freely browse through the website. An example to see the advantage of Session Cookies - Open the Amazon website. Once you open amazon.com add a few items into the cart. Close the browser and open amazon.com in a new tab. You will still see the items in the cart. Now open the setting panel of your chrome browser either from the menu or by hitting ctrl+shift+del. Tick the clear Cookies option. Now when you open the amazon.com website again, there will be no items in the cart. It will be as though you are accessing amazon.com for the first time. The Session Cookies are normally deleted after the session or when the connection is closed. They are not saved on the browser once the session is done[15].

Persistent Cookies, unlike Session Cookies, remain on the browser after the connection has been closed. The duration these Cookies remain on the browser can be from a short duration of one day to more than five years. These Cookies are used to remember the user preferences when the same website is viewed again in the future. Te Persistent Cookies are also called Tracking Cookies. They can also be used to accomplish functionalities like identifying the user, monitor the surfing behavior of the user. Using this information, it can deduce the age of the user. It keeps track of time spent on each web page by the user and also analyze the performance of these web pages. Persistent Cookies or Tracking Cookies remain on the browser for a prolonged time. W3 techs - Web Technology Surveys, did a survey on Cookies, Tracking Cookies and Session Cookies being used. The results said that 46.5% of all websites used Cookies. Among all the websites 36.2% used Session Cookies. 20.2% of websites used Tracking or Persistent Cookies[16][17].

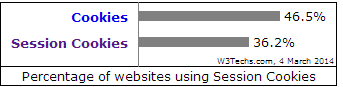


Figure 4 - Percentage using Cookies and Session cookies out of the top 10 million websites[16]

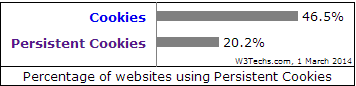


Figure 5 - Percentage using Cookies and Persistent cookies out of the top 10 million websites[17]

Since Cookies can be easily captured and viewed, the main issue is the Security and Privacy concerned with Cookies. Our main concern is not with web servers such as Amazon or Facebook or Google, etc. It is with other unknown and less popular websites. These websites normally outsource their advertising to third party ad servers. These servers are an unknown entity and it is these servers which look to keep a Tracking Cookie on your browser to profile you. Vincent Toubiana and Vincent Verdot of Alcatel-Lucent Bell labs wanted to test the security of the Cookies and analyzed the session cookies for the Google web search of ten volunteers. Analysis of the Cookies alone let them identify 80% of their search history of the ten volunteers[18]. This is just one example to test the security measures for Cookies. There is research also being done to enhance the security of the Cookies. One such example is by Heng Wu, Weiting Chen and Zhongjie Ren in providing a Mac Address Encrypted Key ring for the Cookies[19].

## Online Advertising

The New York Times magazine published an article called "Who do Online Advertisers think you are". The author, Jeffrey Rosen elaborates on Online Advertising and how it being used in the present market. He created a simple test case or experiment to explain the working. The Cookies on two browsers were cleared. Two user profiles were to be created. Over a period of two days, on the Safari browser, a democratic profile was intended to be created. He viewed websites related to Obama's campaign, went through travel websites, looked into buying a car, checked flight ticket prices to Los Angeles, went through Volvos and Birkenstocks. On the Firefox web browser, a Republican profile was favored. The websites viewed were related to Mitt Romney's election campaign, Cadillac sales websites, planning a trip to Hawaii and looking for diamond rings. This kind of browsing went on for two days. There were no specific changes of interest to be noted. On the third day, Jeffrey Rosen noticed that, on a particular website, in the advertisement space was an ad showing for Catholic University, where as on the same website and on the same space the Firefox browser showed Romney's election campaign advertisement[20]. These advertisements were different because of Real time bidding of advertisements. Scott Spencer, Director of Product Management, Ad Exchange, Google gives an apt definition for Real Time Bidding. In summary, he states that real time bidding is the method or the process in which only selected or relevant ads will be provided to consumers[21].

Online advertising has become a reality with the aid of Cookies. Oracle, recently took over Bluekai, a big data for marketing firm. Bluekai aid their partners in serving the right advetising by accumulating data for them. As an example, Jasmine is looking to travel from New York to Hawaii. Upon visiting websites partnering with Bluekai, on her browser, an anonymous Cookie is stored. This is common terms is a tracking cookie monitoring her activity and hence knows she is looking for a trip to Hawaii. Bluekai informs its partner about Jasmine and she is displayed ads relevant to travel and Hawaii[22]. Similar to this, The Facebook Exchange was launched by Facebook to monitor user traffic. Using this data, relevant advertisements were provided. Assume a user looks at a Ford SUV page, Facebook Exchange takes account of this user traffic and displays Ford SUV related or similar ads on the Facebook page. The accuracy of targeting the right people is important. Big premiums are being paid by advertisers for highly accurate targeting[23]. Cookies such as \_\_utma, \_\_utmb, ..., \_\_utmz are Google analytic cookies to monitor the sessions and the user[24].There are also Persistent Cookies stored with the chrome browser to monitor the user[25].

Monitoring users, creating a profile and keeping track of demographic information is what online advertising agencies do. After conducting his simple experiment, Jeffrey Rosen visited Bluekai to learn how the advertisers viewed his profile. The Firefox browser, that is the Republican profile was viewed as someone earning $60000-$74999 from Portland. This profile likes fancy cars, likes the celebrities and Television. The Safari browser, that is the Democratic profile was viewed as a big shot living in Los Angeles or one of the fancier cities like Long Beach or Santa Ana running a company having over 5000 employees under him. This profiles interest would mainly be in advertising and Marketing. Even though the information gathered isn't completely accurate, this is a good example of how Online Advertising has advanced[20].

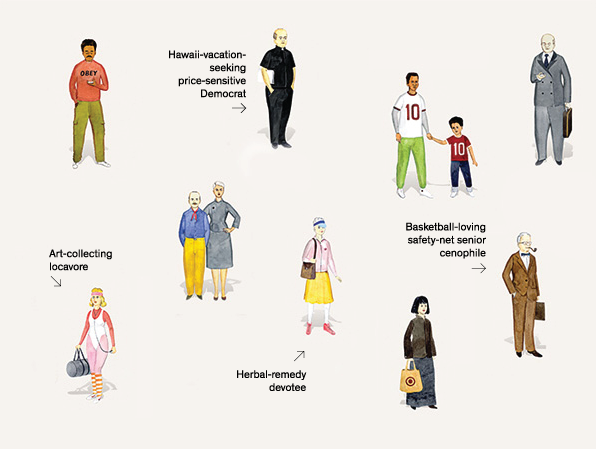


Figure 6 - Based on collecting information from the user, this is an example on how Online advertisers look at them

Online tracking, profiling and advertising depends on recording each and every web activity of the user. The intention is businesses and commercial purposes. Upon accessing a website, a Cookie is stored on the web browser. The activities of the user are monitored using this Cookie. This Cookie keeps track of what pages she visited, time spent on each page, the advertisements viewed, articles read, purchases made, names or items that were searched, the web browser used, the operating system being used and the geographical location. This accumulated information helps behavioral advertising. Behavioral advertising is a system where the acquired information is used to feed the user with relevant advertisements of his or her interest. Based on the assumed interests of the user, it has a selection of categories and relevant ads based on user interests and demographic profile are provided. Web sites usually hire third party ad servers to display the advertisement based on the information being shared with them and the website[26]. Advertisements are displayed by the advertiser based on the web page. The Following figure shows some type of advertisements -



Figure 7 - Online Advertising - Different types of advertisements possible in a web page[27]

Online Advertising and Behavioral advertising is a currently trending field where lots of research and development is taking place. Sophisticated systems and algorithms, methods and tools to facilitate, optimize and manage the campaigns and the advertising market are being developed. Knowledge in machine learning, stochastic optimal control, information retrieval, data mining, natural language processing, and econometrics are needed to develop a solution that is mathematically sound and practical. Analyzing the statistics on spending in this field shows the importance of it. In 2012, around $100 billion was invested into this field and is projected to be $163 billion by 2016[28]. Taking all this into account, it is only logical to assume more importance given to build highly accurate demographic profiles.

Amidst all this development, the issue here is that users are being tracked with or without their will. Can one opt out from being tracked if needed is something to be considered. The United States Federal Trade Commission looked into this issue and concluded that the user can opt out if he does not want to be tracked online. Companies currently customizing advertisements on your browser can be scanned and found by accessing the website - www.aboutadsinfo.com. In this website, the user is given an option to opt out from the relevant advertisements from these sources. This does not mean the user is not being tracked. He is still tracked with the help of Cookies, but he is not given the relevant advertisements for his profile because he has opted out of it. Browser ad-ons can set opt out cookies, block advertising cookies, block request to tracking domains or notify users if they are being tracked. Lorrie Faith Cranor of Carnegie Mellon University conducted a study and tested nine tools to prevent bring tracked - three in websites such as the one provided by the US federal state, two built in tools as part of the browser settings and four tools Ghostery, TACO, Adblock plus and internet explorer tracking protection. None of these tools proved successful in allowing the user to control online tracking and behavioral advertising to suit their preferences[29].

The infringement of the privacy of the user is the main issue here. Any online activity and any task done by the user online is being tracked. Whether the user is comfortable with this or not, he has to put up with it without a choice. Another dilemma is that, if he is not being tracked and a demographic profile is not built for him, he may not enjoy a good online experience. This is a necessity, but must be done without causing a concern to the user.

## Ad serving and Ad Blocking

There are several application or simple plug in such as "No script" to Norton Antivirus that blocks all advertisements and the control is not with the user. In the internet advertising business Google Ad services, Yahoo Ads and Amazon ad services are clearly at the forefront. Google Ad services is so dominant that is earns 41% of all the digital revenue and 51% of all the mobile advertising revenue[30].

The following figure summarizes the different fields and options available to the user. The Netgear Privacy and advertising platform project accomplished Persistent cookie management, blocking advertisements and serving advertisements.

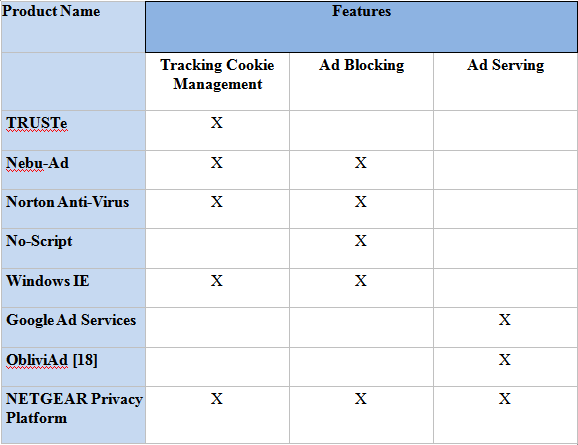


Figure 8 - State of the art tools and their functionality

# IMPLEMENTATION

This Section describes the implementation of the project and the Tools used. The architecture and design of the project.

## Implementation Tools and Requirements

The project will be implemented in C++ programming language. The project was initially implemented using Java and the apache Tomcat server and was run successfully. Initially, the entire application was supposed to be run on Java and ported on the router using the Netgear Genie platform. The Netgear Genie platform consists of a Virtual Machine and a Java compiler on the router to run applications on the router. Due to changes in the management structure at Netgear, the Genie platform and portal was shutdown. The application was then to be ported on to C++. At the center of the application is a Proxy server that enables us capture the HTTP packets. It was implemented using servlets in Java and Boost libraries in C++. Netgear set forth certain functional requirements which were to be implemented by us. These are shown in figure 9. Figure 10 shows the Non functional requirements set by them. Our emphasis in this project is to make sure the functional requirements are met and among the non functional requirements to make sure our code is scalable, that is, less than 2MB.

The implementation was successfully done on the Microsoft Windows 7 and Ubuntu platforms. In Ubuntu, the Eclipse IDE was used. In Windows 7, the code was implemented with Microsoft Visual Studio 10. The application was successfully implemented using the Java and C++ programming languages.

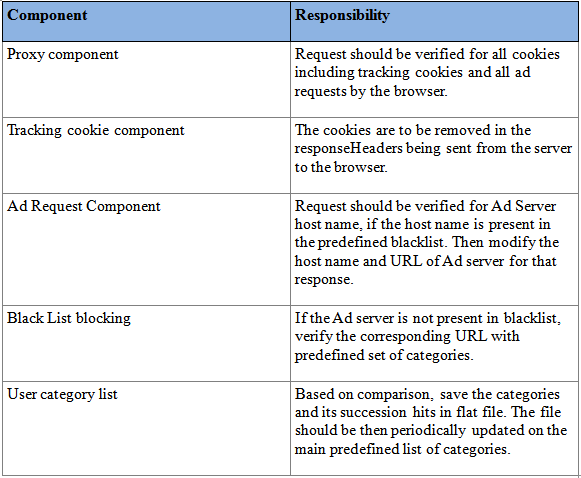


Figure 9 - Functional Requirements set forth by Netgear

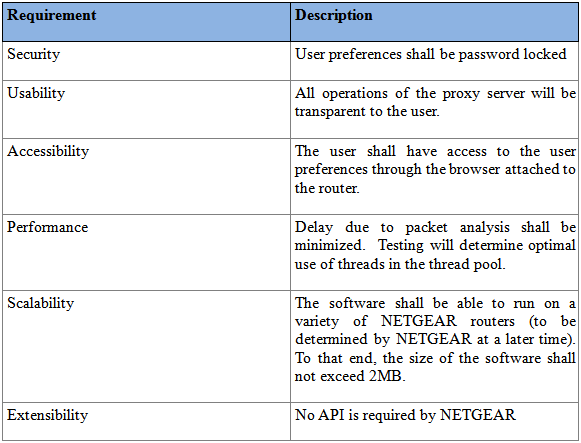


Figure 10 - Non Functional requirements set forth by Netgear

The software tools used over the course of this project are shown in the following figure. The Majority of them have an Open License and are free. The paid tools for this project are paid to as part of the curriculum at San Jose State University. Subversion using SVN were created, hosted and managed by Netgear.

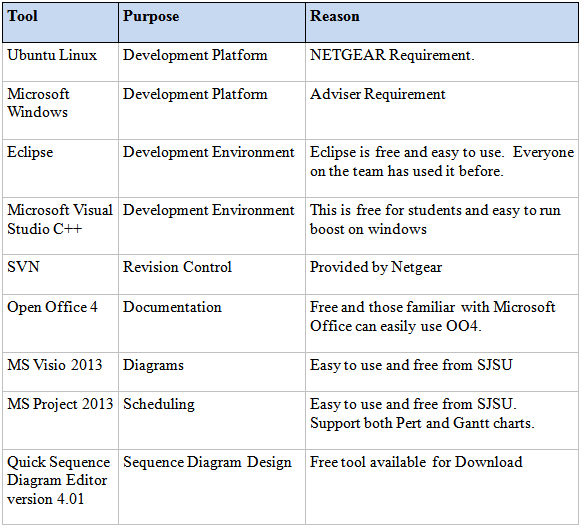


Figure 11 - Tools used in the project

## Architecture

The top view architecture of the Netgear privacy and advertizing platform application is shown in Figure. Following that each component of the application is explained.

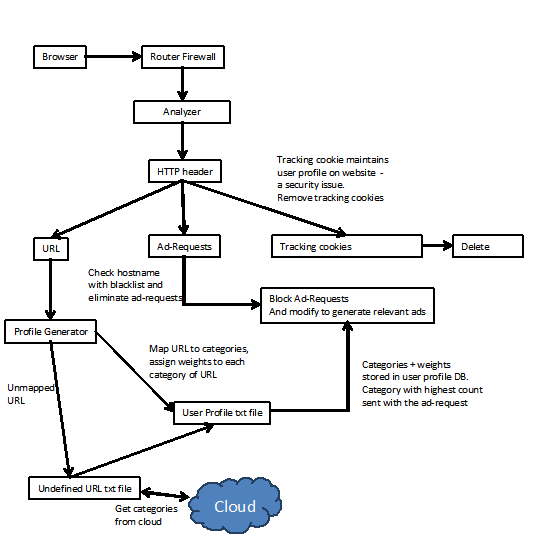


Figure 12 - Netgear Privacy and advertising platform Overview

### Browser

The browser is the user agent in our application. It is here he user accesses the websites. This creates the required Http connections. In the prototype, we need to change the setting on the browser to make connections through our proxy and on port 8080. The browsers used in our prototype were the Microsoft Internet Explorer 11 and Google Chrome version 33.0.1750.154

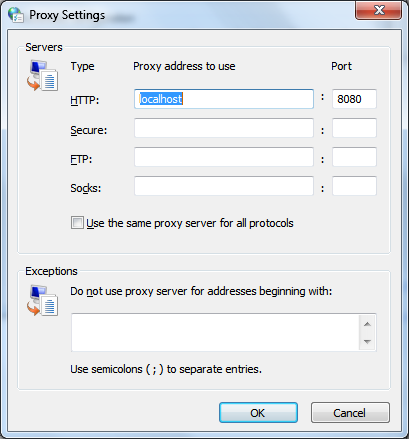


Figure 13 - Proxy setting changes for the Browsers

### Router Firewall

The Router Firewall, which is not a part of our prototype model, is present within the Router. This is a part of the Router Software which is supposed to redirect the HTTP packets within the router to pass through a port number mentioned in our application. In this case, it is 8080.

### Analyzer/Proxy Server

In the prototype the analyzer or proxy server is use to act as a middle man between the web server and the browser. Its job is to intercept all the HTTP packets that are part of the ongoing traffic, and give us access to read this information. Further on from here, this information will be used by the other components. In Java, this component is implemented using servlets. In the C++ implementation, Boost libraries are used to implement this component. Thread pools are created by the proxy server to reduce connection time for multiple requests. This is done instead of creating connection threads at the time of connection. Twenty threads used resulted in low memory usage and yielding the best performance.

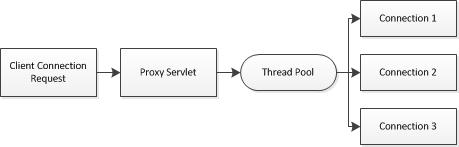


Figure 14 - Proxy Server Overview

### Cookie Component/Tracking Cookie

As already discussed several times in this report, Tracking Cookies are set on the browser and remain there monitoring the user and accumulating date to create a demographic profile for the user. The Proxy server also gives us access to the HTTP Responses that are being sent from the server to the user. The Cookies found in these responses are parsed and should they contain an Expiration date, which the session cookies do not have, the Cookie is eliminated. Thus, this component ensures that the tracking cookies are removed prior to reaching the browser itself. The "Set-Cookie" header entry in the HTTP packet has the Cookie and a Cookie with the expiration date is the Tracking Cookie. Session Cookies have a separate header entry for the expiration date.

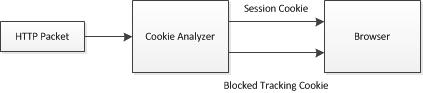


Figure 15 - Cookie Component Diagram

### Advertising Blocking/Ad Request Component

A text file containing a list of black listed advertisement servers is created. From the HTTP Requests captured by the proxy server, the URL of the request made is scanned. If the URL name of the website matches the one of those in the blacklist, the HTTP Request packed is dropped. This prevents irrelevant ads from being displayed on the web page. In the real model, Netgear will create an Ad service component and provide advertisements instead of the Advertisement being blocked. In this prototype, instead of the blocked advertisements, we serve the Netgear Logo. In the C++, implementation, this component is not very stable and the Netgear Logo is only occasionally displayed. In Java, the Netgear Logo was implemented with greater frequency.

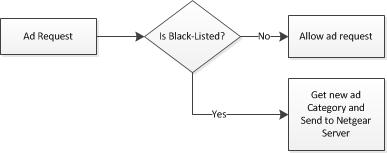


Figure 16 - Advertising Server Component



Figure 17 - Advertisement Request Direction

### URL Component

In the prototype, there are two text files of importance - catmap.txt and catcount.txt. The catmap.txt contains a predetermined list of websites along with their categories. As an example we have taken about 10 websites along with their respective categories. The catcount.txt file contains the category name and the count of the category. Based on the category of the website accessed, its count is increased. In the actual model, the intention is to keep a record of the websites accessed and get the categories for them from a third party provider. This would replace the catmap.txt. The count is then based on the number of hits on these categories. This information will be provided to the Netgear ad servicing component which would serve the relevant advertisement for the category.



Figure 18 - URL category component

# PROJECT DESIGN

The class diagram for the entire project is shown in the figure below. The class diagram shows all the components in the project. The interaction of all components and external servers can be seen in this figure. The BoostProxy.cpp, ProxyServer.cpp and ProxyConn.cpp together enable the working of the Proxy Server or the Analyzer. The TrackingCookieAnalysis.cpp class, is used to find and eliminate the tracking cookies. The URLConnect.cpp is used to establish connection to the website. The CatMap.cpp class reads the contents of Catmap.txt and puts the URL name and category name in two arrays, hostname[] and category[]. The CatCount class, reads the category and count that category is at and stores it in the count[] and category[] arrays. The AdRequestComponent.cpp interacts with the Blacklist.cpp class and compares the URL with the blacklisted servers.

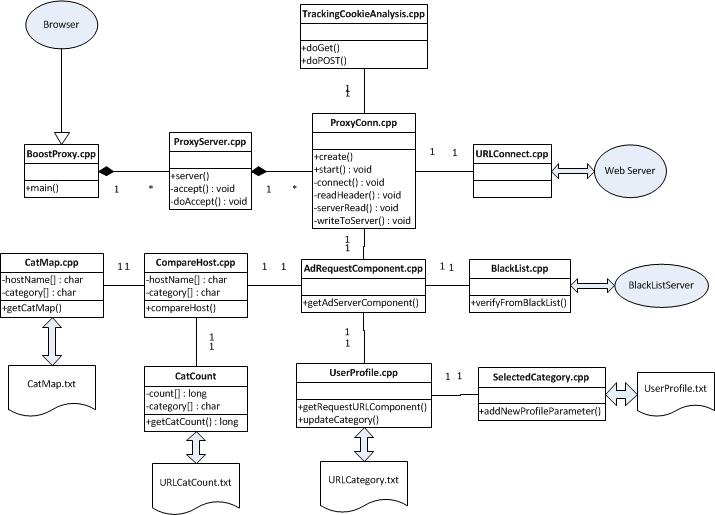


Figure 19 - Netgear Privacy and Advertising Platform class diagram

The project has no client design. The websites can be accessed by using any web browser. This project was mainly tested on Google Chrome and Internet Explorer as mentioned before. When a website is accessed and run through the proxy of this project, some of the blocked advertisements are replaced by the Netgear Logo as shown. The advertisement serving module of this project was to be implemented by Netgear. Since it hasn't yet been implemented, it was decided that a Netgear Logo be used instead. The output is shown in the following figures.

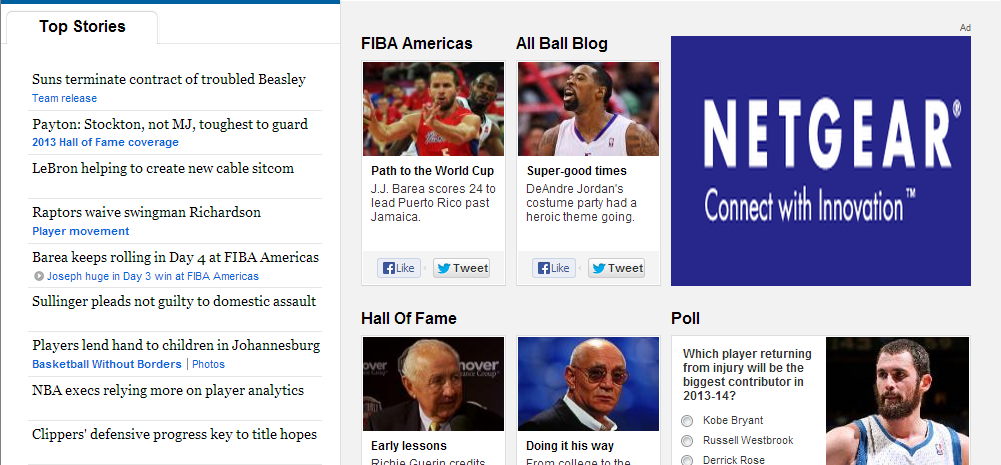


Figure 20 - Advertisements blocked on nba.com and replaced by Netgear Logo



Figure 21 - Advertisements on timesofindia.com replaced by Netgear Logo

## Proxy Server Component

When the user connects to a website, the connection requests, HTTP Requests and HTTP Responses are captured by the Proxy Server, since the Proxy Server is the middle man between the user and the web server. The thread pool creates a new connection and this connection is analyzed. The Request from the Browser is sent to the Ad server component. Here the URL is scanned and compared with the Blacklist. If not, this connection is not blocked and forwarded to the web server. In the HTTP Response, at the proxy server, the tracking Cookie component is called. Here, if the Cookie is a tracking Cookie, it is removed and the Proxy Server forwards the Response to the Web Browser. The following figure shows the sequence diagram of the proxy server component. This figure is for the implementation in Java. It shows the various components in the figure. The tool used to create Sequence Diagrams is the Quick Sequence Diagram Editor version 4.01.

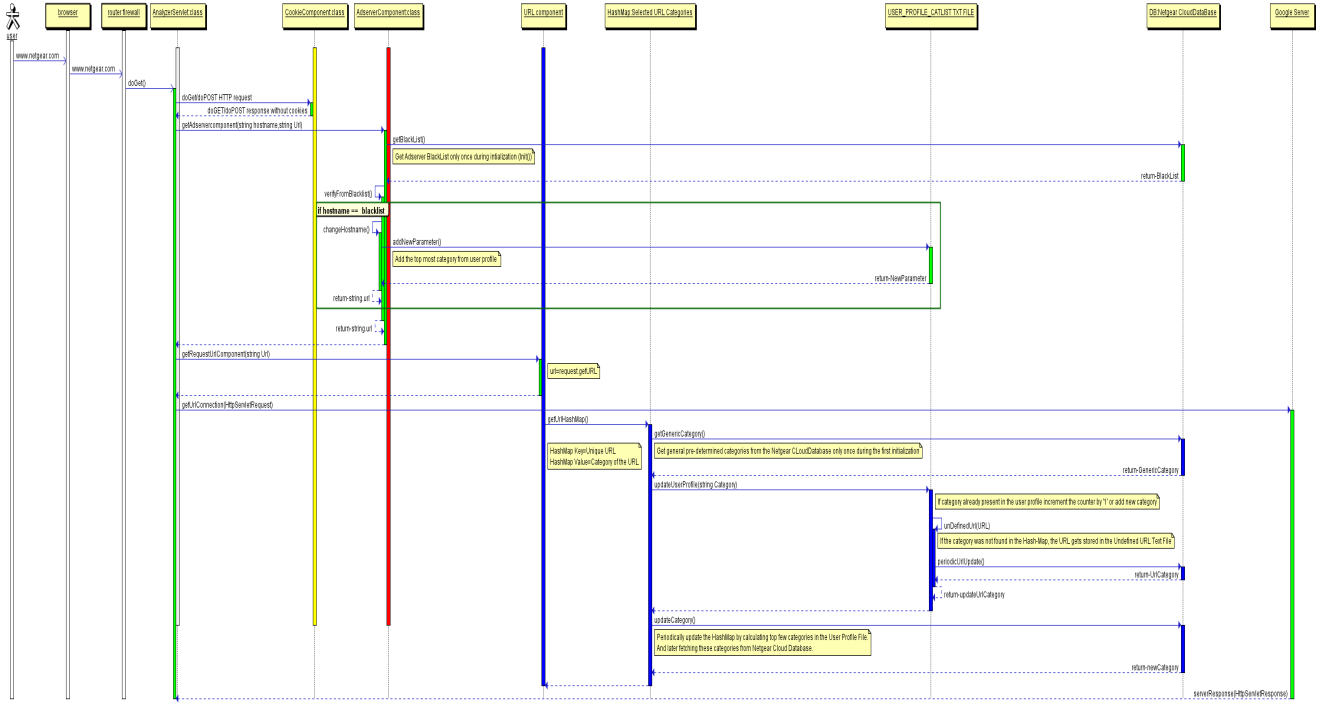


Figure 22 - Sequence Diagram for the Netgear Privacy and Advertising platform project

The following figure shows, the code snippet used to create a pool of threads, that will be created for handling client connections. The Boost::thread\_group method helps in achieving this using thread handling. Each thread is pushed into a deque of Boost::io\_service objects for storage. When there is a request for connection using the client, from the list, the first thread is popped and used for the client connection.

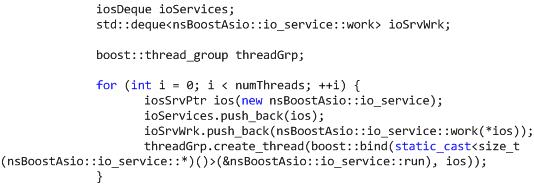


Figure 23 - Code snippet showing Thread Pool Creation

## Tracking Cookie Component

When the user accesses a website a connection is made to the web server. The Web Server reads the HTTP request, services the HTTP Request and creates a HTTP Response. This component as mentioned before, looks at the HTTP Responses received from the web servers. The HTTP Response contains the "SetCookie:" parameter. If the "SetCookie:" value contains a keyword called "expires", this Cookie is a Tracking Cookie. This parameter is cleared. The HTTP Response without the Cookie is passed through to the user, that is the browser. This way, no Tracking Cookie is saved on the browser. The following steps were done to see the Cookies and their values. Open the Google Chrome Browser. Open the Google Chrome Developer Tools from settings or use the Ctrl+Shift+I keys. Open a website, that serves advertisements. Example www.goal.com/en. Click on resources in the developer tools. On the left panel choose Cookies. You can see all the Cookies with their Names, Values and Expiration dates. Now, in the same Developer Tools, select the Network tab and select a packet that is from an ad server. Look at the HTTP Request and HTTP Response. In the Response you should see the Tracking Cookie with the Expires value.

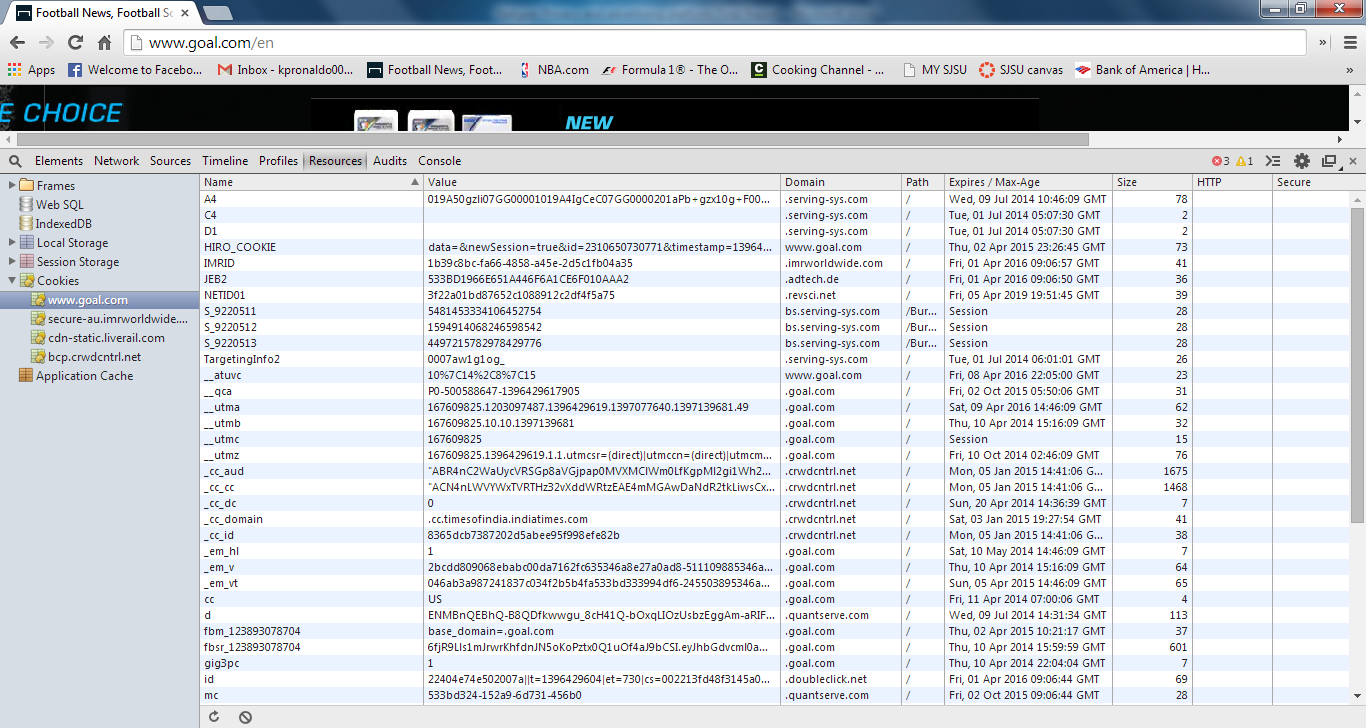


Figure 24 - Chrome Developer Tools showing Cookie Name, Values and Expiration Dates

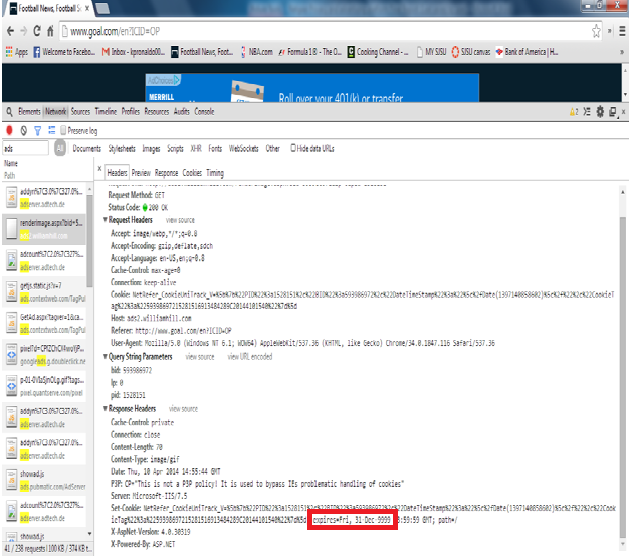


Figure 25 - Adserver packet showing Tracking Cookie in the Response

The Sequence Diagram for the Cookie Blocker used in the Java Implementation is as shown below. This Diagram is self explanatory.

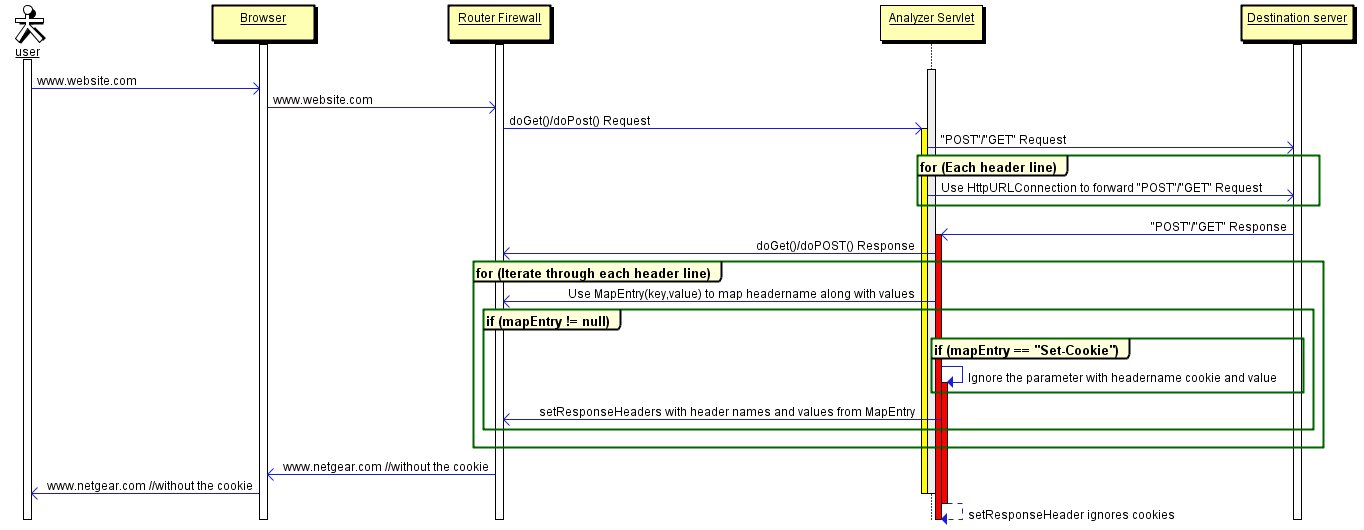


Figure 26 - Sequence Diagram showing Tracking Cookie Component

The HTTP header is analyzed for the Tracking Cookie. In C++, the HTTP header and entries are analyzed using the Boost Client/Server access. The Following Figure shows the Cookie capture using Boost.

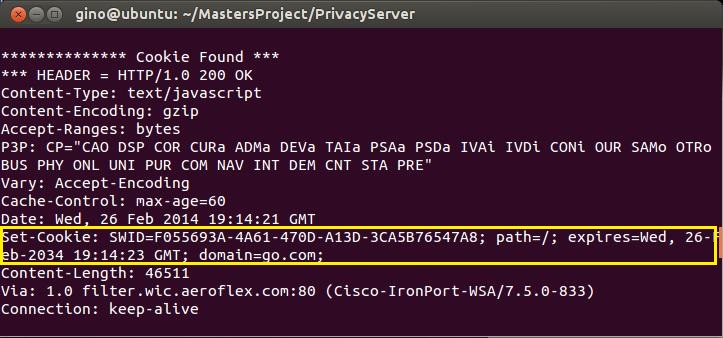


Figure 27 - Set-Cookie shows expiry date and hence tracking Cookie

The Boost Regular expressions library or the Boost regex methods are used to manipulate strings. To the data stream, the HTTP header is reassembled along with its payload and added. The reason the packet isn't completely deleted is that, when the packet is discarded, since the packet has not arrived, another HTTP Request is created requesting that packet and Tracking Cookie. By removing only the SetCookie entry, the web server thinks the Cookie has been sent and the user browser thinks the Cookie has been received.

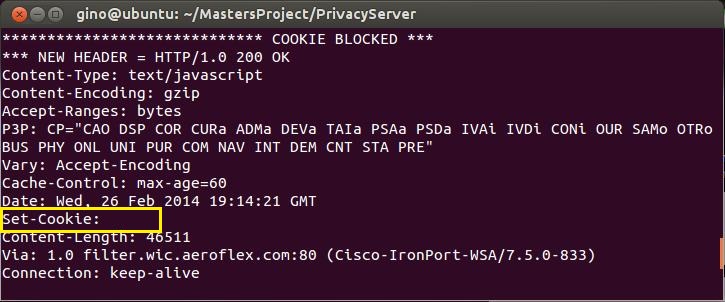


Figure 28 - Tracking Cookie Removed

The Cookie Component leaves session cookies alone. These temporary cookies can be from online activity like online shopping carts and contain a second entry in the http header limiting its use. The Entry “Expires:” sets the limit for the cookies use allowing the browser to remove the cookie at that time. Usually these time limits are for the duration of the user’s session so are of no security risk.

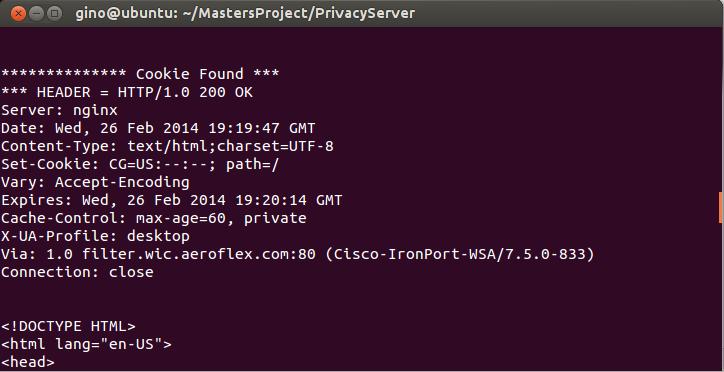


Figure 29 - Session Cookie

In Google Chrome, the developer tools can be used to see the session cookies. The following figure shows the session cookies. The steps are

* Open Google Chrome
* Open Developer Tools
* Access a URL(website)
* Click on the Resources tab.
* Select Cookies on the Left Panel.
* See a cookie with expiration value as Session.

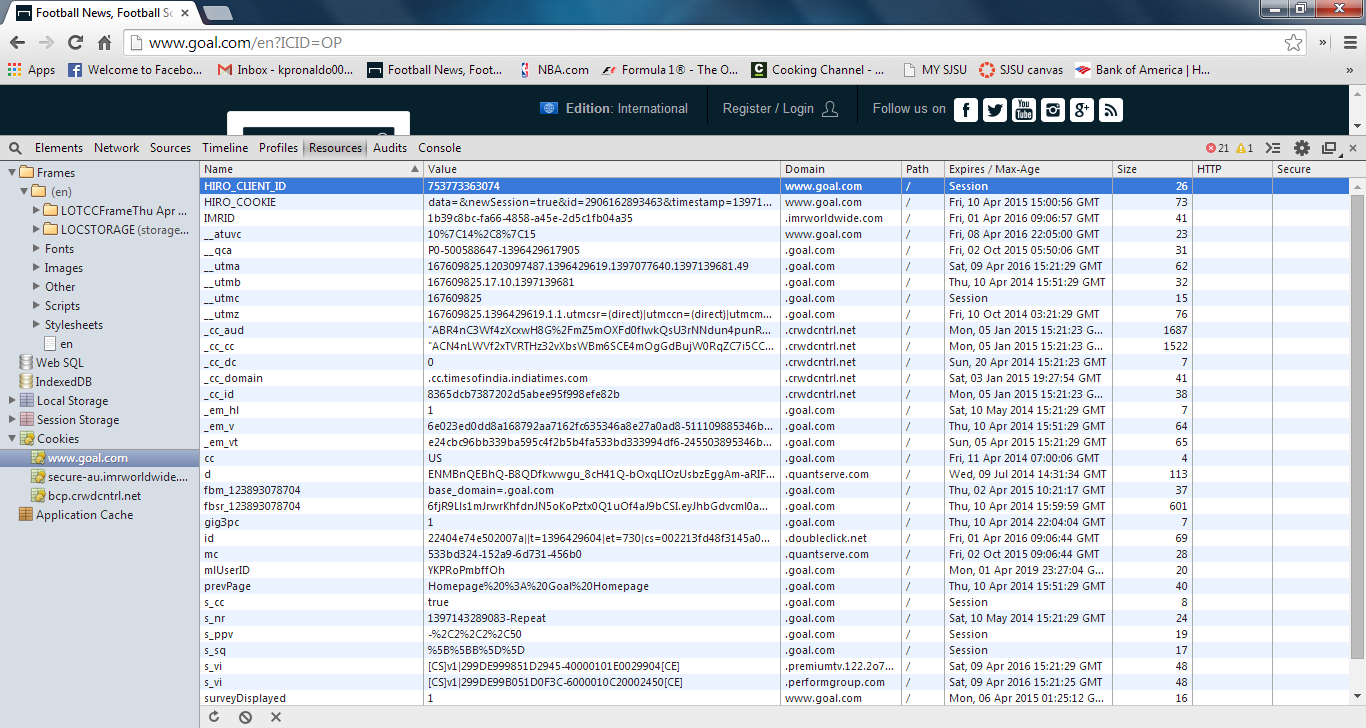


Figure 30 - Session Cookie Highlighted

## Ad Server Component

The figure shows the sequence diagram of the ad server component created again using Quick Sequence Diagram Editor. The URL from the HTTP Request Header is compared with a blacklisted set of Advertising server. A comparison is made. If a match is found, the HTTP Request is dropped. A new Request is sent to the Netgear Ad server on the category from the category list from the Category Component. As mentioned before, the sequence diagram is for the implementation that has been done in the Java programming language.

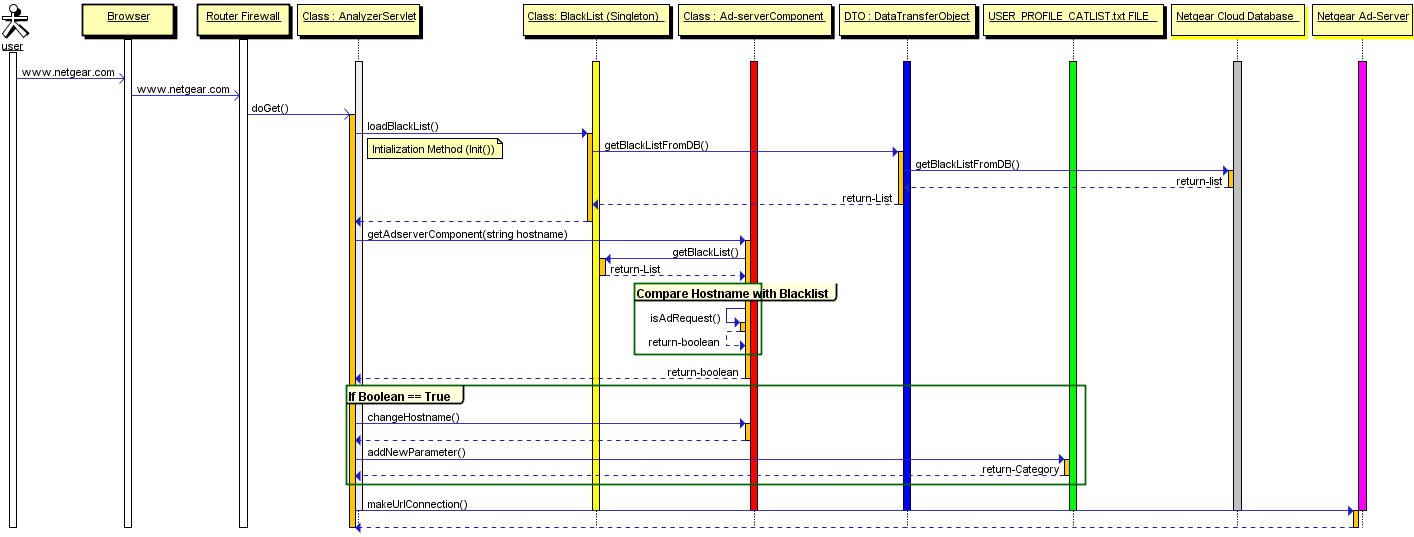


Figure 31 - Sequence Diagram for the Advertising Component

The following shoes the logs for the C++ implementation using Boost. For this particular example, the CNN news website was visited. In our blacklist.txt file, their ad portal, ads.cnn.com was added. The Turner ad server was not added to this list. We could see that the CNN ad server was blocked while the Turner ad server was let through to retrieve the advertisement.

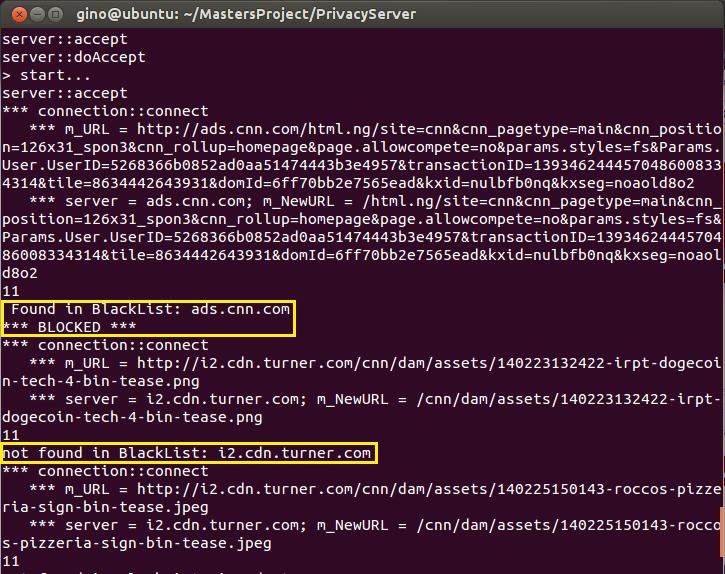


Figure 32 - CNN ad server is blocked while Turner ad server is not blocked

## Category Component

The following figure shows the sequence diagram for the URL category component. This component is used to keep a category of websites and also maintain the number of times the website has been browsed. The sequence diagram and titles "USER\_PROFILE\_CATLIST.TXT FILE", "UNDEFINED\_URL.TXT FILE" and "URL\_CAT\_MAPPING\_TABLE" are unique to the Java implementation of the project.

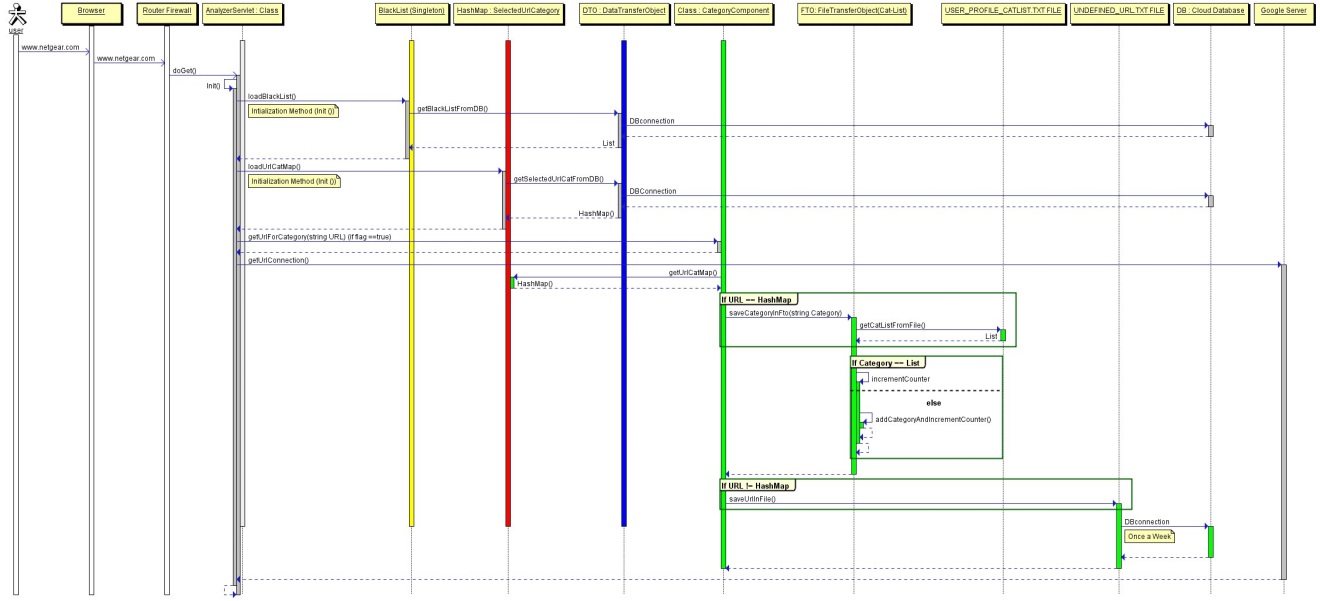


Figure 33 - Category Component Sequence Diagram

**"URL\_CAT\_MAPPING\_TABLE"**

The purpose of this table is to store the website names with its known categories. This is not a part of the modules needed to be created by us. This feature is supposed to be got from a third party such as Zvelo[30]. This table is then used to update the USER\_PROFILE\_CATLIST.TXT file to feed it with the new categories.

**“USER\_PROFILE\_CATLIST.TXT FILE”**

This file is unique to the Java implementation of the code. The purpose of this file is to maintain a list of categories visited by the user. The file has two parameters - The category name and the category count. Each time the a category is hit, the counter is incremented. There are two conditions to update a file.

* Compare URL with the set of categories assumed to have been got from a third party. If there is a match, then increase the counter.
* Else update the UNDEFINED\_URL.TXT. This file in the actual model is supposed to interact with the third party to categorize the URL's and update the USER\_PROFILE\_CATLIST.TXT file with the categories. This also is not part of the project since third party is involved.

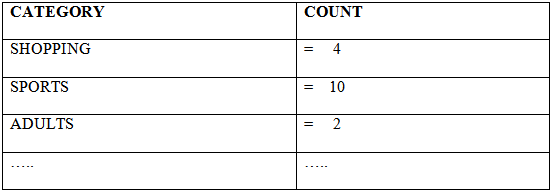


Figure 34 - USER\_PROFILE\_CATLIST.TXT example

**“UNDEFINED\_URL.TXT FILE”**

The purpose of this file is to maintain a list of undefined URLs. This file is supposed to sync up with the third party categorization tool for categorizing the undefined URLs. These categories are then supposed to use URL\_CAT\_MAPPING TABLE and update the USER\_PROFILE\_CATLIST.TXT file and monitor the user.

In the C++ implementation of the code, the category component looks at the catmap.txt file. This file has the URL and the category associated with it. From the URL, we get the category. Based on this category, we update the catcount.txt file. This file contains the count of each category depending on the category that has been hit in the catmap.txt. The catcount.txt and catmap.txt files are shown in the figure below. Since this is a prototype, we have considered only ten categories. In the real model, there will be more categories. In case the category is not present in the file, nothing is done in the prototype.

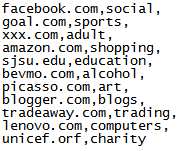


Figure 35 - catmap.txt

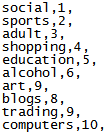


Figure 36 - catcount.txt

If we access any one of the websites in catmap.txt, the corresponding category should increase in the catcount.txt file.

# TEST CASES

## Test Case for Cookie Component

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Step Description | Expected Result | Pass/Fail |
| 1 | Go to Tools 🡪 Developer Tools 🡪 Cookies on Google chrome browser. Access a website | You should be able to see all the cookies of the website. Make a note of the cookies and number of cookies. | Pass |
| 2 | Go to Proxy settings on Google chrome browser. Set LAN setting to name = “localhost” and port number = “8080”. | This step will access website through proxy server. Website should not be accessible because the proxy server is not up. | Pass |
| 3 | Start the proxy server on port 8080. | The proxy server should be up and any web page accesses should show relevant HTTP headers. | Pass |
| 4 | Access a website on the Google chrome browser now. | Should be able to access website. The corresponding Http Headers and Cookies should be seen in the proxy server console window | Pass |
| 5 | Compare the cookies and number of cookies from Step 1. | The number of cookies set on the browser is lesser than before, indicating that the cookies have been blocked at the proxy server. | Pass |

## Test Case for Ad-Server Component

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Step Description | Expected Result | Pass/Fail |
| 1 | Open the browser and open the options window |  | Pass |
| 2 | In the settings of the browser enable the proxy to port 8080 | Web traffic should be stopped until the proxy server is enabled | Pass |
| 3 | Run the proxy code | Web traffic should work fine | Pass |
| 4 | Access a webpage usually filled with ads | The ads shouldn't be seen since they are blocked. | Pass |
| 5 | Look for the Netgear Logo | Certain Advertisements should be replaced by Netgear Logo | Pass |

## Test Case for URL Component

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Step Description | Expected Result | Pass/Fail |
| 1 | Open the browser and open the options window |  | Pass |
| 2 | In the setting s of the browser enable the proxy to port 8080 | Web traffic should be stopped until the proxy server is enabled | Pass |
| 3 | Run the boostproxy code or enable it on the router | Web traffic should work fine | Pass |
| 4 | Access a webpage whose name is in the catmap.txt file | The web page will open | Pass |
| 5 | Open the catcount.txt file | The count against the category of the webpage opened should be incremented by 1 | Pass |
| 6 | If the webpage is not in the catmap.txt file | The category others should increase by 1 and the hostname should be stored for future use | Pass |

# DEPLOYMENT

This section explains the Deployment of the code that has been written. This document has two sections - deploying the code on Java using Apache Tomcat server and Deploying the code using the Boost C++ libraries. Set the browser setting as required following the instructions below.

## Browser Settings

* For Google Chrome[32]:

1. Open Google Chrome Browser, Click "Customize and Control Google Chrome" menu.

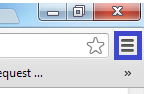


Figure 37 - Settings tab on Google Chrome

1. Click "Settings" button.

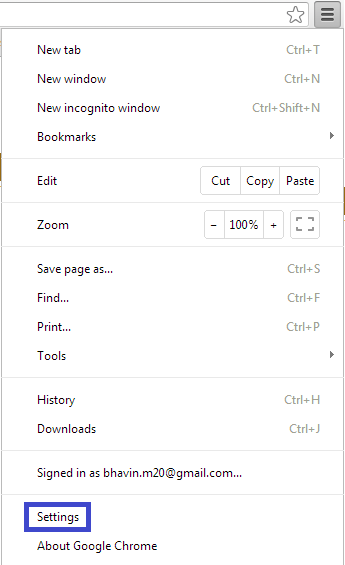


Figure 38 - Setting Drop down Menu

1. Go to “More Advanced Settings” in the bottom of the page, select “Change Proxy Settings” under Network.

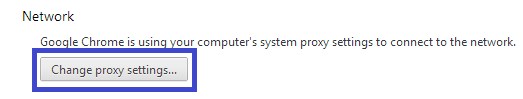


Figure 39 - Change proxy settings option

1. Go to LAN Settings, uncheck “Automatically Detect Settings” box and check “Use a Proxy server for your LAN”.

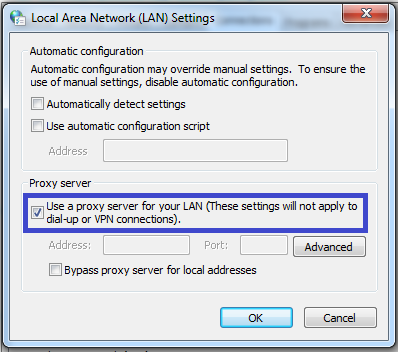


Figure 40 - select your proxy server

1. Click on “Advanced” and set HTTP Proxy Address as “localhost” and port number “80”.

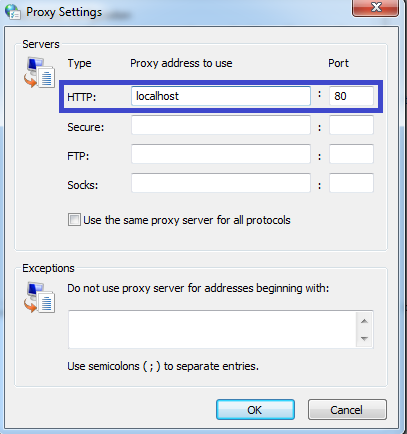


Figure 41 - HTTP packets to go through your proxy server running on port 80

* For Internet Explorer[33]:

1. Open Internet Browser, and click on “Tools”.



Figure 42 - Settings in Internet Explorer

1. Select “Internet Options” from the drop down list.

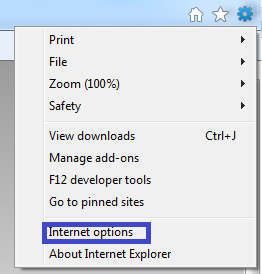


Figure 43 - Choose Internet options

1. Click on “Connections” tab and then follow the instructions same as Google Chrome.

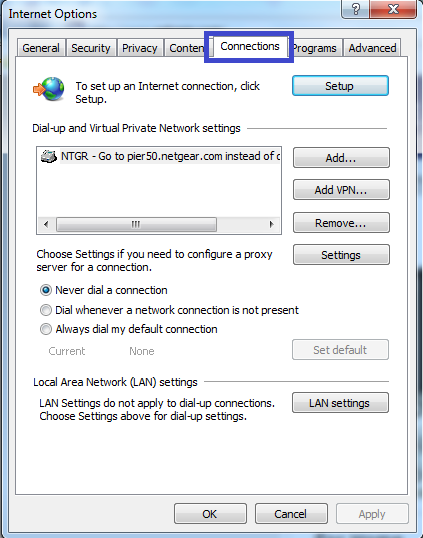


Figure 44 - Connections tab

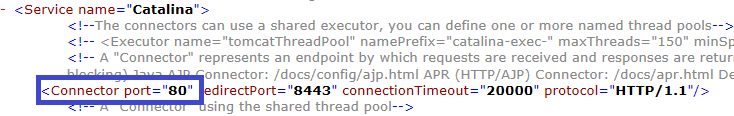
# Deployment on Apache Tomcat Server - Java Application

Requirements -

* Download Apache Tomcat and Install it in your computer.
* <http://tomcat.apache.org/download-70.cgi>[31]
* war file package was created in eclipse IDE.

Detailed Steps -

* After Browser Settings are done, go to the Tomcat location in your computer. Example “C:\tomcat\tomcat7.1\apache-tomcat-7.0.41\webapps”.
* Deploy your MyProxyServlet.war in the webapps folder and rename it as “ROOT.war” and change the originally ROOT folder name as “ROOT-1”.
* After that go to “conf” folder in Tomcat, for example “C:\tomcat\tomcat7.1\apache-tomcat-7.0.41\conf”.
* Open “Server.xml” and change “Connector port = 80”. In Java Application we run it on port 80, not 8080.



* Start the Tomcat Server by double clicking on Startup.bat “C:\tomcat\tomcat7.1\apache-tomcat-7.0.41\bin\Startup.bat”.
* Once the server starts, open the browser and type in any website name, the HTTP packets would flow through this proxy server.

## Deployment using Boost Libraries - C++ application

Requirements -

* Microsoft Visual Studio C++
* Boost C++ libraries version 1.53.0 or higher
* mginw tool chain for installing Boost

Detailed Steps -

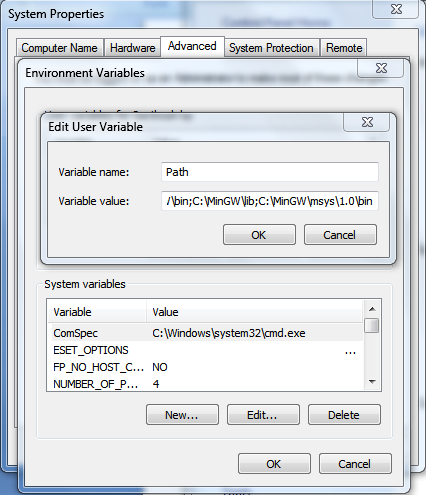
* Download Boost library version 1.53.0 from http://www.boost.org/users/history/version\_1\_53\_0.html[34]
* Extract the Boost Package in to your local folder.
* Open Command Prompt and go to that folder.
* Download and install mginw for windows and set up the environmental variables. Add MinGW\lib and msys\1.0\bin to the environment variables -  
  

Figure 45 - Setting Environment Variables

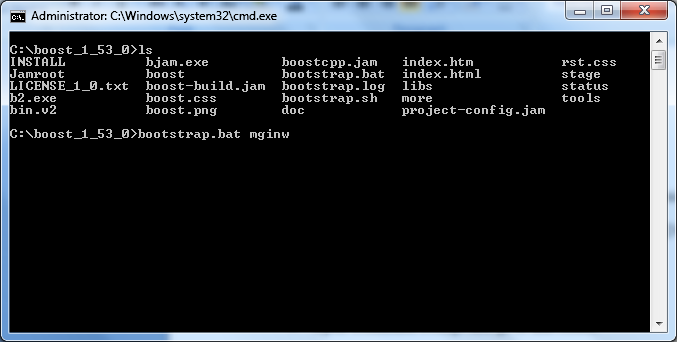
* Run the command **bootstrap mginw**

Figure 46 - Bootstrap command with mginw toolset

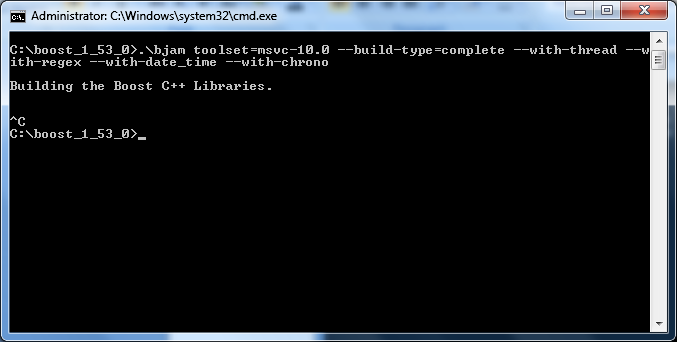
* This will create a boost build that needs to be run to get all the boost library files.
* Enter the command **.\bjam toolset=msvc-10.0 --build-type=complete --with-thread --with-regex --with-date\_time --with-chrono.** This will build the required boost libraries.  ****

Figure 47 - Run bjam with the required options only

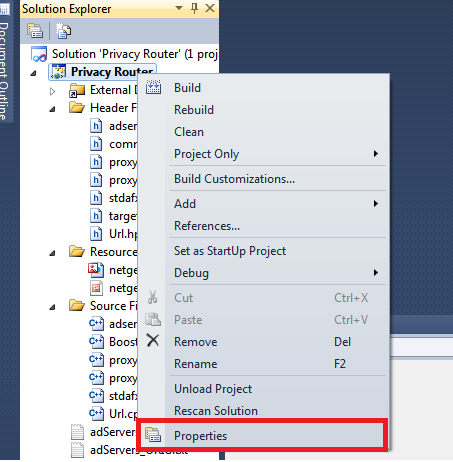
* Now open Visual Studio C++
* From Visual Studio's File menu, select New > Project…
* In the left-hand pane of the resulting New Project dialog, select Visual C++ > Win32.
* In the right-hand pane, select Win32 Console Application (VS8.0) or Win32 Console Project (VS7.1).
* Add all the program code in the solution panel.
* Right-click **Privacy Router** in the Solution Explorer pane and select Properties from the resulting pop-up menu  
  

Figure 48 - Properties in Visual Studio C++

* In Configuration Properties > C/C++ > General > Additional Include Directories, enter the path to the Boost root directory, for example

C:\boost\_1\_53\_0

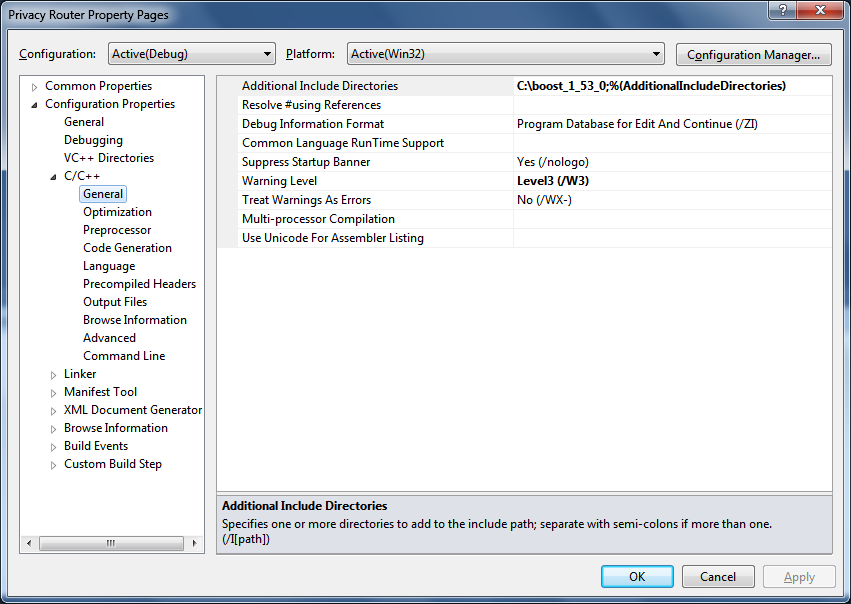


Figure 49 - Additional Include directories option

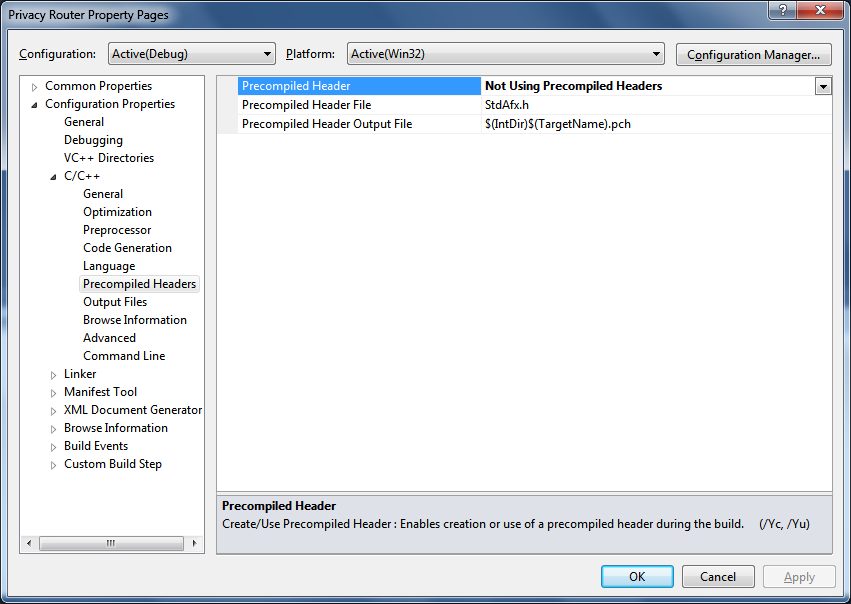
* In Configuration Properties > C/C++ > Precompiled Headers, change Use Precompiled Header (/Yu) to Not Using Precompiled Headers.   
  

Figure 50 - Setting no Precompiled Headers

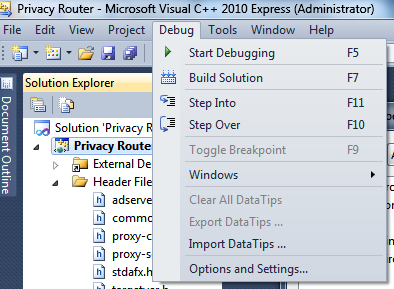
* From the Debug menu, select Build Solution or hit F7. Make sure the build goes through successfully in the console.
* Run the Code using Start Debugging Option or hit F5.  
  ****

Figure 51 - Build and Run options under the Debug tab

# SUMMARY

The internet has become a vast resource of knowledge in this world. It is ever growing and at the same time becoming a part of the users daily errands. Ever since the internet has become easily accessible on mobile cell phone devices, it has definitely become a part of our daily life. With over a billion users, it is also an ideal platform for advertising. Identifying a user among millions and sending him the right advertisement is a tough task. But a very valuable one. Every year, millions of dollars are spent in the online advertising industry to ensure advertising is effective. With the importance of the advertising industry being so high, the users privacy concern is something of major concern to anyone. Any information that can be obtained on the user is useful information. Creating a demographic profile is that much more important than worry about his or her privacy.

In the "Netgear Privacy and Advertising Platform" Project, we have come up with a logical and optimal solution for the user. His or her privacy is given utmost importance. The user cannot be tracked any more by any of the third party ad servers or unreliable ad servers. The Tracking Cookies, they try to place on the user's browser are intercepted by the application. Since his internet experience shouldn't be affected, he is being monitored, but by a much more reliable a safe entity, his home router. The user has complete control over the profile that is created for him. Since this private profile is used to deliver the advertisements relevant to the user, he feels safer because he is not tracked and gets the required and relevant advertisements he needs.

# RESULTS AND CONCLUSIONS

The application was run as a standalone application on the Windows platform and Ubuntu platforms. When the websites were accessed, the advertisements were blocked. The Netgear Logo was displayed in its stead on certain scenarios. The Netgear Logo was displayed with greater consistency when implemented using Java programming language.

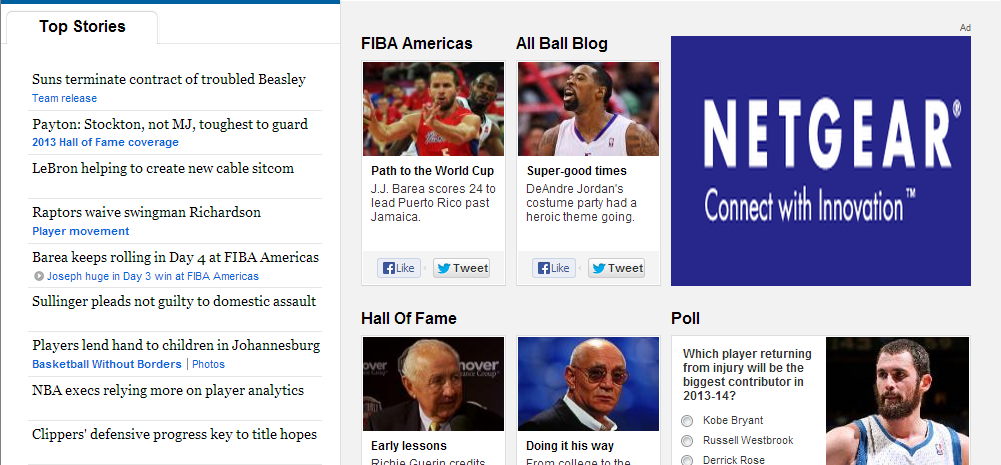


Figure 52 - Advertisements blocked on nba.com

The Cookies are also blocked. A majority of the Tracking Cookies are removed. In the figure, only certain Cookies such as \_\_utma,\_\_utmb, etc. are seen. These are the Google Analytics Cookies for Google Chrome which are sent as part of the HTTP requests. To conclude, the application provides privacy by eliminating Cookies, replacing ads and categorizing websites.

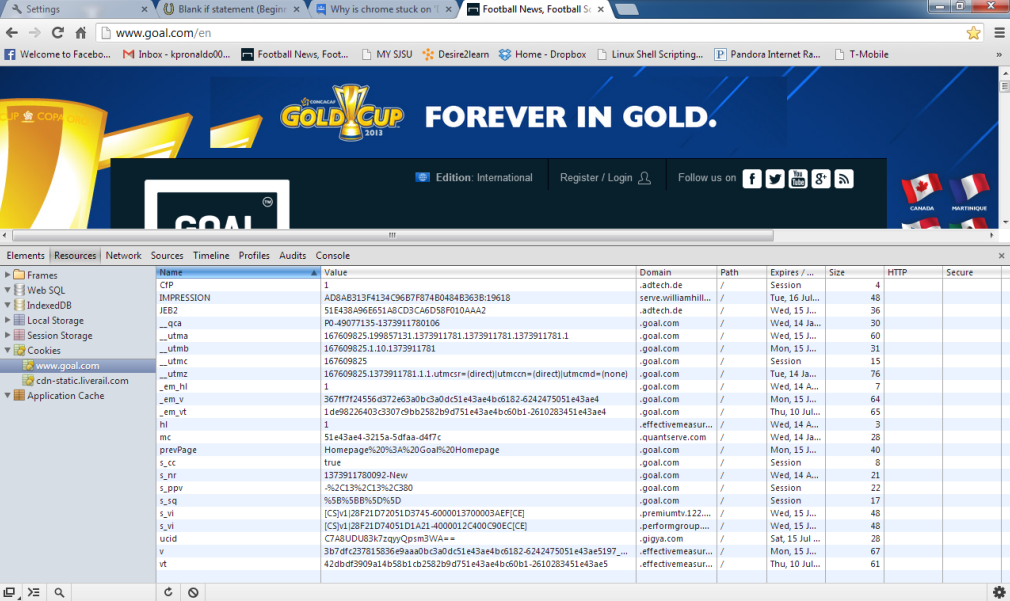


Figure 53 - Cookies before running the application

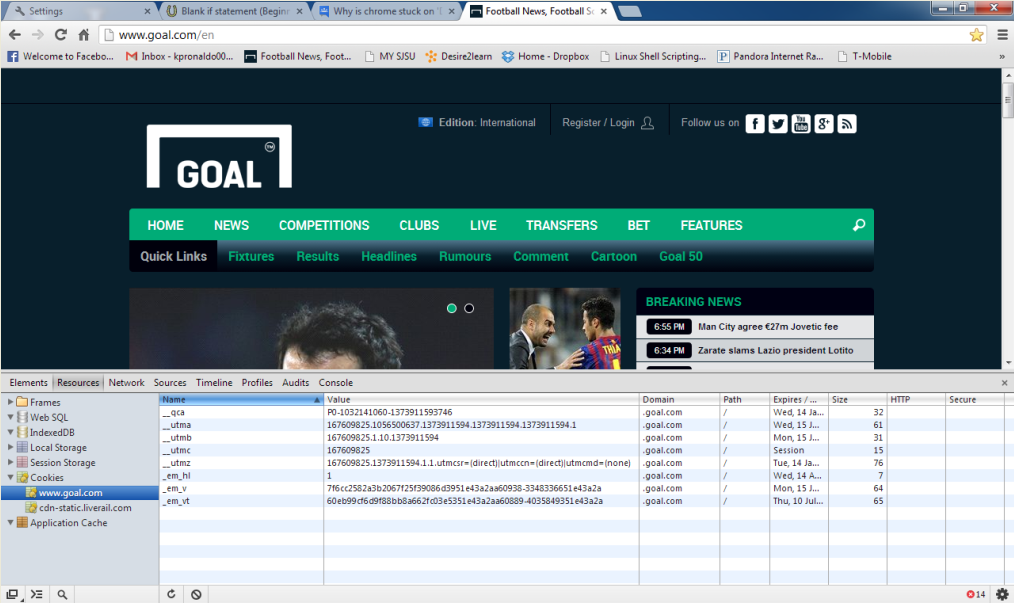


Figure 54 - Cookies blocked after running through the application

# RECOMMENDATIONS FOR FUTURE RESEARCH

This project is a prototype of the model. The real world model will have a Netgear Advertising Server. Generating advertisement categories and displaying those advertisements for the advertisement server components of the application should be done by this Netgear advertising server. Access times and delay times should be given priority in the design of this server.

Implementing a web categorization algorithm instead of using a third party service will be a smart decision. If the algorithm is small enough, it could be put on the router and run. Else it can be one of the features of the Netgear Advertising server mentioned before. Some of the algorithms used in implementation of Webpage classification are k-Nearest Neighbor algorithm(kNN algorithm), Bayesian algorithm, Support Vector Machine(SVM), Neural Networks, Decision Trees[35]. Researchers in Department of Computer Science and Engineering at the Mississippi State University created an automated document classification system that can be used for Webpage classification called Webdoc[36]. Web page classification can be subdivided in to subject classification, functional classification, sentiment classification, and other types of classification. Subject Classification marks the webpage based on the subject like arts, business, sports, etc. Classifying the webpage based on the role it plays such as a personal homepage, course page, admission page, etc. Classification can also be done based on genre and so on. Classification can be also on the number of classes as shown in figure. They can be binary or multiclass classification. Classification can also be hierarchical or flat classification[37]. One of these techniques can be used for the classification depending on which is more efficient.

Another recommendation is to used Machine Learning Algorithms for creating the URL list of categories. Some of the algorithms helpful are Naive Bayes and SVM[38].



Figure 55 - Classification types[37]



Figure 56 - Flat and Hierarchical classification[37]

# DIVISION OF WORK

|  |  |
| --- | --- |
| **Contribution of work** | **Name of the person** |
| Abstract | Bhavin Niranjan Mehta |
| Introduction | Bhavin Niranjan Mehta |
| Hypothesis | Bhavin Niranjan Mehta |
| State Of Art | Santhosh Kumar Pattabhiraman |
| Implementation | Santhosh Kumar Pattabhiraman |
| Project Design | Santhosh Kumar Pattabhiraman |
| Test Cases | Santhosh Kumar Pattabhiraman |
| Deployment | Bhavin Niranjan Mehta |
| Summary | Bhavin Niranjan Mehta |
| Results and conclusions | Santhosh Kumar Pattabhiraman |
| Recommendations | Santhosh Kumar Pattabhiraman |

# REFERENCES

[1] ["Tracking Cookie." *Endpoint, Cloud, Mobile & Virtual Security Solutions*. N.p., n.d. Web. 10 Oct. 2013. <http://www.symantec.com/security\_response/writeup.jsp?docid=2006-080217-3524-99>.](http://www.bibme.org/)

[2] Aleecia McDonald, Lorrie Faith Cranor, America's Attitudes About Internet Behavioral Advertising Practices, Carnegie Mellon University, Pittsburg, PA, October 4, 2010

[3] ["Analytics." *Google Official Website â“ Web & Reporting*. N.p., n.d. Web. 25 Sept. 2013. <http://www.google.com/analytics/>.](http://www.bibme.org/)

[4] ["About Adblock Plus." *About Adblock Plus*. N.p., n.d. Web. 10 Oct. 2013. <https://adblockplus.org/en/about>.](http://www.bibme.org/)

[5] ["Change Internet Explorer Privacy settings - Microsoft Windows Help."*windows.microsoft.com*. N.p., n.d. Web. 30 Sept. 2013. <http://windows.microsoft.com/en-us/windows7/change-internet-explorer-privacy-settings>.](http://www.bibme.org/)

[6] ["Networking Basics: What You Need To Know." Cisco. http://www.cisco.com/cisco/web/solutions/small\_business/resource\_center/articles/connect\_employees\_and\_offices/networking\_basics/index.html (accessed April 6, 2014).](http://www.bibme.org/)

[7] [Lammle, Todd. "Internetworking." In CCNA Cisco certified network associate study guide, 1 - 41. 5th ed. San Francisco, Calif.: SYBEX, 2005.](http://www.bibme.org/)

[8] [Mir, Nader F.. "Routing and Internetworking." In Computer and communication networks, 171 to 185. Upper Saddle River, NJ: Pearson Hall, 2007.](http://www.bibme.org/)

[9] ["Router Manufacturers Contact List." Router Manufacturers Contact List. http://home.windstream.net/slashdot/manufacturers.html (accessed April 6, 2014).](http://www.bibme.org/)

[10] Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1", RFC 2616, June 1999.

[11] Kristol, D. and L. Montulli, "HTTP State Management Mechanism", RFC 2965, October 2000.

[12] http cookies: standards, privacy and politics by david m kristol, Bell Labs, Lucent technologies, ACM Transactions on Internet Technology, Vol. 1, No. 2, November 2001

[13] ["Cookies: Frequently Asked Questions." Cookie FAQ. http://www.aboutcookies.org/default.aspx?page=5 (accessed January 26, 2014).](http://www.bibme.org/)

[14] ["Amazon.com Privacy Notice." Amazon.com Help:. http://www.amazon.com/gp/help/customer/display.html?nodeId=468496 (accessed February 4, 2014).](http://www.bibme.org/)

[15] ["What is a cookie? - All about Cookies." What is a cookie? - All about Cookies. http://www.allaboutcookies.org/cookies/ (accessed Jan 20, 2014).](http://www.bibme.org/)

[16] ["Usage of Session Cookies for websites." Usage Statistics of Session Cookies for Websites, April 2014. http://w3techs.com/technologies/details/ce-sessioncookies/all/all (accessed March 6, 2014).](http://www.bibme.org/)

[17] ["Usage of Persistent Cookies for websites." Usage Statistics of Persistent Cookies for Websites, April 2014. http://w3techs.com/technologies/details/ce-persistentcookies/all/all (accessed March 6, 2014).](http://www.bibme.org/)

[18] Vincent Toubiana and Vincent Verdot, Show me your cookie and I'll tell you who you are, Alcatel-Lucent Bell Labs-Application domain, Aug 2011

[19] Heng Wu, Weiting Chen and Zhongjie Ren, Securing cookies with a MAC address encrypted key ring, 2010

[20] [Rosen, Jeffrey. "Who Do Online Advertisers Think You Are?." The New York Times. http://www.nytimes.com/2012/12/02/magazine/who-do-online-advertisers-think-you-are.html?pagewanted=all&\_r=1& (accessed March 5, 2014).](http://www.bibme.org/)

[21] ["AdExchanger: News and Views on Data-Driven Digital Advertising." AdExchanger News and Views on DataDriven Digital Advertising. http://www.adexchanger.com/online-advertising/real-time-bidding/#google (accessed March 6, 2014).](http://www.bibme.org/)

[22] ["How Does It Work?." BlueKai:. http://www.bluekai.com/consumers\_howdoesitwork.php (accessed March 10, 2014).](http://www.bibme.org/)

[23] ["Facebook Exchange: A New Way For Advertisers To Target Specific Users With Real-Time Bid Ads." TechCrunch. http://techcrunch.com/2012/06/13/facebook-exchange/ (accessed March 11, 2014).](http://www.bibme.org/)

[24] ["Google Analytics." Cookie Usage on Websites. https://developers.google.com/analytics/devguides/collection/analyticsjs/cookie-usage (accessed March 12, 2014).](http://www.bibme.org/)

[25] ["Types of cookies used by Google Privacy & Terms Google." Types of cookies used by Google Privacy & Terms Google. https://www.google.com/policies/technologies/types/ (accessed March 12, 2014).](http://www.bibme.org/)

[26] ["TECHNOLOGY LAW CORNER Online Tracking, Profiling and Targeting - Behavioral Advertisers Beware!." Online Tracking, Profiling and Targeting. http://www.ecommercetimes.com/story/73966.html (accessed March 13, 2014).](http://www.bibme.org/)

[27] Ashish Kumar Singh and Vidyasagar Potdar, "Blocking Online Advertising – A State of the Art", Digital Ecosystems and Business Intelligence (DEBI) Institute, Curtin University of Technology

[28] Richard E. Chatwin, "An Overview of Computational Challenges in Online Advertising", 2013

[29] Lorrie Faith Cranor, "Can Users Control Online Behavioral Advertising Effectively", April 2012

[30] ["Zvelo - We categorize the web."*Website, URL and Content Categorization*. N.p., n.d. Web. 15 June 2013. <https://zvelo.com/>.](http://www.bibme.org/)

[31] ["Apache Tomcat." *- 7 Downloads*. N.p., n.d. Web. 10 Jan. 2014. <http://tomcat.apache.org/download-70.cgi>.](http://www.bibme.org/)

[32] ["Google Chrome Tutorial." *Google Chrome Proxy Settings*. N.p., n.d. Web. 10 Jan. 2014. <http://www.googlechrometutorial.com/google-chrome-advanced-settings/Google-chrome-proxy-settings.html>.](http://www.bibme.org/)

[33] ["Change proxy server settings in Internet Explorer." *windows.microsoft.com*. N.p., n.d. Web. 18 Jan. 2014. <http://windows.microsoft.com/en-us/windows/change-internet-explorer-proxy-server-settings#1TC=windows-7>.](http://www.bibme.org/)

[34] "Boost C++ Libraries." Version 1.53.0. N.p., n.d. Web. 19 Mar. 2014. < http://www.boost.org/users/history/version\_1\_53\_0.html>

[35] W.A Awad, Machine Learning Algorithms in web page classification, Math.& Computer Science Dept., Faculty of Science, Port Said University, Egypt.

Scientific Research Group in Egypt (SRGE), October 2012

[36] Yong Wang, Julia Hodges, Bo Tang, Classification of Web Documents Using a Naive Bayes Method, Department of Computer Science & Engineering, Mississippi State University, 2003

[37] Xiaoguang Qi and Brian D, Web Page Classification: Features and Algorithms, Lehigh University, ACM Computing Surveys, Vol. 41, No. 2, Article 12, Publication date: February 2009.

[38] M. Indra Devi, Dr R.Rajaram and K.Selvakuberan, Machine Learning Techniques for Automated Web Page Classification using URL Features, Thiagarajar College of Engineering, International Conference on Computational Intelligence and Multimedia Applications 2007