

Project Number: CEW-0701

Extension Based Privacy Protection

A Major Qualifying Project Report:
submitted to the Faculty of the
WORCESTER POLYTECHNIC INSTITUTE
in partial fulfillment of the requirements for the
Degree of Bachelor of Science

by

Christopher Gianfrancesco

Adam Fiske
and

David Marsh

March, 2007

Approved:

Craig E. Wills, Advisor

Abstract

Recently, Internet privacy has become a growing concern among Internet users. These users should have a way to prevent unwanted content from being downloaded onto their computers. This project looks at one possible solution to this problem. The project team developed a completely user-defined whitelisting tool which prevents said content from being downloaded. This tool was tested subjectively and objectively across a number of different websites and was found to have a great impact on a user's browsing experience.

Table of Contents

Abstract	1
Table of Figures	4
Table of Tables	5
1. Introduction	6
2. Background	9
2.1 General History of Firefox Extensions	9
2.2 Pagestats	10
2.3 NoScript.....	10
2.4 Adblock.....	11
2.5 Adblock Plus.....	11
2.6 Summary.....	11
3. Design	13
3.1 Summary.....	14
4. Implementation	15
4.1 Firefox Extensions	15
4.2 Javascript Services.....	15
4.3 How Adblock and Adblock Plus Work	16
4.3.1 <i>Adblock Structure</i>	17
4.3.2 <i>Adblock Plus 0.7 Structure</i>	18
4.4 Our Extension	18
4.5 Migration from Adblock to Adblock Plus	20
4.6 Implementation Issues	20
4.7 Summary.....	21
5. System in Action	22
5.1 Choosing the level of allowable objects	22
5.2 Viewing the results of a page load.....	23
5.3 Allowing / Disallowing Specific Hosts	24
5.4 Summary.....	25
6. Results	26
6.1 Rating System.....	26
6.1.1 <i>Rating 1 Example</i>	27
6.1.2 <i>Rating 2 Example</i>	28
6.1.3 <i>Rating 3 Example</i>	29
6.1.4 <i>Rating 4 Example</i>	30
6.2 Analysis	31
6.2.1 <i>News sites</i>	31
6.2.2 <i>Forum Sites</i>	33
6.2.3 <i>Webcomics</i>	34
6.2.4 <i>Shopping Websites</i>	36
6.2.5 <i>Flash Websites</i>	37
6.2.6 <i>Information Websites</i>	39
6.2.7 <i>Search Engines</i>	41
6.3 User Surveys.....	42
6.4 Summary.....	43
7. Conclusions	44
7.1 Future Work.....	44
Works Cited	46
Appendices	47

Appendix A: User Surveys	47
<i>Survey 1</i>	47
<i>Survey 2</i>	47
<i>Survey 3</i>	47
Appendix B – Data Gathered	49
<i>First Party</i>	49
<i>Second Party</i>	50
<i>Third Party</i>	51

Table of Figures

Figure 4-1 Adblock / Adblock Plus Program Flow.....	17
Figure 4-2 Our Extension Flow Control.....	19
Figure 5-1 Preferences Window.....	22
Figure 5-2 Page Load Window	23
Figure 5-3 Action Buttons.....	24
Figure 5-4 Allowing and Blocking Hosts.....	25
Figure 6-1 www.xkcd.com operating at “Allow All”	27
Figure 6-2 www.xkcd.com operating at “First Party”.....	27
Figure 6-3 www.ign.com operating at “Allow All”	28
Figure 6-4 www.ign.com operating at “First Party”	28
Figure 6-5 http://news.bbc.co.uk at “Allow All”	29
Figure 6-6 http://news.bbc.co.uk at “First Party”.....	29
Figure 6-7 http://news.google.com at “Allow All”	30
Figure 6-8 http://news.google.com at “First Party”.....	30
Figure 6-9 News Sites – First Party.....	32
Figure 6-10 News Sites – Second Party	32
Figure 6-11 Forums – First Party	33
Figure 6-12 Forums – Second Party.....	34
Figure 6-13 Webcomics – First Party.....	35
Figure 6-14 Webcomics – Second Party	35
Figure 6-15 Shopping Sites – First Party.....	36
Figure 6-16 Shopping Sites – Second Party.....	37
Figure 6-17 Flash Game Sites – First Party	38
Figure 6-18 Flash Games Sites – Second Party.....	38
Figure 6-19 Information Sites – First Party.....	40
Figure 6-20 Information Sites – Second Party	40
Figure 6-21 Search Sites – First Party.....	41
Figure 6-22 Search Sites – Second Party	42

Table of Tables

Table 6-1 News Sites Data	31
Table 6-2 Forum Sites Data.....	33
Table 6-3 Webcomics Data	35
Table 6-4 Shopping Data.....	36
Table 6-5 Flash Sites Data	38
Table 6-6 Information Sites Data	39
Table 6-7 Search Sites Data	41

1. Introduction

In recent years, the Internet has become a household convenience, with nearly anybody able to access it in one way or another. Unfortunately, many have sought to exploit this fact, creating spyware, unwanted advertisements and more to unleash upon the Internet's widespread user base. While there are tools to help combat these exploits, not all unwanted content can be taken into account. Even if this were not the case, preventing the download of extraneous content is highly preferable to discovering it once it is already on a user's system. It may have already caused some damage or the user may just not have wanted it downloaded in the first place.

In addition to this unwanted, content which sites may be able to put onto a computer, we must also be wary of what can be retrieved from one's computer as well. These days, an increasing number of services are made available online (such as shopping, credit card/bank account management, and more). With these services, an increasing amount of sensitive data is being transferred through our browsers. Sites may monitor users and gather information about their browsing patterns and habits. With this data sites can generate ads based on a user's specific Internet history and this content would be invasive along with being unwanted. Thus, we must be exceedingly careful about the sites we visit and the external content attached to those sites, as we are sometimes unsure of exactly how much a website can learn about us without our consent or knowledge.

To combat the increasing threat of malicious content and privacy invasion, we must be more cautious about what we do and do not download onto our computers. An idea must be formed about what types of content can be trusted to be safe, and what content should be held to a higher level of scrutiny. Once it is decided which content falls under which category,

appropriate action can be taken to better protect the Internet user.

We propose that content offered on any given webpage can be divided into three distinct categories: first-party, second-party and third-party. First-party content is any content that comes from exactly the same domain name as the page the user requested. That is, if one navigates to www.cnn.com, only objects whose domain is exactly www.cnn.com are considered to be first-party content. Second-party content is any which shares the primary part of the domain name of the requested page, generally a secondary server used by the host. For instance, i.walmart.com is the image host for www.walmart.com, and is considered a second-party provider for any www.walmart.com page. Finally, third-party content is any object whose domain name does not fit the criteria for first- or second-party status. Therefore, third-party content will come from a site whose domain name is entirely different than the domain of the viewed webpage.

With these definitions in mind, the team decided that they would examine the capabilities of the Firefox web browser. One of Firefox's key features that separates it from other browsers, such as Internet Explorer, is the ability to create extensions and plug ins that can change the way the browser works. The team chose to explore the option of writing an extension for Firefox. One which recognizes the status of all content requests and responds accordingly could be a powerful tool in providing users with an extra level of security and privacy.

With this extension, the user is able to specify their level of trust in the sites they visit by the first-, second- and third-party definitions. Thus, one could choose only to trust content from the originating server; allow the server to contact secondary servers under the same host; or even allow a site to contact any tertiary servers. By giving users the option to decide for themselves, the project team believes this plug-in effectively filters out unwanted content while at the same time providing an overall safer and pleasant browsing experience.

The extension has also been evaluated for its effectiveness and usability on a variety of web sites. Between rigorous trials by the project team as well as subjective end-user testing, the plausibility of filtering out certain levels of content on commonly-used web sites has been evaluated. These tests have determined the plug-in's ability to remove most of the extraneous, unwanted, and possibly distrusted content found on many web sites, while still preserving the overall usability of the sites as much as possible.

The remainder of this paper describes the process the team went through to design and implement the above described plug-in. Chapter 2, Background, discusses relevant work done in the area of privacy and content control on the Internet in the forms of other extensions created. Chapter 3, Design, discusses the overall design and approach we took with our extension, as well as other alternatives that we considered. Chapter 4, Implementation, describes in greater detail the final design and the process taken to implement it. Chapter 5, System in Action, leads the reader through examples of different features of our extension. Chapter 6, Results, details the team's findings from testing the extension using defined objective ratings. Chapter 7, Summary and Future Work, restates, in a concise form, our extension and its outcomes as well as outlines future research and implementations that could be based off of this particular project, or in areas related to security and content control.

2. Background

The goal of this project is to write an extension for the Firefox web browser that allows for more intelligent filtering options for third party content than currently existing extensions. However, at the onset of this project, none of the group members had any prior experience creating extensions for Firefox, and only minimal experience writing in Javascript, the language used for Firefox extensions. Thus, the first step before the team could proceed in writing the extension was to study existing Firefox extensions to gain knowledge about how they work and how they are written.

2.1 General History of Firefox Extensions

Firefox was designed as an extensible browser so that all sorts of customizations and added functionality could be added. Extensions may take many forms, as well – they can exist as additional buttons or menus for the user to operate, features that run in the background as one browses, or even fundamentally change the interface of Firefox. For the purposes of this project, we studied a number of extensions which focused on observing and modifying the way webpages were requested and loaded as a user moved from page to page.

Three major extensions were studied for this purpose: the Pagestats extension, used to track requests made by a loading webpage (Dedeo, 2007); Adblock, which uses simple whitelisting and blacklisting logic to filter content (Adblock, 2007); and NoScript, another content blocker that focuses on Javascript objects (NoScript, 2007). They helped supply a working knowledge of the Javascript language and Firefox extensions in general, as well as having many features directly relevant to the proposed extension. While Adblock ended up

being the biggest contributor to the rest of the project, all three had their own particular contribution to the group's research.

2.2 Pagestats

Pagestats, an extension that logs page requests and responses, offered some useful insight, but was not as influential as the other two extensions. Pagestats tracks webpage loading statistics by observing outgoing HTTP requests and incoming responses, and it was originally thought that this would be a useful tool in writing the extension. After further study, however, Adblock uses the same types of event listeners as Pagestats, and in a way more relevant to this project, so Adblock ended up being a more considerable contributor. Pagestats did prove to be of additional worth later, when additional functionality was added to the extension.

2.3 NoScript

NoScript focuses on stopping Javascript from being executed, as opposed to disallowing Javascript objects from being downloaded. However, the logic used in NoScript is similar to what is needed for this project, so it is still a valuable resource. NoScript is able to analyze from where the Javascript is being executed by comparing its address to the current active site's address to determine whether or not the content is first-party or third-party. By examining the details of how NoScript accomplishes this filtering method, the team can determine how to produce a similar algorithm for this project.

2.4 Adblock

The Adblock extension was of most use particularly because it was quite similar in functionality to the extension to be written. Adblock chooses to block or allow certain content objects based on pattern-matching to certain blacklist or whitelist patterns, defined before the web browsing occurs. It uses an event listener to determine when an object request is being made, then compares the URL for that request to the black- and whitelist patterns, stored in the form of regular expressions. While this project is focusing on a more dynamic, per-page set of patterns to block or allow, the same general principles apply, and so Adblock is a good source of knowledge on how to proceed in writing our extension.

2.5 Adblock Plus

Some time after researching Adblock, we also discovered that an enhancement to the basic extension was also being offered, known as Adblock Plus, specifically Adblock Plus 0.7 (Palant, 2007). This extension has all of the key functionality available in Adblock, but with a number of additional features as well. Of particular interest are its abilities to employ whitelisting techniques to always allow certain sites, and to employ blocking rules to a site specifically, rather than globally. In addition, Adblock Plus features a menu system that aided us in creating our user interface.

2.6 Summary

Studying this set of extensions provided us with both a toolbox of knowledge about Firefox extensions and a starting point to work from when writing our extension. We now had

access to vital information about how extensions were written, installed and operated, and had multiple examples of extensions which work in the same realm as our project does. With this phase of preparation under our belts, we moved on to the designing and creation of our actual extension.

3. Design

There were a number of different directions we could have taken our extension. Before we decided on extending Adblock, we looked at several different options. Firstly, we could write an entire extension from scratch. This way it could do exactly what we wanted to do, and all the functionality would have been designed and implemented by us. However, the learning curve for such an endeavor seemed unrealistic given our time-frame. Also, we felt that if the resources are already out there for us to us freely we should take advantage of them. So we decided to simply add onto an already existing extension. We looked at the three extensions Pagestats, NoScript, and Adblock. To extend Pagestats to do what we hoped would be almost a full implementation of a new extension, so this option was discarded quickly. NoScript, as described in the previous section, is an extension that blocks Javascript execution using whitelisting rules. Ways of extending this extension would include making it block Javascript objects as opposed to Javascript execution, and once these objects were blocked, it would be simple to block other types of objects as well. Also, it would be useful to add some blacklisting functionality. Thirdly, Adblock is an extension that already blocks all objects aside from simply Javascript, ads or any other type of content being downloaded into a browser.

Based on these choices, we chose to extend Adblock Plus 0.7, an alternate version of Adblock, into an application that takes the security and privacy of users to even further and more customizable extents. Mainly, our extension is a whitelisting program that blocks content from being downloaded onto a user's computer. The basic options are to select from which types of sites to allow content: first party sites only, first and second, or first, second and third party sites. Secondary to this basic choice a user can also whitelist or blacklist specific sites. At this point, it is up to the user's discretion to explicitly allow their browser to download objects from sites they

feel necessary. Exact details on how this works will be explained later in Section 5, System in Action. Through these choices a user can customize our extension to meet their particular Internet browsing experience.

3.1 Summary

We chose this approach because it seemed like the most obvious extension of current Internet privacy and security options available to users. Adblock is a good tool for blocking ads and other objects, but the way a user must go about it can be awkward. In many instances, the objects would be downloaded onto the user's computer before the user explicitly blocked them. In this way, Adblock was essentially a blacklisting tool. Of course, with Adblock Plus, one could subscribe to lists of known ads and therefore have them blocked, creating a small whitelisting feature. However, our extension turns Adblock into a fully user-defined whitelisting tool, where almost all content must be explicitly *allowed* rather than *blocked* by the user. In this way, our extension is safer as extraneous content will not be downloaded unless the user allows it to happen.

Of the three extensions we looked at, Pagestats was the least useful and practically un-extensible for the kind of tool we planned to make. NoScript was a little more useful, but as it only stopped execution of code rather than blocking content, the best choice for the extension was Adblock. In particular, we chose Adblock Plus 0.7 to extend into a fully user-defined whitelisting tool. The details of how we implemented this extension will be described in the next section.

4. Implementation

This section discusses the implementation of the final design settled on in the previous section. An overview of Firefox extensions and services precede a discussion of implementation details for Adblock, Adblock Plus and the modifications made to it.

4.1 Firefox Extensions

The core functionality of all extensions is written in Javascript, a client side scripting language used primarily in web pages to provide dynamic content. A scripting language is a programming language that is not compiled, but interpreted by the program it is written for.

All graphical user interfaces (GUIs) are written in XUL, a cross-platform, XML compliant markup language written for designing user interfaces. The backend to these GUIs is written in Javascript.

4.2 Javascript Services

While some extensions create functionality from the bottom up, other extensions seek to improve existing Firefox operations. This improvement is done using the Javascript services provided by Firefox. These services tap into normal Firefox operations and either manipulate, augment or replace these operations, depending on the application. The two services crucial to the success of this project are defined below.

The first service, known as *Components.interfaces.nsIContentPolicy*, is a service used to monitor downloaded objects. Any object referenced in the HTML that must be downloaded is first passed through this service. Extensions that have requested the use of this service are

notified of each download. Specific information about the object, such as the requesting page or where the object is being downloaded from, is also provided. A decision is then made by the extension using the interface and a vote of yes or no is cast. This vote determines whether the object will be downloaded or not. It is here that Adblock and Adblock Plus do their blocking.

The second service, known as *preferences-service*, is a service useful for accessing and modifying Firefox preferences. Firefox stores all user and extension preferences in either a string, an integer or a boolean. Each preference has a name associated with it, typically starting with the name of the extension, to eliminate conflicting naming schemes. It is using this service that our extension stores all relevant user data and preferences.

4.3 How Adblock and Adblock Plus Work

Adblock and Adblock Plus both use a series of rules to determine which objects to block and which objects to keep. These rules are maintained as a series of regular expressions that are matched against every object brought through the *nsIContentPolicy* service. If the regular expression is a blacklisting expression, any match will be rejected. If the expression is a white-listing expression, any match will be accepted. Figure 4-1 shows the basic flow of control through Adblock and Adblock Plus:

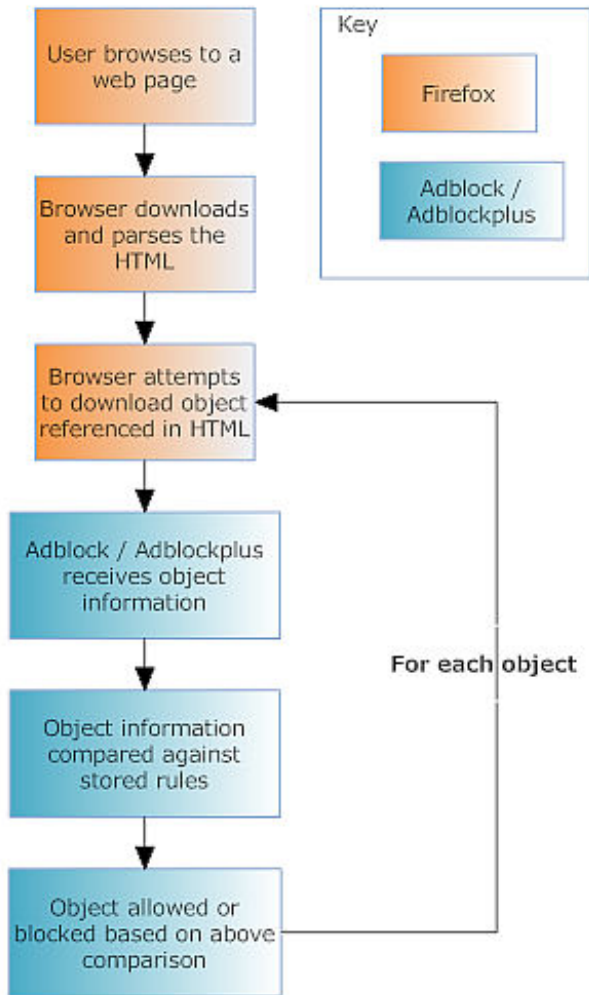


Figure 4-1 Adblock / Adblock Plus Program Flow

The user browses to a webpage, causing Firefox to download the appropriate HTML document from the web server. This document is parsed and the necessary objects referenced in the HTML (images, Javascript, CSS) are prepared for download. This is where Adblock and Adblock Plus step in. They receive relevant information about the object, such as where it is coming from and the page that requested it. The regular expressions mentioned above are run against the object, looking for a match. If a match is encountered, the appropriate action is taken. If no match is found, the object is allowed through.

4.3.1 Adblock Structure

Adblock is a simple extension with few options. The blacklisting/whitelisting rules are kept in a comma separated list managed by the *preference-service*. The source code, while uncommented, is not difficult to work through due to the forced structure the *nsIContentPolicy* interface gives the extension. The simple nature of the extension means there is less content to sift through when first learning the inner workings of Adblock.

4.3.2 Adblock Plus 0.7 Structure

Adblock Plus 0.7 is a more complicated, totally rewritten version of Adblock. The rewrite was performed in order to make some major changes to the way Adblock was organized, such as to allow rules to be downloaded from the internet. Due to the nature of the rule download service, the rules are stored in a file as opposed to the *preference-manager*. The added complexity increases the number of files and the amount of code that must be understood. However, since Adblock Plus is still using the *nsIContentPolicy* interface to block objects, an understanding of Adblock makes understanding Adblock Plus easier.

4.4 Our Extension

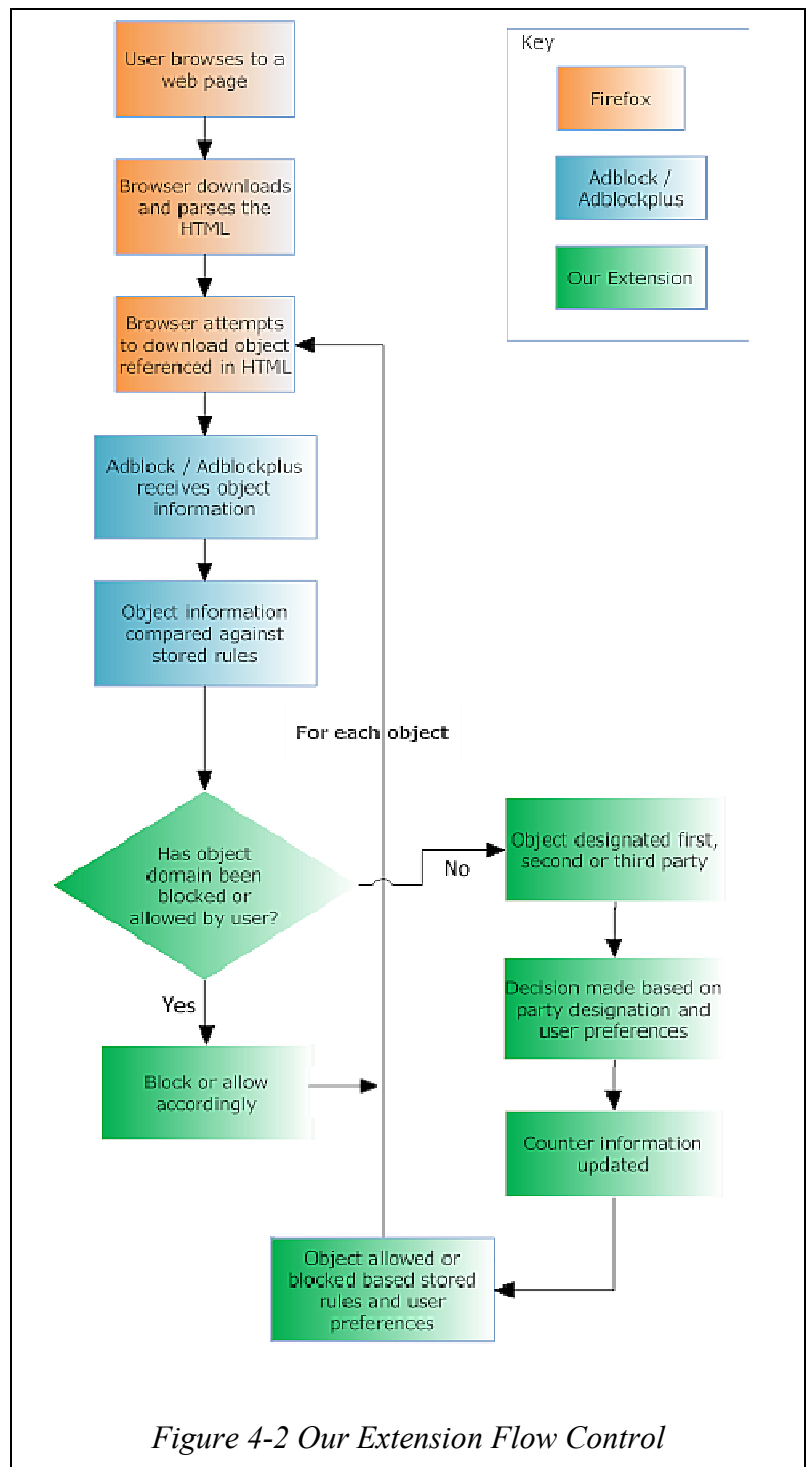
Unlike Adblock and Adblock Plus, rules are not used to determine whether an object is loaded or not. Instead, the extension blocks content based on whether it is first, second or third party content. To do this, it utilizes the information provided by the *nsIContentPolicy* service.

For content to be considered first party, it must come from the same server as the base request. To determine if this is the case, the host name of the request origin (original HTML document) and request object (image, Javascript, etc) are compared directly. If the object is a first party object, these names will match. If no match is found for first party, second party is tested next. For content to be considered second party, it must come from another server within the same domain. The request object host name from before is taken and stripped down to the primary domain and compared with the request origin host name. For example, *images.google.com* is stripped to simply *google.com*. If they match, then the object is a second

party object. At this point, if the names still do not match, then the object is a third party object.

Depending on the level the user has specified, the object is slated to be blocked or allowed. However, this decision is not final. The user also has the option of allowing or disallowing specific hosts, regardless of their classification as first, second or third party objects. If a host has been specifically allowed or disallowed by the user, the vote of yes or no is cast based entirely on the users decision. Only if the user has not made a decision on the host in question will the first, second or third party classifications be used to block the content.

After the final decision is made, a tally of hosts and the number of objects allowed through from each host is updated and made available to the user after the entire page loads. Figure 4-2 shows where in the flow of the Adblock / Adblock Plus decision making our extension



was placed:

4.5 Migration from Adblock to Adblock Plus

The first prototypes of the extension were based on Adblock. Adblock's smaller code base allowed us to develop the basic functionality for our extension without sifting through unnecessary code. The prototype was developed to the point where content could be classified and blocked based on party determination and users could select their preferences before the decision to move to Adblock Plus was made.

Migrating the code from Adblock to Adblock Plus was relatively problem free. Since our extension did not rely on any specific pieces of Adblock, the code written for Adblock Plus did not require our prototype code to be modified. Since the structure of Adblock and Adblock Plus is more or less the same, the original Adblock prototype code was easily placed in the corresponding place within Adblock Plus.

4.6 Implementation Issues

An issue arose when coding the user interface portion of the extension. The issue stemmed from the fact that the user interface and core logic pieces are separate from each other. This separation means that information known by one piece cannot not be directly shared with the other piece. This proved to be problematic when trying to get the information concerning blocked objects from the core logic to the user interface.

The most efficient solution to this problem was to use the preferences service provided by Firefox. This service is available to all branches of Firefox, regardless of whether they can

communicate with each other or not. For the core logic to communicate with the user interface, two comma separated lists are maintained. One is a listing of all hosts contacted during the page load. The other is a list of numbers corresponding to how many objects were allowed from each host. At the beginning of a page load, this information is cleared. During the page load, the information is repopulated. At the end of the page load, the information is made available to the user interface for display.

4.7 Summary

In this chapter, implementation and functionality details were discussed from a programmer's point of view. Firefox offers an efficient, cross-platform framework for developing extensions to the browser. Utilizing this framework and several of the services provided by Firefox, Adblock created an extension to block ads by utilizing rules. Attempting to improve on this design, Adblock Plus was written to provide additional functionality and a better way to manage the rules. Our extension augmented the logic used by Adblock and Adblock Plus by examining where the object comes from in relationship to where the user is currently browsing. The next section will provide information and instruction to the users of the extension.

5. System in Action

Despite the complicated inner workings of our extension, using it only requires knowledge of a few simple windows. The first window manages the level of allowed content as well as the regular expressions and general Adblockplus options. The second window is used to view the results of a page load as well as dictate which hosts are to be expressly allowed or disallowed.

5.1 Choosing the level of allowable objects

Selecting which party level may be loaded is the basic function of this extension. This choice is made by selecting one of the three highlighted radio buttons from the following window, shown in figure 5-1.

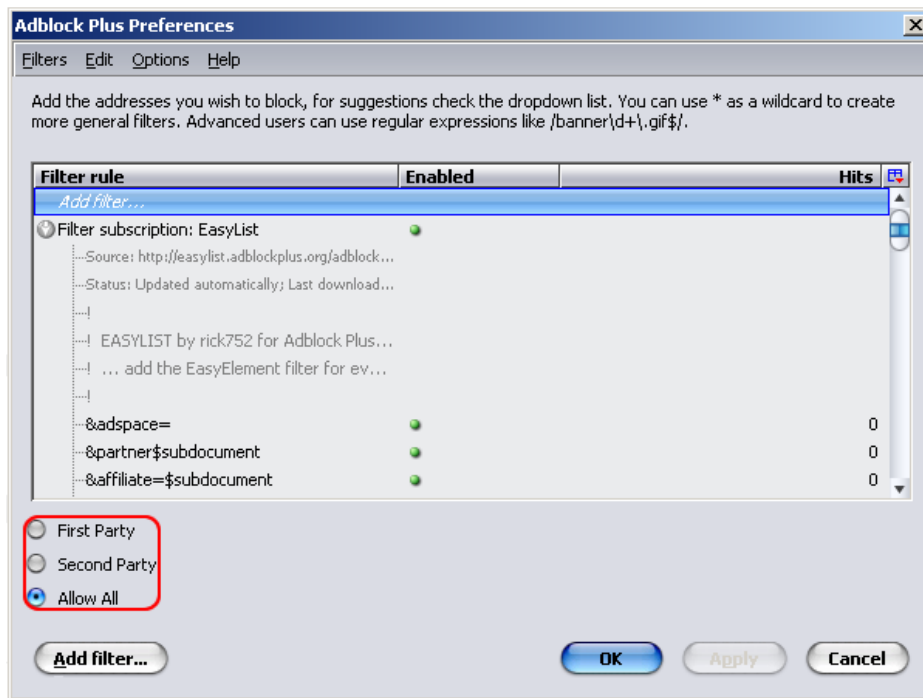



Figure 5-1 Preferences Window

Selecting “Allow All” means that all traffic will be allowed through, and no objects will

be blocked. Another selection only allows traffic at that level or lower through. For example, a selection of “Second Party” allows first and second party content through and any third party content is blocked.

5.2 Viewing the results of a page load

Once the page has loaded, the user can view the results of the load by clicking on the  icon, located in the bottom right portion of the browser. These results includes all servers whose content was allowed through as well as a counter of how many objects were loaded from each server.

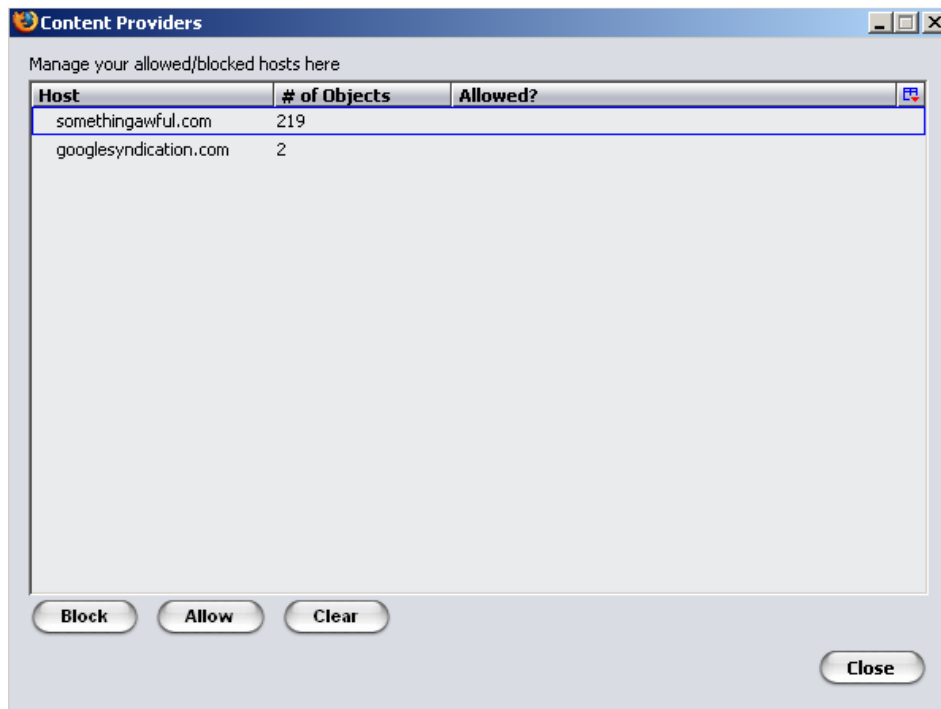


Figure 5-2 Page Load Window

Using the window shown in figure 5-2, the user can experiment with different party settings to see what objects are getting through for a particular page.

5.3 Allowing / Disallowing Specific Hosts

A problem arises when a website loads legitimate content from another website. For example, cnn.com, loads some news content from cnn.net. According to our extension, objects coming from cnn.net through cnn.com is third party. If your settings are set such that third party content is allowed, then there is no problem. However, if a user is browsing using a setting of first or second party, the desired content is blocked. Alternately, if a user browses using a setting of “Allow All”, ad content will be allowed through.

To combat this problem, the user can allow or disallow specific hosts across the board, regardless of their classification as first, second or third party content. The process for doing so is simple. Select the appropriate host from the list and click one of the action buttons highlighted in red, as shown in Figure 5-3.

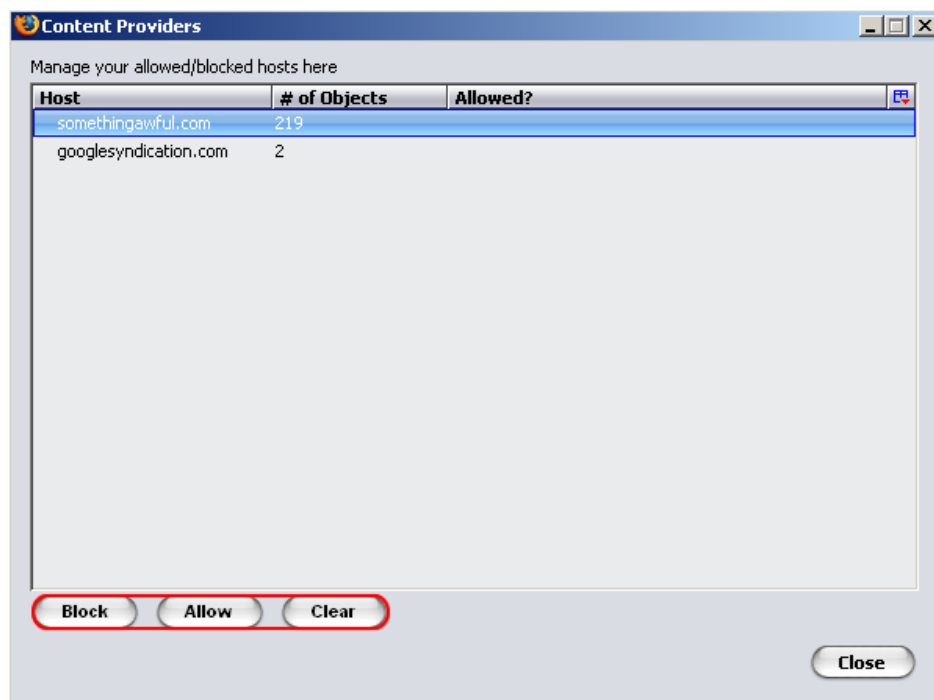


Figure 5-3 Action Buttons

Currently, neither host has a designation. After making a selection, the window indicates

the choice in the “Allowed?” column, as shown in Figure 5-4.

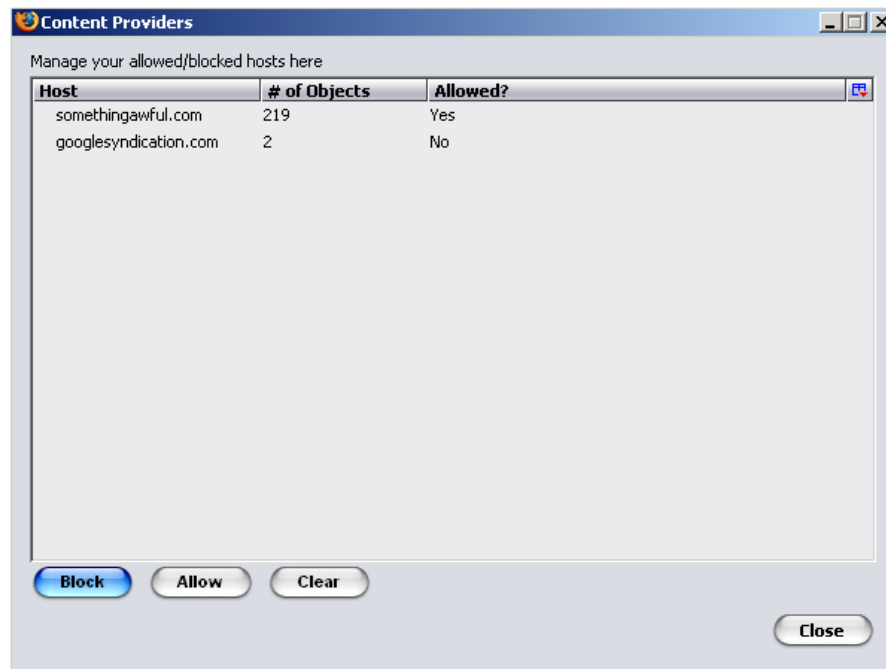


Figure 5-4 Allowing and Blocking Hosts

This choice can be overridden with the opposite choice or any designation can be cleared so that the host will be dealt with on a case by case basis.

5.4 Summary

This extension can be easily manipulated with the knowledge of only a few windows. The preferences window allows users to select their party level. The page results window not only shows the object/server counts, but also allows a user to allow or disallow a particular host. Next, we will discuss the results of testing the extension against various sites on the Internet.

6. Results

Following the creation of the extension, the next step was to determine how it affected different pages on the Internet. We compiled a list of various pages and divided them into categories: News, Forums, Web Comics, Shopping, Flash Games, Information and Search Engines. For each page we rated it based on a subjective rating system and recorded the number of objects downloaded and number of servers contacted. The analysis of this data allowed us to draw conclusions about the best setting for each category as well as general information about what percentage of downloaded objects are necessary for using the pages within a category.

6.1 Rating System

To understand the results of our analysis, it is important to understand how our rating system works. Each page visited was assigned a number 1-4 reflecting the usefulness of the page. This assignment occurred for first and second party level blocking.

A rating of 1 indicates that the website has been rendered useless. Whatever the primary function of the page was, our extension prevented that function from happening. A rating of 2 indicates that the website can be used for its primary function, but there are severe impediments that make doing so difficult. A rating of 3 indicates that the page can be used for its primary function, although there are a few abnormalities that prevent it from being perfect. A rating of 4 indicates that the website can perform its function with no problem. All pages being evaluated at Third Party are automatically given a rating of 4.

6.1.1 Rating 1 Example

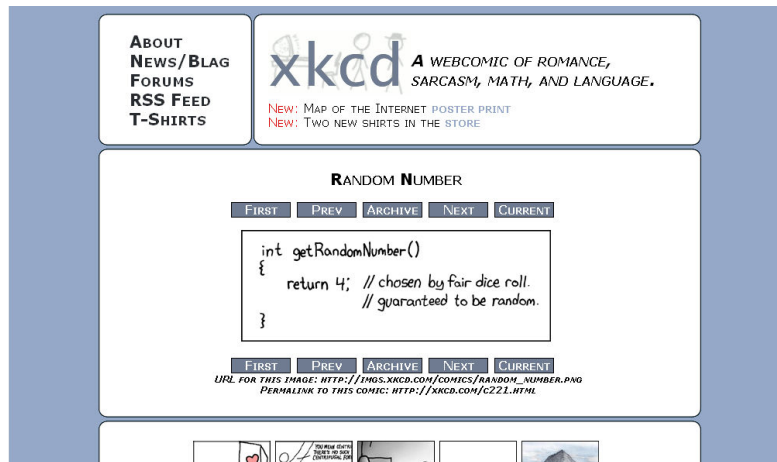


Figure 6-1 www.xkcd.com operating at “Allow All”

Figure 6-1 shows a webcomic known as xkcd. The purpose of the website is to display the image seen in the center to the user. Operating at “Allow All”, the image is clear and visible. However, if operation is switched to “First Party”, the page loads as shown below. Since the image is no longer loaded, the whole point of the website has been nullified and is given a rating of 1.



Figure 6-2 www.xkcd.com operating at “First Party”

6.1.2 Rating 2 Example

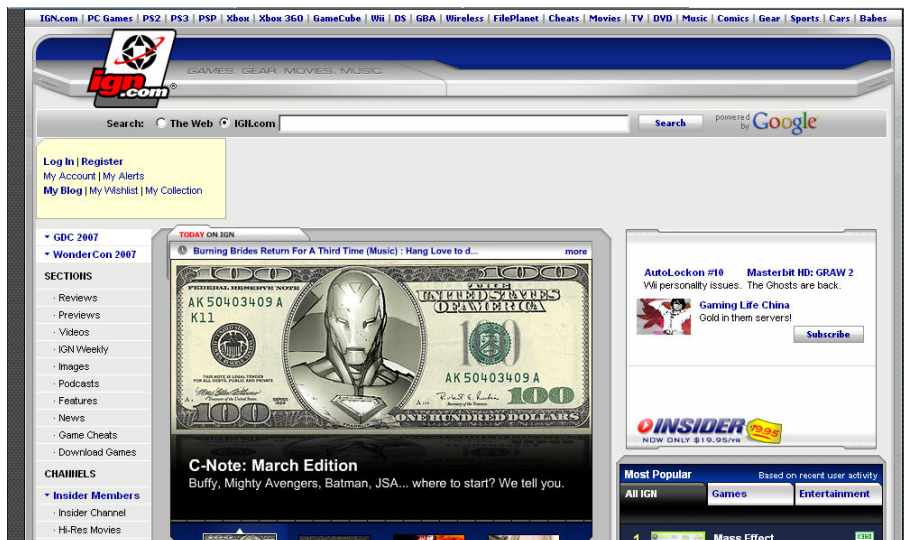


Figure 6-3 www.ign.com operating at “Allow All”

Figure 6-3 shows a gaming news site known as IGN. At a setting of “Allow All”, the page content is organized and easy to follow. At a setting of “First Party”, the content is no longer organized and it is difficult to locate a particular item, as shown in Figure 6-4. Despite the difficulties, it is still possible to read the news. Therefore, IGN operating at a setting of “First Party” receives a rating of 2.



Figure 6-4 www.ign.com operating at “First Party”

6.1.3 Rating 3 Example



Figure 6-5 <http://news.bbc.co.uk> at "Allow All"

For the British news site news.bbc.co.uk in Figure 6-5, only allowing first party content does not have as devastating an effect as it did on IGN. The above and below images are similar in layout and organization, with the bottom missing a few images. While this is far easier to browse than IGN, Figure 6-6 shows that some text is overlapping other text. It is because of this overlapping that news.bbc.co.uk receives a rating of 3 when set on "First Party".



Figure 6-6 <http://news.bbc.co.uk> at "First Party"

6.1.4 Rating 4 Example

The screenshot shows the Google News homepage with the 'Allow All' setting. The top navigation bar includes links for Web, Images, Video, News, Maps, and more. The main content area features a 'Top Stories' section with the following articles:

- Survivors say Indonesia jet shook before landing** (CTV.ca - 1 hour ago)
- White House Dismisses Rumors of Pardon for Libby** (NPR - 2 hours ago)

The right sidebar contains a 'Personalize this page' section with various news snippets and an 'In The News' list.

Figure 6-7 <http://news.google.com> at “Allow All”

A quick comparison of the images in Figures 6-7 and 6-8 reveals no significant differences. Like the other examples, however, the page in Figure 6-7 is set to “Allow All” and the Figure 6-8 is set to “First Party”. Due to the fact that there is no discernable difference between the two, a setting of “First Party” nets Google News a rating of 4.

The screenshot shows the Google News homepage with the 'First Party' setting. The layout is identical to Figure 6-7, but the top stories and sidebar content are different:

- Sabotage Unlikely In Indonesian Plane Crash** (All Headline News - 1 hour ago)
- Bush's Chavez Challenge** (The National Interest Online - 3 hours ago)

The right sidebar also shows different news snippets and an 'In The News' list.

Figure 6-8 <http://news.google.com> at “First Party”

6.2 Analysis

The analysis for each section follows a template, making comparisons between each section easier. In each section, the number of first and second party objects downloaded are divided by the number of third party objects downloaded. This division gives us what percent of the total objects were downloaded at a particular setting. Taking this percentage and plotting it against our subjective rating show what party setting is needed to achieve a usable web experience with the least amount of content downloaded. In this graph setup, the optimal results would be pages with high ratings but low downloaded object counts. This result indicates that the page can be successfully viewed without downloading excessive unwanted content.

6.2.1 News sites

News sites are highly trafficked websites, usually containing both news and ad content. This content can be spread out over first second and third party servers. Images necessary for news stories can be stored in the same partied servers as some unnecessary content. This mixture makes them good choices to test our extension. Table 6-1 contains information about how many objects were downloaded for and the rating given to various news sites.

Table 6-1 News Sites Data

Site	1 st Party Objects	2 nd Party Objects	3 rd Party Objects	1 st Party Rating	2 nd Party Rating
news.yahoo.com/	0	0	100	2	2
www.cnn.com/	0	1	367	2	2
www.nytimes.com/	5	89	134	2	4
news.bbc.co.uk/	7	145	152	3	4
www.msnbc.msn.com/	14	68	83	3	4
www.boston.com	0	214	248	2	4
news.google.com/	48	48	50	3	4
theonion.com	86	87	107	4	4
www.ign.com	0	168	218	2	3

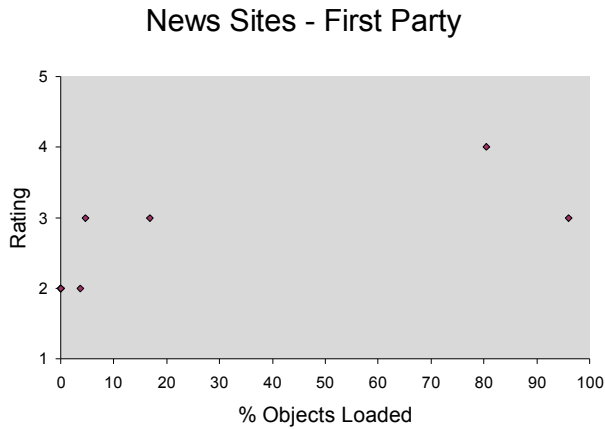


Figure 6-9 News Sites – First Party

to view successfully on first party only.

When set on “Second Party”, the ratings are generally higher and a majority of the objects were loaded (see Figure 6-10). Based on these results, allowing second party content brings the loaded object percentage and usability rating up, providing a more complete browsing experience than “First Party.”

Based on figures 6-9 and 6-10, we recommend the extension be set to “Second Party” while browsing news websites.

Although figure 6-10 does not display

optimal results, it displays high ratings at a reduced object count when compared to “Allow All” and better overall results when compared to “First Party.”

When set on “First Party”, the spread for news sites was large. As you can see in Figure 6-9, ratings varied from 2 to 4 and the percentage of objects downloaded ranged from 0 to 100. Based on these results, these news websites have a varied amount of content coming from second and third party servers and are therefore difficult

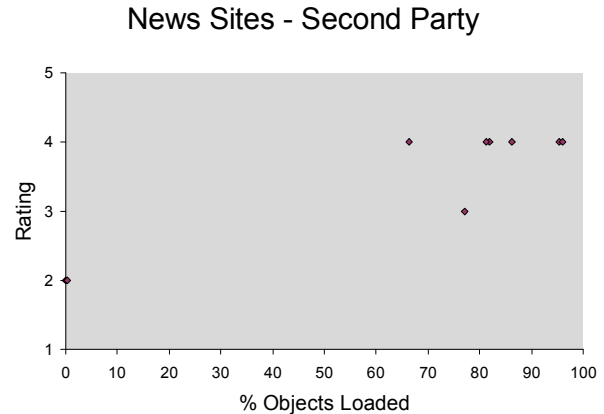


Figure 6-10 News Sites – Second Party

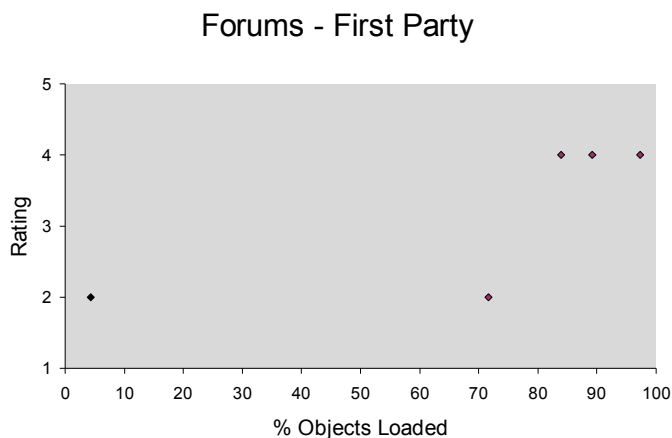
6.2.2 Forum Sites

Forum sites, to save bandwidth, force their users to host images on third party servers. Often, these same forums will use ads to supplement whatever monthly income they have. This arrangement presents a unique opportunity to examine a set of sites with third party user content as well as advertisements. Table 6-2 contains information about how many objects were downloaded for and the rating given to various forums.

Table 6-2 Forum Sites Data

Site	1 st Party Objects	2 nd Party Objects	3 rd Party Objects	1 st Party Rating	2 nd Party Rating
forums.somethingawful.com/	2	38	46	2	4
forums.tabit.net/	21	23	25	4	4
my.wpi.edu	106	104	109	4	4
www.ultimatemetal.com/forum/	429	427	481	4	4
forums.cgsociety.org/	169	223	236	2	4

Figure 6-11 indicates that some forums were unaffected while a few forums were rendered difficult to use. Most of the time, forums are only about the text, so not downloading a



few images will not matter. However, a few of the forums tested relied on the use of images for navigation, and the use of first party only ruins the experience.

Figure 6-11 Forums – First Party

Figure 6-12 shows that all forums tested were brought up to a rating of 4, the highest rating possible. The object count now ranges from 0.8 to 1, a much smaller spread than first party sites. However, the highly rated first party results did not change significantly in rating or object count when tested with second party.

Based on these results, it is our recommendation that forums be browsed with a setting of “Second Party.” However, in the case of a few forums, it may be acceptable to use “First Party” without any

detrimental effects. Finally, due to the large amount of third party member content it would be acceptable to utilize the “Allow All” when viewing certain threads.

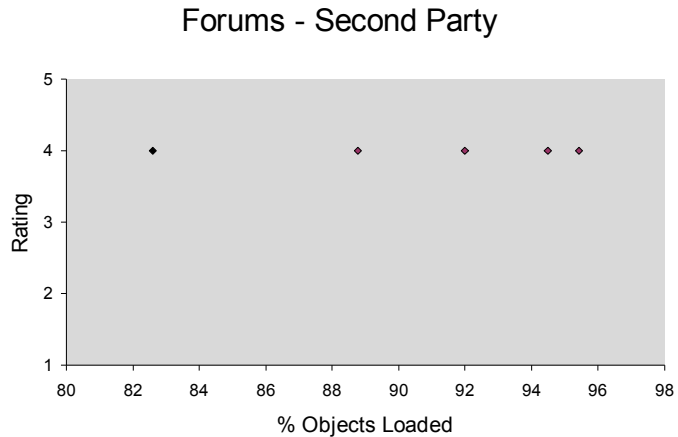


Figure 6-12 Forums – Second Party

6.2.3 Webcomics

Webcomics are an increasingly popular way for artists to get their art viewed on the Internet. Since the content of a webcomic is the image, and not the text, these pages are a good test of how our extension handles pages whose primary content is made up of images. Table 6-3 contains information about how many objects were downloaded for and the rating given to various webcomics.

Table 6-3 Webcomics Data

Site	1 st Party Objects	2 nd Party Objects	3 rd Party Objects	1 st Party Rating	2 nd Party Rating
www.ctrlaltdel-online.com/	71	71	84	4	4
www.vgcats.com/	29	28	41	4	4
www.penny-arcade.com/	27	27	41	4	4
www.questionablecontent.net/	4	17	32	4	4
xkcd.com/	0	19	22	1	4
pbfcomics.com/	16	16	19	4	4

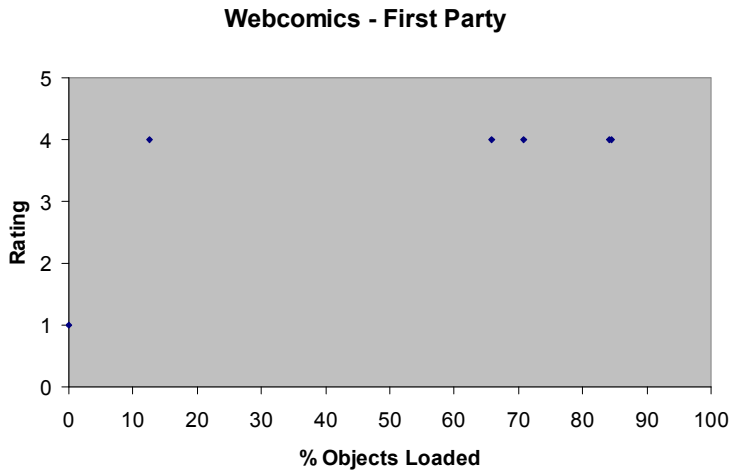


Figure 6-13 Webcomics – First Party

the actual comic. This, however, was a rare occurrence.

In Figure 6-14, most of the webcomics increased their downloaded object count and retained their usability ratings. The one comic that did not work using “First Party” was fixed by using “Second Party.”

Based on these results, we recommend that webcomics be browsed using the “First Party” setting. For the rare comics that do not work under this setting, it would be necessary to increase the allowed level to “Second Party.”

As can be seen in Figure 6-13, webcomics score high in terms of First Party usability ratings. The downloaded content, however, is fairly spread out. Only one website in particular was rendered useless by the lack of

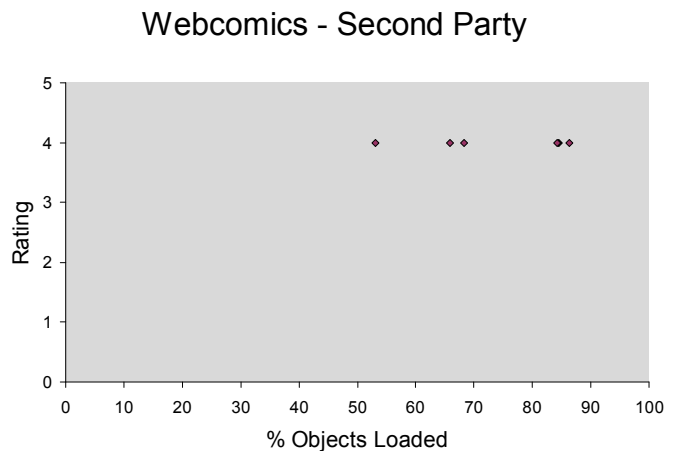


Figure 6-14 Webcomics – Second Party

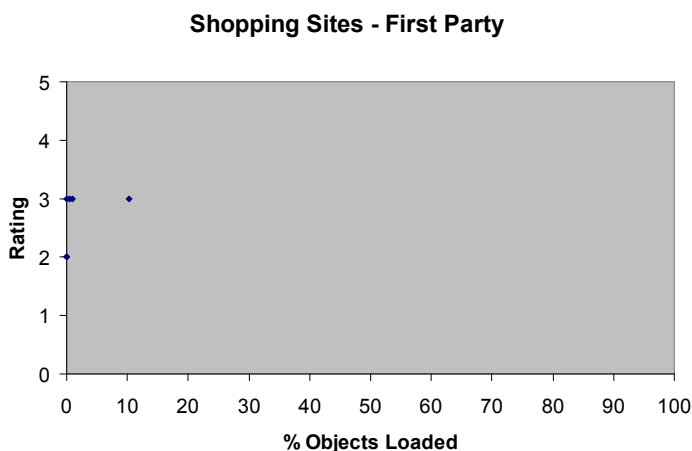
6.2.4 Shopping Websites

Shopping websites have a unique mix of image and text for their content. Both of these contents are needed to browse the website successfully. However, shopping websites can also contain unwanted content such as ads. Often times it is hard to block the ads but still allow the necessary content, for they may both be hosted on second or third party servers. Table 6-4 contains information about how many objects were downloaded for and the rating given to various shopping websites.

Table 6-4 Shopping Data

Site	1 st Party Objects	2 nd Party Objects	3 rd Party Objects	1 st Party Rating	2 nd Party Rating
http://www.amazon.com/	0	8	82	3	3
http://www.target.com/gp/homepage.html	0	2	196	2	2
http://www.ebay.com/	0	0	167	2	2
http://www.newegg.com/	1	190	193	3	4
http://www.walmart.com/	1	35	108	3	4
http://www.half.ebay.com/	32	1	312	3	3

As Figure 6-15 shows, allowing first party content only results in very low object counts. However, none of the shopping sites were rendered useless. In fact, the majority of them



received a rating of 3. These results actually fit best with what we considered optimal. Although none of the websites was fully functional with a rating of 4, these results still show that shopping websites can be useful

Figure 6 15 Shopping Sites – First Party

even with a setting of “First Party.”

Allowing second party content also, we see a jump in several of the shopping websites object counts and ratings (see Figure 6-16). The results are still very similar to the “First Party” results, however, some of the websites now receive ratings of 4 with considerably more content downloaded.

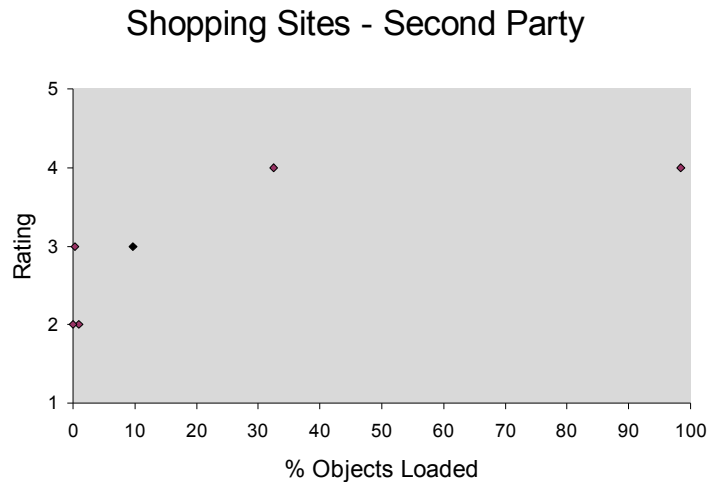


Figure 6-16 Shopping Sites – Second Party

Based on these results it can be surmised that between settings of “First Party” and “Second Party”, not much difference will be found in a website. However, this does not mean that “Second Party” is not better. We recommend that if a user is browsing a shopping website with our extension, it would usually be best to allow all content for a full browsing experience, as neither “First Party” or “Second Party” gave us any particularly excellent results. However, if a user is willing to sacrifice some of the content, they can bump it down to “Second Party” or perhaps even “First Party” depending on the particular site.

6.2.5 Flash Websites

Flash sites contain objects that are neither text nor images. However, this content should not matter for our extension, but it is a good way to test flash objects in particular. As the main goal of these websites is to provide flash content to the user, it is useful to find out whether or not the sites will work on “First Party” or “Second Party” settings. We, as users, would like to

block unwanted content, because flash-centric sites are notoriously overrun with unnecessary ads. Table 6-5 contains information about how many objects were downloaded for and the rating given to various flash websites.

Table 6-5 Flash Sites Data

Site	1 st Party Objects	2 nd Party Objects	3 rd Party Objects	1 st Party Rating	2 nd Party Rating
www.addictinggames.com	12	73	88	1	4
www.albinoblacksheep.com	88	87	91	1	1
www.freearcade.com	175	174	201	3	3
www.newgrounds.com	250	235	280	4	4

As Figure 6-17 shows, allowing first party content only causes a scattering of results.

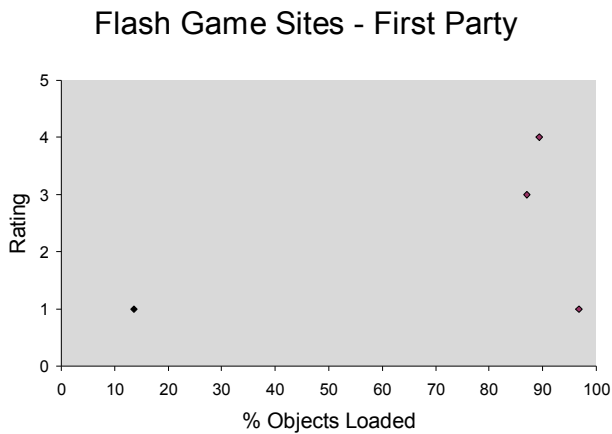


Figure 6-17 Flash Game Sites – First Party

Some websites have low ratings with high object counts, others have low ratings with low counts, and even others have high ratings with high object counts. It would be hard to recommend such a setting for anyone who visits multiple flash websites regularly.

Setting the extension to also allow second party content causes the sites to simply download more objects (see Figure 6-18). This, however, does not necessarily make the sites any easier to view, as the ratings given still range from 1 to 4. This is most likely due to most of the flash content being hosted on third party servers.

From these results, our best

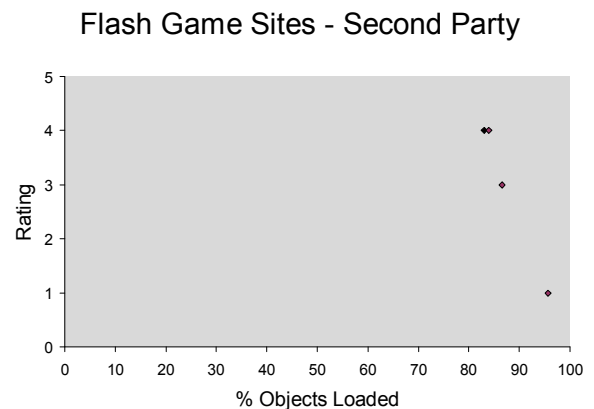


Figure 6-18 Flash Games Sites – Second Party

recommendation for users who regularly browse multiple flash sites would be to set our extension on a setting of “Allow All,” otherwise the flash content will most likely be inaccessible. For advanced users, however, in order to block the extraneous ads that regularly plague flash websites, we recommend a lower setting along with our extension’s user-defined whitelisting features.

6.2.6 Information Websites

Information websites are websites that can be used as large repository of information. They may contain encyclopedic knowledge about a specific topic or all topics, or they may be social networking websites. Many of these websites have a lot of objects being displayed that are all relevant to the topic being explored. Table 6-6 contains information about how many objects were downloaded for and the rating given to various information websites.

Table 6-6 Information Sites Data

Site	1 st Party Objects	2 nd Party Objects	3 rd Party Objects	1 st Party Rating	2 nd Party Rating
www.imdb.com	0	105	108	3	4
www.facebook.com	0	67	79	2	4
www.allmusic.com	80	95	137	3	4
www.myspace.com	0	29	41	2	3
en.wikipedia.org	17	17	35	3	3

As Figure 6-19 shows, allowing first party content only for information sites generates mostly ratings of 3. While the content percentages range between 0-1, the ratings remain mostly

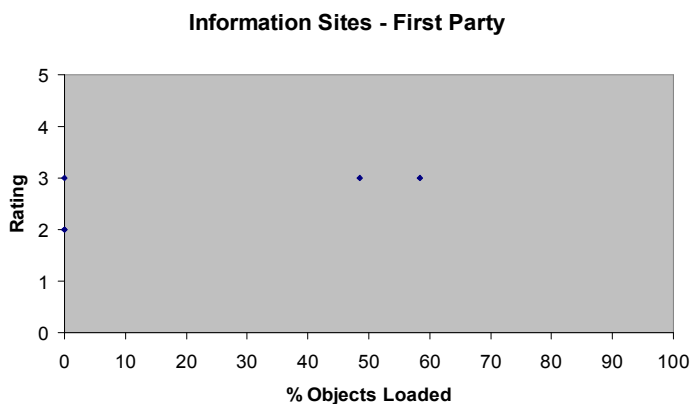


Figure 6-19 Information Sites – First Party

constant. Only a few times are 2s given.

Adding second party content on top of this eliminates all ratings of 2 and even adds some 4s (see Figure 6-20). These sites are almost fully functional without having to allow third party content. Although, this

setting does increase the object counts, in some cases almost to 100%, it is still a better outcome than the “First Party” setting. It is doubtful that second party servers would contain any unwanted content.

Based on these results, we would recommend that users who regularly browse information sites and use our extension use a setting of “Second Party.” This will allow for all the needed content to be downloaded while keeping out third party content such as ads.



Figure 6-20 Information Sites – Second Party

6.2.7 Search Engines

Search engines tend to gather all of their content from other websites, usually third party servers. However, this content is often processed into first or second party objects. In other cases, all the content is hosted by the search engine, such as youtube.com. Table 6-7 contains information about how many objects were downloaded for and the rating given to various search engines.

Table 6-7 Search Sites Data

Site	1 st Party Objects	2 nd Party Objects	3 rd Party Objects	1 st Party Rating	2 nd Party Rating
www.google.com	3	3	6	4	4
images.google.com	25	24	28	4	4
www.youtube.com	86	117	113	3	4

As Figure 6-21 shows, allowing only first party content, our browsing experience is hardly hindered. Mostly ratings of 4 were given, and 50% or more of the content was

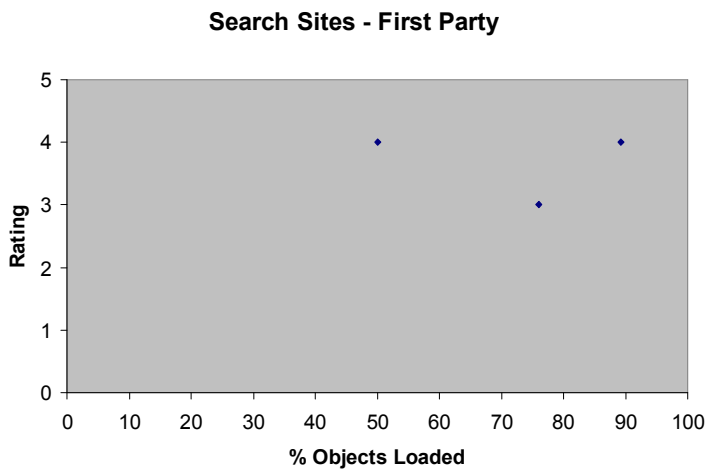


Figure 6-21 Search Sites – First Party

downloaded. Usually, almost all of the content was hosted on first-party servers.

When adding second party content to the allowed objects, the search engines became almost perfectly rendered, with all ratings of 4 and practically 100% of the content downloaded (see Figure 6-22). However, this does not necessarily make the “Second Party” setting better than the “First Party” setting, as the latter setting still produced ratings of mostly 4.

While we would recommend a setting of “First Party” for users using mainly search engines, this recommendation is rather frivolous, as search engines are used primarily as a gateway to many other sites. Only in the case of a site like youtube.com could we make such a recommendation, as all its content is hosted on its own servers.

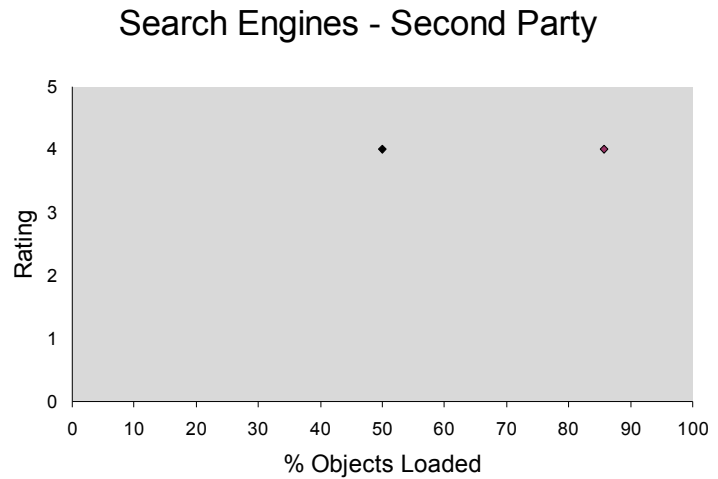


Figure 6-22 Search Sites – Second Party

6.3 User Surveys

Along with our own testing, we allowed our extension to be used by several outside users and give us feedback. Most of this feedback was positive, and many users said they would continue to use our extension. There were a few small complaints people had with the extension. One user suggested that we should replace all of the normally visible blocked content with an indicator that some object used to be there. However, this feature was not able to be implemented in our short development cycle and is something to be looked at for future releases of the extension. Other complaints simply stemmed from users not fully understanding what our extension was meant to do, and could have been caused by users not reading the readme file contained with the extension.

6.4 Summary

Based on all our findings, the best overall setting we would recommend to users of our extension would be “Second Party.” As second party servers are still affiliated with the first party host, it is highly doubtful that these servers will contain unwanted or malicious content. However, in a select few cases third party content may be necessary to download in order to view a webpage correctly, such as Flash websites, for oftentimes the content is not contained on first or second party servers. In many cases, allowing only first party content was also sufficient, however the rating usually suffered by at least a point, and the websites where it didn’t were few and far between. As first party content is contained within the “Second Party” setting, this setting still remains our best recommendation.

7. Conclusions

This project set out to find a potential solution for the wide range of undesirable content and privacy concerns experienced when browsing the Web. Our attempt involved writing a Firefox extension that filters content based on server location, differentiating between content offered from the site a user intended to visit and content provided by sites from other domains. In doing so, we hoped that extraneous content could be avoided while maintaining the usability of most sites.

What we found was that the solution is not that simple. Second-party and third-party content is often essential for a website to display properly, sometimes even to perform its primary function. Blindly filtering out this content causes too many issues for it to be worth applying in such a general sense.

On the other hand, there were sets of websites that were able to perform just as well as usual with most or all of the extraneous content removed. This supports the idea that domain-based content filtering does have a use in web browsing for certain sites.

7.1 Future Work

A number of improvements can be made to the extension to improve its usability. For one, we have determined that there is no one setting for level of content control that works well with all websites, so a more dynamic approach must be taken. Similar to the existing white and black lists, one could maintain a list of websites and their preferred filter setting: “First Party,” “Second Party” or “Allow All.” Some simple interface modifications to allow users to change these settings for a website would make the extension more usable.

In addition, Adblock Plus supports filter rules which a user can subscribe to. These filter

rules are automatically modified as updates are made at some central location. If we establish a similar list of our extensions preferences, such as typical party settings and allowable hosts for commonly used web sites, this list could be made available to the public. This list would allow users to get up and running with the extension very quickly, as well as providing a positive example of how our extension should be used.

Finally, one of the user surveys recommended a visual cue indicating what content had been blocked and where, such as a placeholder image for an ad that had been blocked. While not directly related to the issue of privacy and content control, this is something that would enhance the user experience and should be investigated later, time permitting.

Works Cited

Adblock, 2007. The Adblock Project. New York, NY: Mozdev Community Organization, Incorporated. On-line. Available from Internet, <http://adblock.mozdev.org/>, accessed 2006-2007.

Dedeo, Scot, 2006. Pagestats Extension. Worcester, MA: Worcester Polytechnic Institute. On-line. Available from Internet, <http://web.cs.wpi.edu/~cew/pagestats/>, accessed 2006-2007.

NoScript, 2007. NoScript. Palermo, Italy: InformAction. On-line. Available from Internet, <http://noscript.net/>, accessed 2006-2007.

Palant, Wladimir, 2007. Adblock Plus: Save your time and traffic. On-line. Available from Internet, <http://adblockplus.org/en/>, accessed 2006-2007.

Appendices

Appendix A: User Surveys

Survey 1

1. Which features of the extension did you like, and why?

It blocked ads which is always good.

2. Which features of the extension did you dislike, and why?

When set to 2nd party, it blocked most ads but it blocked a lot of images that I wanted to see.

3. Did you encounter any websites which were difficult or impossible to use while running our extension? If so, please list the sites and what the problems were.

www.amazon.com had all its images blocked while set to 2nd party.

4. Would you continue using this extension in the future? If not, what changes would make you consider using it?

Regular adblock works fine for me.

Survey 2

1. Which features of the extension did you like, and why?

I liked the fact that you could choose to allow only first or second party content because third party content is often irrelevant.

2. Which features of the extension did you dislike, and why?

None.

3. Did you encounter any websites which were difficult or impossible to use while running our extension? If so, please list the sites and what the problems were.

Nope.

4. Would you continue using this extension in the future? If not, what changes would make you consider using it?

Probably. It would be nice to have a feature which block's all third party sites except image hosting sites like imageshack.

Survey 3

1. Which features of the extension did you like, and why?

I like that it is one click away on the bottom of the screen so that I can easily add/remove hosts.

2. Which features of the extension did you dislike, and why?

Can't really tell when/where things are blocked. It'd be nice to have maybe some small text show up to replace the ads.

3. Did you encounter any websites which were difficult or impossible to use while running our extension? If so, please list the sites and what the problems were.

I didn't find any, and if there were it would be simple to add the host to

the allowed websites list.

4. Would you continue using this extension in the future? If not, what changes would make you consider using it?

Yes I'm going to continue using it, there really isn't any reason not to, it hasn't caused any problems and is a lot easier to use than updating my hosts file all the time.

Appendix B – Data Gathered

First Party

	Number of Objects	Number of Servers	Rating
News			
news.yahoo.com/	0	0	2
www.cnn.com/	0	0	2
http://www.nytimes.com/	5	1	2
news.bbc.co.uk/	7	1	3
www.msnbc.msn.com/	14	1	3
www.boston.com	0	0	2
news.google.com/	48	1	3
theonion.com	86	1	4
www.ign.com	0	0	2
Forums			
http://forums.somethingawful.com/	2	1	2
http://forums.tabit.net/	21	1	4
my.wpi.edu	106	1	4
http://www.ultimatemetal.com/forum/	429	1	4
http://forums.cgsociety.org/	169	1	2
Webcomics			
http://www.ctrlaltdel-online.com/	71	1	4
http://www.vgcats.com/	29	1	4
http://www.penny-arcade.com/	27	1	4
http://www.questionablecontent.net/	4	1	4
http://xkcd.com/	0	0	1
http://pbfcomics.com/	16	1	4
Shopping			
http://www.amazon.com/	0	0	3
http://www.target.com/gp/homepage.html	0	0	2
http://www.ebay.com/	0	0	2
http://www.newegg.com/	1	1	3
http://www.walmart.com/	1	1	3
http://www.half.ebay.com/	32	1	3
Flash Games			
www.addictinggames.com	12	1	1
www.albinoblacksheep.com	88	1	1
www.freearcade.com	175	1	3
www.newgrounds.com	250	1	4
Information			
IMDB	0	0	3
www.facebook.com	0	0	2
allmusic.com	80	1	3
myspace.com	0	0	2
en.wikipedia.org	17	1	3
Seach Engines			
www.google.com	3	1	4
images.google.com	25	1	4
www.youtube.com	86	1	3

Second Party

	Number of Objects	Number of Servers	Rating
News			
news.yahoo.com/	0	0	2
www.cnn.com/	1	1	2
http://www.nytimes.com/	89	1	4
news.bbc.co.uk/	145	1	4
www.msnbc.msn.com/	68	1	4
www.boston.com	214	1	4
news.google.com/	48	1	4
theonion.com	87	1	4
www.ign.com	168	1	3
Forums			
http://forums.somethingawful.com/	38	1	4
http://forums.tabit.net/	23	1	4
my.wpi.edu	104	1	4
http://www.ultimatemetal.com/forum/	427	1	4
http://forums.cgsociety.org/	223	1	4
Webcomics			
http://www.ctrlaltdel-online.com/	71	1	4
http://www.vgcats.com/	28	1	4
http://www.penny-arcade.com/	27	1	4
http://www.questionablecontent.net/	17	1	4
http://xkcd.com/	19	1	4
http://pbfcomics.com/	16	1	4
Shopping			
http://www.amazon.com/	8	1	3
http://www.target.com/gp/homepage.html	2	1	2
http://www.ebay.com/	0	0	2
http://www.newegg.com/	190	1	4
http://www.walmart.com/	35	1	4
http://www.half.ebay.com/	1	1	3
Flash Games			
www.addictinggames.com	73	1	4
www.albinoblacksheep.com	87	1	1
www.freearcade.com	174	1	3
www.newgrounds.com	235	1	4
Information			
IMDB	105	1	4
www.facebook.com	67	1	4
allmusic.com	95	1	4
myspace.com	29	1	3
en.wikipedia.org	17	1	3
Search Engines			
www.google.com	3	1	4
images.google.com	24	1	4
www.youtube.com	117	1	4

Third Party

	Number of Objects	Number of Server	Rating
News			
news.yahoo.com/	100	4	4
www.cnn.com/	367	6	4
http://www.nytimes.com/	134	8	4
news.bbc.co.uk/	152	3	4
www.msnbc.msn.com/	83	4	4
www.boston.com	248	9	4
news.google.com/	50	1	4
theonion.com	107	6	4
www.ign.com	218	5	4
Forums			
http://forums.somethingawful.com/	46	3	4
http://forums.tabit.net/	25	1	4
my.wpi.edu	109	1	4
http://www.ultimatemetal.com/forum/	481	6	4
http://forums.cgsociety.org/	236	2	4
Webcomics			
http://www.ctrlaltdel-online.com/	84	4	4
http://www.vgcats.com/	41	3	4
http://www.penny-arcade.com/	41	4	4
http://www.questionablecontent.net/	32	5	4
http://xkcd.com/	22	1	4
http://pbfcomics.com/	19	1	4
Shopping			
http://www.amazon.com/	82	2	4
http://www.target.com/gp/homepage.html	196	2	4
http://www.ebay.com/	167	4	4
http://www.newegg.com/	193	4	4
http://www.walmart.com/	108	5	4
http://www.half.ebay.com/	312	4	4
Flash Games			
www.addictinggames.com	88	5	4
www.albinoblacksheep.com	91	1	4
www.freearcade.com	201	8	4
www.newgrounds.com	280	11	4
Information			
IMDB	108	3	4
www.facebook.com	79	2	4
allmusic.com	137	9	4
myspace.com	41	3	4
en.wikipedia.org	35	2	4
Seach Engines			
www.google.com	6	1	4
images.google.com	28	1	4
www.youtube.com	113	3	4