

INTRODUCTION

Welcome to SurfReport

SurfReport delivers statistics about your web site traffic, customized to fit your needs.

In addition to giving you the flexibility to generate reports at any time, SurfReport offers the ability to filter information and create multiple charts. SurfReport's results can be displayed on the web or exported by email for easy integration into spreadsheet or database applications.

SurfReport measures the following statistics for the period contained in the log files:

- total number of unique visitors
- total number of visits, including multiple visits by unique visitors
- total number of page hits - offers the option of viewing the results by URL or by the file's title
- total number of hits
- day-by-day breakdown of visits, page hits, and total hits
- average number of pages hit by each visitor
- average number of visitors accessing your site each day
- average number of page hits per day
- average number of total hits per day
- average number of total hits per hour



Introduction

- average percentage of total hits per hour
- the most frequent “x” visitors who have accessed “your site” with at least “y” hits *
- the most frequent “x” domains accessing “your site”
- the most frequent “x” countries accessing “your site”
- the most frequent “x” referrer pages to “your site”
- the most frequent “x” files that were accessed at least “y” times on “your site”
- all visitors who have accessed particular file(s) containing keyword(s)
- all visitors whose name matches a particular substring or keyword
- a list and hit count of visitors excluded from the report (see filters)
- a list and hit count of files excluded from the report (see filters)
- a list of error codes listed by name and frequency
- a summary of the log files found by SurfReport and used to generate the current report; including name, number of hits, number of excluded hits and number of errors for each log file.

SurfReport's filters can:

- exclude all visitors matching particular keyword(s) or substring(s)
- exclude all visitors except those matching particular keyword(s) or substring(s)
- exclude all files matching particular keyword(s) or substring(s)
- exclude all files except those matching particular keyword(s) or substring(s)

* Both “x,” “y,” and “your site” are variables that you can set.



Introduction

A Simple Roadmap to Logs, Formats, Analyzers, and Pitfalls

SurfReport is a server-based analysis application designed to be run with a Web browser. All execution takes place on the server and results are passed back to the browser in HTML and in easy-to-read charts. Since it runs on the server where the logs reside, the total elapsed time to obtain an analysis is anywhere from 2 to 100 seconds on average. By combining a unique hashing algorithm in C, and by eliminating download times and import processes, SR is the fastest commercial product on the market for obtaining actionable reports and graphs on Web traffic.

Once installed on the server, there are no further complications in obtaining a real time analysis from any work station with a Web browser. SurfReport eliminates the inherent cost in dedicating a client work station and the waste associated with file transfer time to the work station from the server.

A common problem people have is the installation of server-resident software. SurfReport is no exception. To install SurfReport, you need to understand the ins and outs of your Web Server (Apache, Netscape, et al), operating system, and directory tree to your logs and document root. Follow the installation procedure carefully when you install SurfReport. For your convenience, Bien Logic offers installation included in the cost of the single domain license.

SurfReport analyzes the access log in NCSA format referred to as the common log format. The access log in this format is the only standardized log. All others have meta defin-



Introduction

itions or are proprietary in nature specific to an HTTP server. The NCSA common log format contains one record per line. All HTTP servers can produce one. Here is an example record:

```
p25.ts1.walrus.com - - [22/Sep/1996:11:06:26 -0700] "GET /cgi-bin/surfreport.cgi/check/surfoutput/5297 HTTP/1.0" 302 -
```

An extension to the common log format supported by SurfReport is the combined log format where the record contains referrer and browser information:

```
shawano-51.dialup.frontiercomm.net - - [30/Dec/1996:02:43:17 -0800] "GET /index.cgi HTTP/1.0" 200 - "http://www.flirt.com/forum1.cgi?view=comments" "Mozilla/2.01 (Win95; I; 16bit)"
```

SurfReport will automatically detect the combined log format as well and provides statistical analysis on referring pages and browsers. Bien Logic wrote and published an article for Dr. Dobbs Sourcebook, September/October 1996 entitled, "Analyzing Web Traffic with Perl." Consult this article if you are interested in a more detailed discussion of this topic. (<http://www.ddj.com>)

Let's talk briefly about the technical architecture. SurfReport combines the efficiency of C and the flexibility of Perl (Practical Extraction and Reporting Language). Essentially, Perl is a very powerful yet user-friendly language for formatting reports. SurfReport uses this language to format reports and to generate the user interface rendered by a Web browser. For crunching the data, it uses C which has proven to be lightning fast. C is also used



Introduction

for generating charts which is another reason the program is so fast. If you are a Perl programmer, you can tweak SurfReport for minor reformatting as you see fit provided that you propagate your changes across releases of SurfReport.

SurfReport is principally an application for marketing and trend analysis. Through ranking and charting, decision-making becomes instantaneous instead of a laborious endeavor combing through numbing mounds of tabular reports. Ranking, as opposed to sorting, is the key. Charting using pie and bar charts enhance cognition. Also, by combining this week's log with last month's log for the same period, it is possible to perform interesting trend analysis. SurfReport can analyze an unlimited number of distinct logs in a single run! Finally, for advanced users, SurfReport may be run as a background job under the UNIX cron facility. This makes it possible to automate the entire production process, including e-mail delivery of HTML and tab-delimited reports.

SurfReport is available for NT servers and many versions of UNIX and http servers. We are constantly updating SurfReport through engineering changes, platform testing, and, yes, bug fixing. Our mission is to provide you with the best analysis tool available. Enjoy using SurfReport!



Introduction

TABLE OF CONTENTS

Introduction	i
• Welcome to SurfReport	i
• A Simple Roadmap to Logs, Formats, Analyzers, and Pitfalls	iii
Part 1: Installing SurfReport	1
• Initial Considerations	1
• UNIX Installation	1
Check List Before You Install	1
Uncompressing <code>surfreport.tar</code>	3
Running the Installer	4
Using your WEB browser to complete the installation	6
Configure SurfReport's Form Interface	9
Other Sources of Documentation	10
Removing SurfReport	10
• Windows NT Installation	11
Part 2: SurfReport's Settings	13
• Deliver my SurfReport	13
• Name my SurfReport	14
• Daily Totals	14
• Period Totals	15
• Daily Averages	15
• Hourly Averages	15



Introduction

• The Most Frequent Visitors That Have Accessed Your Site	16
• The Most Frequent Companies/Domains That Have Accessed Your Site	17
• The Most Frequent Countries That Have Accessed Your Site	17
• The Most Frequent Referrer Pages ToYour Site	18
• The Most Frequent Files In Your Site That Have Been Accessed	18
• Specific Visitor Report	19
• Visitor Filter	19
• Specific File Report	19
• File Filter	20
• Browsers	20
• Operating Systems	20
• Summary of HTTP Errors	21
• Process Files For	21
• Access Log Files	21
• Password	21
Part 3: Generate a SurfReport	22
• Using Your Default Settings	22
• In-Depth Analysis	23
Part 4: Advanced Tasks	24
• Overall Architectural Flow	24
• Using Visitor and File Filters	26
• Logging Only Page Hits to Your Site	28
• User Browser and Platform Preferences	29



Introduction

• Sorting by Country or Organization	30
• Running Reports on an Unattended Basis	32
• Using SurfReport with Many Virtual Hosts	33
Part 5: Charts	34
• Basic Chart Options	35
Chart Type	35
Orientation	36
Dimensions	36
Colors	38
• Advanced Chart Options	39
Titles and Labels	39
Axes	40
Part 6: Glossary of Terms and Definitions	41
Appendix A: Summary of HTTP Errors	54
Appendix B: Country Codes	56
Index	62



PART 1: INSTALLING SURFREPORT

Initial Considerations

SurfReport uses a mixture of Perl and C code and works on nearly all UNIX and NT servers. For more information regarding your particular host computer, see the accompanying text documents that come with your SurfReport files.

UNIX INSTALLATION

Checklist Before You Install

Please take a couple of minutes to review the following information. Checking this information now will help make SurfReport's installation run smoothly.

- 1) The file **surfreport.tar** should be in your HTTP server root directory (where your html documents reside). The root directory is the top level for your html files and in it you should find a file named index.html (this file, commonly called the "home page," is the page seen when you type http://www.company_name.com in the location line of your web browser). Move **surfreport.tar** to this level if it is not already there.



Part 1: Installing SurfReport

2) Some permissions must be set prior to installing SurfReport.

Files: The log files should have permissions set to 755

Directories: The cgi folder (cgi-bin) must have permission to run cgi scripts.

Note: In most cases, the permissions for these files and directories are already set to the correct permissions.

3) You will need to use a web browser to complete the final steps of SurfReport's installation.

4) You need to know the location of the following folders and paths:

- a) CGI path/folder
- b) HTTP server root directory - where your html documents reside
- c) Perl Path
- d) Access log Folder - where your access logs are kept
- e) Server/host URL, if different than your own
- f) Web site name(Domain name) and URL
- g) URLs to graphics (Only if you plan to customize SurfReport's look. You can use a custom banner and a custom background).



Part 1: Installing SurfReport

Uncompressing `surfreport.tar`

The following steps will uncompress `surfreport.tar` *

- 1) Open a Telnet session.
- 2) Go to the directory where `surfreport.tar` resides. You can verify this by typing:
% `ls`<return> You should see a file named `surfreport.tar` in the list that this command generates.
- 3) Type: % `tar -xvf surfreport.tar`<return>
- 4) Verify that `surfreport.tar` was uncompressed by typing % `ls`<return> In the list that follows, you should see that a directory named `surfreport` has been created. You'll also notice that `surfreport.tar` still exists. Uncompressing a tar file does not destroy the tar file itself. Leave `surfreport.tar` here for right now. If you wish, you can delete it after the installation.

* Note: to denote a Unix command, file name, or directory in this manual we will use bold text in a Courier font. When asked to type a Unix command, type the portion after the % sign (the % sign represents the Unix command line prompt). For example, instead of typing % `tar xxx` simply type `tarxx` . Words enclosed in <> are keys on your keyboard. For example, instead of typing <return> , simply press the return or enter key.



Part 1: Installing SurfReport

RUNNING THE INSTALLER

Now that you've uncompressed `surfreport.tar`, you're ready to install SurfReport. You should be in the directory where `surfreport` resides (check step 4 in **UNCOMPRESSING** `surfreport.tar` if you're uncertain about this).

Type: % `surfreport/install`

After typing this command you'll see a message welcoming you to SurfReport. Pay special attention to the following message, **“You will be installing SurfReport for the following platform...”** Make sure that the platform that you have and the platform SurfReport thinks you have are the same. If your version of SurfReport is not the correct platform, stop now and go to <http://software.bienlogic.com/SurfReport/> to download the correct version of SurfReport. Continue with the installation if you have the correct version of SurfReport.

To finish this part of the installation, SurfReport needs to know three things: HTTP Server CGI location, Path to PERL Interpreter, and Path to SENDMAIL. You will have to provide these locations if SurfReport cannot find them on its own.

At any time during this part of the installation, you can type 9 to ABORT installation of SurfReport. If you do need to abort installation, you can continue at some later time by repeating the steps in *Part 1: Installing SurfReport*.



Part 1: Installing SurfReport

SurfReport needs the following information:

```
(1) HTTP Server CGI location [unknown]
(2) Path to PERL Interpreter [ /bin/perl ]
(3) Path to SENDMAIL [ /usr/lib/sendmail ]

(9) ABORT Installation of SurfReport
Current Directory /export/httpd/htdocs/clients/mikep
SurfReport Home Directory /export/httpd/htdocs/clients/mikep/surfreport

Type 1,2,3 or 9 to make your selection. Type 0 to continue [0]
```

A screen like this will appear when you continue with the installation.

After typing 0 to continue, SurfReport will ask you to enter your name. *This name cannot be changed.*

Once you have confirmed your name, SurfReport will generate a unique public key for your copy of SurfReport. Make sure to record this 16-digit alpha-numeric code. If you pay for SurfReport, Bien Logic will issue a private key so you may use SurfReport continuously.

Finally, SurfReport will ask you to provide your email address and any comments that you may have.

Congratulations! You've just completed the second part of SurfReport's installation. The two remaining parts need to be completed using your web browser.



Part 1: Installing SurfReport

```
***SurfReport 2.0 Registration
***
***Enter your name:
(will show as the registered user name in SurfReport 2.0 and cannot be
changed!)
phillips
***Use the name "phillips" ? (y/n)
(if you answer "y" you will not be able to change it again!)
y
***Your public key for this install of SurfReport 2.0 is:
32A1DFS1036532DF
***
You can use SurfReport 2.0 for 30 days until you receive your private key.
Your private key will arrive as soon as we receive your payment.
*** This key will now be sent to BienLogic, Inc.
*** Enter your e-mail address:
(the same address you gave us when you downloaded SurfReport 2.0)
phillips@bienlogic.com
*** Is this correct: "phillips@bienlogic.com" ? (y/n): y
*** Any comment ? (y/n)
```

Provide your name and email address to complete the second part of SurfReport's Installation. Make sure to record your public key code for future reference.

USING YOUR WEB BROWSER TO COMPLETE THE INSTALLATION

You'll need to use your web browser to personalize SurfReport. First, you'll configure `surfreport.cgi`, and then you'll set your individual preferences for your report.



Part 1: Installing SurfReport

If you're already anxious to use SurfReport, don't worry, these last two parts shouldn't take long. In fact we've already set default preferences for you. You can simply confirm these preferences and launch SurfReport. As you become more familiar with SurfReport, you'll want to go back and customize these preferences to your needs. These preferences are explained in further detail in *Part 2: SurfReport's Settings*.

CONFIGURE `surfreport.cgi`

This part is actually pretty easy. After you have finished running the installer, you will see a screen that looks like the one below. This screen just tells you to go to your cgi-bin (the one you gave the path to earlier) and open `surfreport.cgi` using your browser.

SurfReport will have already attempted to provide accurate information for your configuration. However, you should still check to make sure that SurfReport has the correct information.

```
Phase 2: To complete the installation, please point your browser to
http://your.domain.com/your-cgi-bin/surfreport.cgi
```

```
After you have completed the configuration,
come back here and press the enter key.
```



To finish the installation go to your cgi-bin and open up `surfreport.cgi` with your web browser.



Part 1: Installing SurfReport

Check:

- 1) The path to your access logs
- 2) The URL to your server
- 3) The URL to your web site
- 4) The URL to your SurfReport folder

site's name

example Top 40 Visitors That Have Accessed Bien Logic's Site at Least Once

Fill in the name of your web site to personalize your SurfReport.

When this information is correct, fill in your site name in the appropriate field, and then click on the "Update Configuration" button. **You need to click on the "Update Configuration" button at least once.** After you've done this, you've completed this part of the installation. To finish the installation, click on the "Configure SurfReport's form interface" link.

Note: **You need to click on the "Update Configuration" button at least once.** Click on the "Configure SurfReport's form interface" link to finish the installation.

Update Configuration

Configure SurfReport's form interface



Part 1: Installing SurfReport

CONFIGURE SURFREPORT'S FORM INTERFACE

This final section is even easier than the previous one. Take a minute to review the standard settings. If they look good to you, simply click on the “Make Default” button. These settings can be changed at any time. **You must click on the “Make Default” button at least once to save these settings.**



Make Default

Click the “Make Default” button to save settings.

After you click on the “Make Default” button, you’ll come to a page that looks like the one below. Now you’re ready to try SurfReport. You can launch SurfReport at any time by going into your `surfreport` folder and opening `surfreport.html` with your web browser.

You can change the configuration settings or default settings at any time by opening `surf-report.cgi` with your web browser (remember that this file is in your `cgi-bin` folder).

SurfReport has changed the defaults in your interface form

Now, you can:

- [Try SurfReport](#)
- [Make more changes to the form](#)
- [Make more changes to your configuration](#)

After you've clicked on the “Make Default” button, you're ready to try SurfReport.



Part 1: Installing SurfReport

OTHER SOURCES OF DOCUMENTATION

Online sources of documentation can be found at:
<http://software.bienlogic.com/Surfreport/technical.html>.

Dr. Dobbs Sourcebook, September/October 1996, contains the article “Analyzing Web Traffic with Perl” by Frederic Bien, Michel Delory, Richard McGillis.

There are many other documents regarding traffic analysis, log formats, etc. Many are excellent sources that describe access and referrer log analysis.

REMOVING SURFREPORT

The following steps will remove SurfReport from your server:

- 1) Open a Telnet session.
- 2) Go to the directory where `surfreport` resides.
- 3) Type: `% surfreport/deinstall<return>` SurfReport will first ask you to confirm your request and then it will ask you to provide the path to your `cgi-bin` directory. You can confirm that the `surfreport` directory has been removed by typing: `% ls<return>` You'll notice that `surfreport` does not show up.



Part 1: Installing SurfReport

WINDOWS NT INSTALLATION

The installation of SurfReport for NT differs from UNIX only in so far as handling the distribution file, install.exe.

Move install.exe to a temporary directory on your server and run it. Answer the prompts which include the following with example values:

```
Web document root directory=\inetpub\wwwroot
Web cgi-bin directory=\inetpub\wwwroot\cgi-bin
mail server=mail.bienlogic.com
mail address=lane@bienlogic.com
mail name=lane
Web URL=http://www.ntsitem.com
```

The NT installation should complete these three important steps:

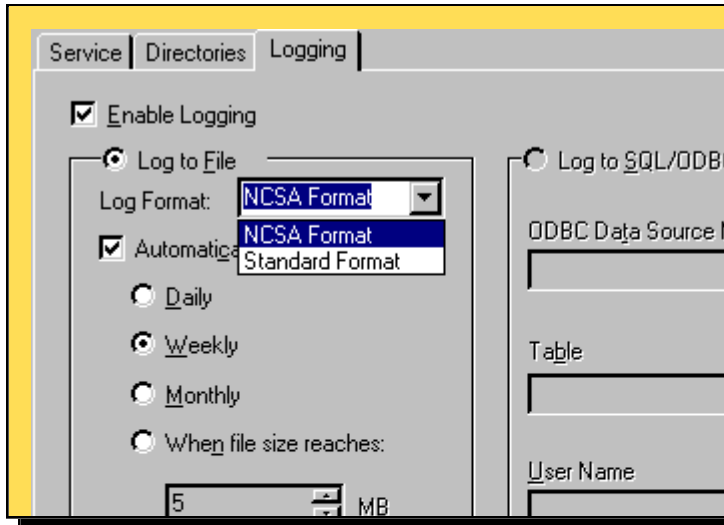
- Create the SurfReport directory tree within your Web document tree and copy to it the distribution files and copy surfreport.cgi to the cgi-bin directory you specified.
- Copy to your system32 directory perl.exe, perl100.dll and blat.exe which provide scripting and e-mail support.
- Update your registry to identify cgi as a Web file type that is executable by the perl interpreter.



Part 1: Installing SurfReport

Your Web server in turn must recognize that the cgi-bin directory is a valid directory for running scripts terminated by the suffix, cgi.

After these steps are complete, reboot the server to allow the registry entries to take effect. Go back to page 15, “**CONFIGURE surfreport.cgi**” and use your Web browser to configure SurfReport’s configuration files **and** form interface. The URL address will be <http://your.domain.com/your-cgi-bin/surfreport.cgi/config>



The screenshot shows the configuration interface for SurfReport, specifically the 'Logging' tab. The 'Enable Logging' checkbox is checked. Under the 'Log to File' section, the 'Log Format' dropdown menu is open, showing 'NCSA Format' selected. The 'Automated' checkbox is checked, and the frequency is set to 'Weekly'. The 'When file size reaches' field is set to 5 MB. The 'Log to SQL/ODBC' option is not selected.

A Note About Microsoft’s Information Server

This server is completely compatible with SR, however there is one pitfall. Be sure to configure your cgi-bin directory as readable and executable and be sure to enable NCSA style log formats.



PART 2: SURFREPORT'S SETTINGS

This section explains both the settings on surfreport.html and the corresponding results displayed on surfoutput.html.

Deliver my SurfReport: In addition to being displayed on your computer screen, SurfReport's results can be sent to one or more e-mail addresses. By default, SurfReport's results are sent to your computer screen. If you do not want the results to be sent to your screen, make sure that the box to the left of "On the web" is not checked. If during SurfReport's configuration you opted to have your results sent to a permanent html file, the results will also be sent to an html file with the name you specified. Check the box to the left of "By e-mail to:" if you want SurfReport's results mailed to someone. You will also have to provide one or more e-mail addresses in the text field. If you put more than one e-mail address, separate the addresses by a comma and a space. The results will be sent in a format similar to the one below.

**Deliver my
SurfReport**

On the web

By e-mail to:

Spreadsheet format

In the above example, SurfReport's results will be sent to the web as well as being e-mailed to webmaster@yourcompany.com.

Part 2: SurfReport's Settings

SurfReport's results can also be sent by e-mail in tab-delimited format for easy integration into spreadsheet or database applications. Check the box to the left of "Spreadsheet Format" if you would like this option. Do not check both "By e-mail to:" and "Spreadsheet Format" unless you want two versions of SurfReport's results. If you send SurfReport's results by e-mail, you will not have access to the graphical capabilities available on the web. Note: the file size of your results may be very large depending on the parameters you have selected and may cause problems with systems that have limits on e-mail transmissions.

Name my SurfReport: By default your SurfReport results will be named surfoutput.html. However, you can name this file anything you want. If you do not change the name, the file will be rewritten each time you run SurfReport.

Name my SurfReport .html

Daily Totals: Daily Totals consist of the total number of visits (Visits), the total number of page or html hits (Pages), and the total number of hits (Hits) listed for each day of the log period. Daily Totals are automatically sent as part of your report.

Daily Totals for YourSite.com

<u>Date</u>	<u>Visits</u>	<u>Pages</u>	<u>Hits</u>
Mon Oct 21 1996	2789	2972	94057
Tue Oct 22 1996	1818	2169	73191
Wed Oct 23 1996	2991	4162	120110
etc...			



Part 2: SurfReport's Settings

Period Totals: This feature measures totals for Unique Visitors, Visits, Pages, and Hits for the entire period. SurfReport tracks unique visitors based on each computer's IP address. An additional visit will be logged when more than 40 minutes passes between any two entries of a unique visitor. For example, if Henry Jackson uses the same computer to go to your site on Monday, Tuesday, and Wednesday, SurfReport will record three visits from one unique visitor. Period Totals are automatically sent as part of your report.

Period Totals

<u>Unique Visitors</u>	<u>Visits</u>	<u>Pages</u>	<u>Hits</u>
13733	20240	28792	791939

Daily Averages: SurfReport computes the daily averages for Pages/Visitor, Visitors, Pages, and Hits. To do this, SurfReport divides the period totals by the number of days in the period. Pages/Visitor shows the average number of pages each of your visitors reviewed (rounded to a whole number).

Daily Averages

<u>Pages/Visitor</u>	<u>Visitors</u>	<u>Pages</u>	<u>Hits</u>
2	2217	4113	113134

Hourly Averages: This breaks down the total number of hits by hour for the period of your report. It also shows you the percentage of hits by hour. In the example below, there were a total of 5973 hits to the site between the hours of 10am and 11am (This number happens to be based on a seven-day log period, but it could be for any period length). The site received 5.3% of its total hits from 10am to 11am.



Part 2: SurfReport's Settings

Hourly Averages

<u>Time</u>	<u>Hits</u>	<u>Percentage</u>
10 am to 11 am	5973	5.3%
11 am to 12 pm	1818	5.5%
12 pm to 1 pm	5359	5.3%
etc...		

Top “X” Visitors That Have Accessed “Your Site” At Least “Y”Time(s):* Visitors (based on each computer’s IP address) are listed from top to bottom starting with those with the most frequent hits to those with the least frequent hits. You’ll notice that while some visitors have many hits, they only have a couple visits. If you want to see the visitors with the most visits, enter a large number for the “Y” variable. For example, if you enter 20 for the “Y” variable, SurfReport will only list visitors who have visited your site at least 20 times.

Visitor Report Top visitors who have accessed my site with at least time(s)
Top companies that accessed my site
Top countries that accessed my site

* For the previous settings, X, Y, and Your Site are variables that you can set. X represents the amount of items that will be listed. Y represents the minimum number of hits. Your Site is the name that you provide in the site name field on surfreport.cgi.

Part 2: SurfReport's Settings

Top “X” Companies/Domains That Have Accessed “Your Site”: This feature ranks companies from top to bottom starting with those with the most frequent hits to those with the least frequent hits. SurfReport defines a company** or domain as anything to the left of the two-digit country code suffix (excluding www). If you do not wish to view this option, you can enter 0 for the “X” variable.

Top 100 Companies/Domains Accessing YourSite.com

5892 ix.netcom.com
5767 proxy.aol.com
5118 ctron.com
2504 compuserve.com
etc...

Top “X” Countries That Have Accessed “Your Site”: This lists countries from top to bottom starting with those with the most frequent hits to those with the least frequent hits. SurfReport determines the country based on a site’s two-digit country code suffix. Companies/domains in the United States omit the country code suffix. If you do not wish to view this option, you can enter “0” for the “X” variable.

Top 100 Countries Accessing YourSite.com

58180 US Commercial
40165 Network
19750 US Educational
1836 Canada
etc...



Part 2: SurfReport's Settings

Top “X” Referrer Pages To “Your Site”: This lists files from top to bottom starting with those with the most frequent hits to those with the least frequent hits. *This section will only work for http servers that record log entries in combined log format.* If your server does not use combined log format, then there should be a “0” in the “X” variable. A referrer page is the file name of the page that linked to your site. For example, if there is a link to your site from the URL <http://www.coolsites.com/links.html>, then www.coolsites.com/link.html will show up as a referrer page.

File Report Top files in my site that have been accessed at least time(s)

Display HTMLfiles only All files
 Document titles File names

Top “X” Files In “Your Site” That Have Been Accessed At Least “Y” Time(s): This lists files from top to bottom starting with those with the most frequent hits to those with the least frequent hits. This feature offers you the option of analyzing all of the files (includes html, cgi, gif, jpg, and txt files) or just html files. In addition, SurfReport can display the files either by their file name, such as `index.html`, or by the file’s title, such as “Bien Logic’s Homepage.” The file’s title is based on the text between the `<title>` `</title>` tags in the html page header.

Part 2: SurfReport's Settings

Specific Visitor Report: This will allow you to see which files in your site have been accessed by a particular visitor or visitors. You can input the name of a visitor by machine name in its entirety, or by any substring within the name, and you will get a listing of all the files that were accessed by those visitors containing that word or string. The results will list total files (i.e. including files looked at multiple times), as well as unique files.

Visitor Filter: This section is used to filter out a particular visitor or set of visitors from the report. It is very useful for eliminating hits from within your organization, so that the results will be showing only accesses from outside of your organization. Any visitor or list of visitors that contains the string you enter will be excluded from the report. If you have a very large number of visitors you wish to exclude, it may be easier to click the radio button "Exclude all visitors but these" and list the names of the companies which you want included. All others will be omitted from the report.

Specific File Report: This will allow you to see which clients have accessed a particular file or files within your site. You can input the name of your file in its entirety, or by any part of the file, and you will get a listing of all visitors that have accessed a file containing that word or string. Note that the file name must be the URL name, and not the actual title of the html document. The results will list total clients (i.e. including repeat customers), as well as unique clients.

A detailed discussion of SurfReport's filtering capabilities can be found in **Part 4: Advanced Tasks:** *Using Filters to Customize Your SurfReport Results.*



Part 2: SurfReport's Settings

File Filter: This box is used to filter out accesses to unwanted files. Any file that contains the string you enter will be excluded from the report. This feature comes in handy with .cgi files (which are files that process forms and other types of requests, but are never viewed by a browser). This filter's behavior can be reversed just like the Visitors filter. A prime use of this reversed file filter is to provide reports based on users folders. If you are a service provider, you can analyze your entire log and issue a report for one or a list of specific users based on their web folder names.

Browsers: For servers that use combined log format, SurfReport will list from top to bottom the highest to lowest percentage of browser types and versions used to access your site during the log period. All browsers that represented less than 1% of the users are lumped together with the label "less than 1% each."

Browsers

31.8%	Netscape Navigator 2.x
19.1%	MS Internet Explorer
18.9%	Netscape Navigator 3.x
2.5%	AOL Browser
etc...	

Operating Systems: For servers that use combined log format, SurfReport will list from top to bottom the highest to lowest percentage of operating systems used to access your site during the log period.



Part 2: SurfReport's Settings

Summary of HTTP Errors: This section lists the HTTP Errors that were encountered during the log period. SurfReport lists from left to right the frequency of the error, the error code, and the name of the error.

Process	<input type="radio"/>	The entire period available		
Files for	<input checked="" type="radio"/>	The period:	From	
				Oct
				1
				96
		To		Nov
				1
				96

Process Files for: SurfReport gives you the option of analyzing the entire period contained in your log files or only a selected portion of the period. If you want to analyze only a portion of the log period, you must click on the radio button for “The period.” Then you need to select the start and end dates for the period.

Access Log Files: This is the name of the file or files that SurfReport will perform its analysis upon. Any valid uncompressed log file that matches the name you specify will be used to generate the report. Multiple files separated by a space or comma are acceptable.

Password: This is the password that allows SurfReport to operate. The default password is “SurfReport.” If you change this password in your configuration file, you’ll need to change it here to reflect the new password.

PART 3: GENERATE A SURFREPORT

Using Your Default Settings

After you have completed the steps in *Part 1: Installing SurfReport*, you can generate a SurfReport at any time by going to `surfreport.html` with your web browser. Remember that this html page is in the folder or directory called `/surfreport/`.

`surfreport.html` will display the default settings that you selected when you installed SurfReport. If you wish to use these settings, click on the “Launch SurfReport” button to generate a new report. Make sure that the string appearing in the “Access Log Files” box is a valid reference to the file or files you wish to process.

SurfReport may take a couple of minutes to process depending on the size of your log files. The approximate time needed to process your request will be displayed on your screen. Feel free to open a new browser and surf the Net while you wait. If you do not want your results displayed on the web and are only having the results sent by e-mail, you do not have to wait for anything. SurfReport will complete its analysis and mail the results automatically.



Part 3: Generate a SurfReport

In-Depth Analysis

After creating a general report, you may want to perform in-depth analysis. Remember that you can generate as many reports as you want.

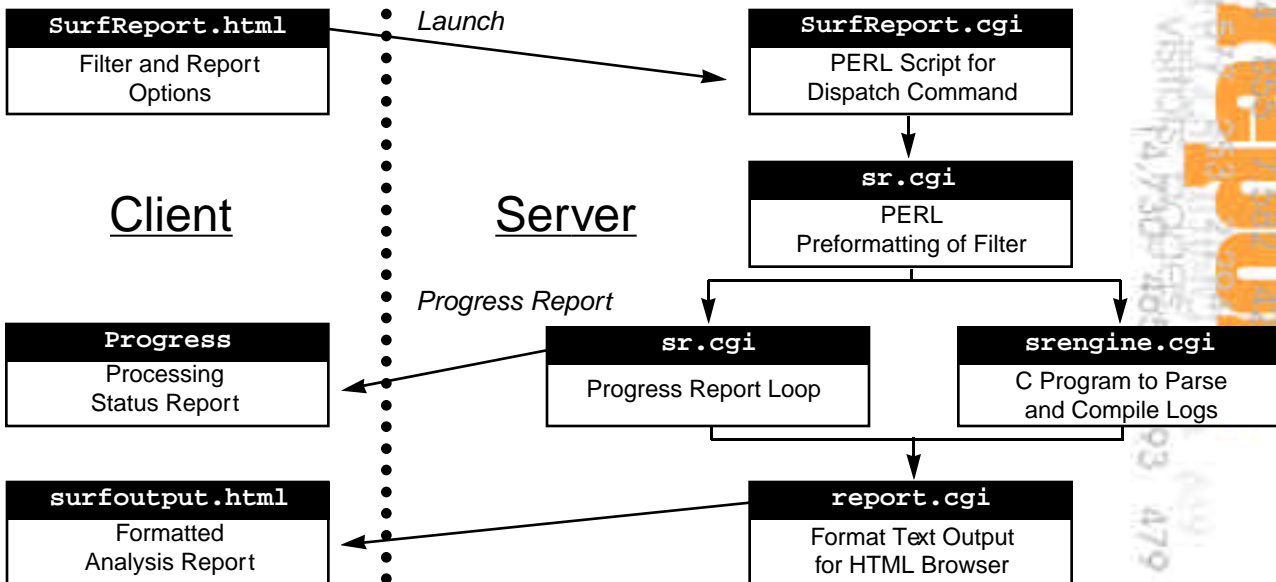
You can perform specific visitor or file analysis by using the “Specific Visitor Report,” “Visitor Filter,” “Specific File Report,” and “File Filter” functions. You can also narrow the focus of your report by having SurfReport analyze a particular time period.

To create a specialized report, change the settings on **surfreport.html** and click on the “Launch SurfReport” button.



PART 4: ADVANCED TASKS

Before showing you some examples of how SurfReport can be used for difficult analysis problems, it is helpful to understand the overall architectural flow:



Part 4: Advanced Tasks

Notice that SurfReport uses extensively Perl scripts to parse, format and present to the Web browser the compilation of data from srengine.cgi, a C program. Also, it is worth pointing out that surfreport.cgi is the only program that is located in the cgi-bin directory of your web host. All other programs are found in the cgi directory of the surfreport hierarchy.

Why is this important? It does provide a very layered approach to the application. This makes it possible for example to easily run SurfReport from the command line. It also makes it easy to perform small formatting changes in the output although we do not recommend fooling around with the Perl code.

Scenario One: “I’d like to be able to analyze how often visitors reached my site and I’d like to eliminate from the report other visitors. Furthermore, I’d like to see what files were accessed and also eliminate other files from the report. How can I do this?”

To accomplish this task, let’s take a close look at the visitor and file filters:



Part 4: Advanced Tasks

Using Filters to Customize Your SurfReport Results

While filters are useful in giving you a report in which your own company's accesses are eliminated, you may find other uses of the filters to give a customized report. I have listed three possible examples below.

Visitor Filter: This section is used to filter out a particular visitor or set of visitors from the report. It is very useful for eliminating hits from within your organization, so that the results will be showing only accesses from outside your organization. Any visitor or list of visitors that contains the string you enter will be excluded from the report. If you have a very large number of visitors you wish to exclude, it may be easier to click the radio button "Exclude all visitors but these" and list the names of the companies which you want included. All others will be omitted from the report. For example, you can have a specific report on Japanese visitors by typing "* .jp" in the filter box, checking the exact match button and the "Exclude all visitors but these" button.

File Filter: This box is used to filter out accesses to unwanted files. Any file that contains the string you enter will be excluded from the report. This feature comes in handy with .cgi files (which are files that process forms and other types of requests, but are never viewed by a browser). This filter's behavior can be reversed just like the Visitors filter. A prime use of this reversed file filter is to provide reports based on users folders. If you are a service provider, you can analyze your entire log and issue a report for one or a list of specific users based on their web folder names.



Part 4: Advanced Tasks

The two reversed filters can be used in conjunction. This allows you to get a report on a subset of visitors viewing a subset of files. For instance, to get a report on Belgian visitors accessing the “belgian_beers” folder on your site you could type:

Visitor Filter box: *.be

Check : **Exact, Exclude all visitors but these**

Visitor Filter	<input type="radio"/> Exclude visitors matching these names:	<input type="text" value="*.be"/>
	<input checked="" type="radio"/> Exclude all visitors but these:	<input type="text" value=""/>
		<input type="radio"/> Substring <input checked="" type="radio"/> Exact <input type="checkbox"/> Case sensitive
		<input type="checkbox"/> Exclude numerical addresses

File filter box: belgian_beers/

Check : **Substring, Exclude all files but these**

File Filter	<input type="radio"/> Exclude files matching these names:	<input type="text" value="belgian_beers/"/>
	<input checked="" type="radio"/> Exclude all files but these:	<input type="text" value=""/>
		<input checked="" type="radio"/> Substring <input type="radio"/> Exact <input type="checkbox"/> Case sensitive

Part 4: Advanced Tasks

Example One: Logging Only Page Hits to Your Site

Consider looking at a page for casting agents that has embedded about 80 gif images into one html page. Besides taking a long time to load, one “hit” to this page would log over 80 entries to the log file. While this is rare, it is nonetheless true that “Hits” may not be a completely accurate measurement of your website usage. While the Daily Totals, Period Totals and Daily Averages section of SR keeps separate information on page hits and all hits, the remaining sections default to outputting data based only on total hits. This can be remedied by using the File Filter.

File Filter	<input type="radio"/> Exclude files matching these names:	<input type="text" value="*.htm, *.html"/>	<input type="checkbox"/>	
	<input checked="" type="radio"/> Exclude all files but these:	<input type="text"/>	<input type="checkbox"/>	
		<input type="radio"/> Substring	<input checked="" type="radio"/> Exact	<input type="checkbox"/> Case sensitive

In the File Filter section, select “Exclude all files but these:” and type in *.htm, *.html. Now when you run SurfReport, you will see the sections in Daily Totals, Period Totals and Daily Averages for Pages and Hits are exactly the same. All the other sections (e.g. Top 40 visitors to your site) will be measuring only page hits and eliminating hits to gifs, jpgs, cgi scripts etc. hence giving you a bit more accurate account of the visitors to your site.

Part 4: Advanced Tasks

Example Two: *Determining your user's browser and platform preferences.*

Sometimes you might find it desirable to find out a particular user's web-browser preference or machine type. If, for example, you know that a particular company has been checking out your site is using 90% Macintosh machines, and 10% Intel based machines, then you may change your marketing strategy towards them to emphasize Mac based products, etc. For this particular application, you must have your server configured to produce the Combined Log Format, which holds information about the User Agent (browser and machine type). The two checkboxes "Show browsers used by my visitors" and "Show operating systems used by my visitors" should be checked. Then, put the name of the user in the Visitor Filter section and click the "Exclude all visitors but these" radio button.

Browsers Show browsers used by my visitors

Operating Systems Show operating systems used by my visitors

When you launch SurfReport, you will be getting results for only the specific user you entered. If you are very specific about that user, e.g. if you put in "elvin.bienlogic.com" rather than just "bienlogic", you can find out the exact type of machine that elvin is, Macintosh, UNIX based, Windows based etc. In this particular case you will see the following:

Part 4: Advanced Tasks

Operating Systems

100.0% Macintosh 68K

So now you know that elvin.bienlogic.com is a 68K based Macintosh. (It is indeed a Mac Quadra 800). You will also find out which browsers are preferred by the user of elvin:

Browsers

81.1% Netscape Navigator 3.x

18.9% Netscape Navigator 2.x

Example Three: *Finding all users from a particular country or organization type that have accessed your site.*

Let's say you've just run SurfReport on your log files and discover from the Countries section that there were 28 hits from visitors in Peru. Now you become curious as to who in Peru was interested in your site, and would like to find out these domains and the files that they looked at. You will first need the country code for Peru which can be found in Appendix B. You will see that Peru's country code is "pe".

Visitor Filter	<input type="radio"/> Exclude visitors matching these names:	<input type="text" value="*.pe"/>	<input type="checkbox"/>
	<input checked="" type="radio"/> Exclude all visitors but these:	<input type="text"/>	<input type="checkbox"/>
	<input type="checkbox"/> Substring	<input checked="" type="radio"/> Exact	<input type="checkbox"/> Case sensitive
	<input type="checkbox"/> Exclude numerical addresses		

Part 4: Advanced Tasks

Now, return to SurfReport, and go to the “Visitor Filter” section and select “Exclude all visitors but these:” Now, put `*.pe` in the input box and choose the “Exact” radio button. The idea here is that you want a report that will only include visitors that have the string “xxxxx.pe” in their names, but want to exclude `www.petsupplies.com`, `www.peaches.net`, etc. By using the “Exact” feature along with a wildcard in the filters, you can get all visitors that contain “xxxxx.pe” but with nothing after the “.pe” which should indeed be all visitors from Peru. After you launch SR you should notice that the Top Countries section now has only one country listed in it - Peru.

You should also get a separate listing of all the visitors from Peru (which will probably be less than 28 since the above section counted the number of hits, not visitors) which accounted for all 28 hits. Now you would like to know which files these Peruvians were interested in so you return to the SR input page. You can now go to the “Specific Visitor Report” section and again input `*.pe` with “Exact” and run SR again. There will now be a section on your output that says:

Visitors matching the string

`*.pe`

accessed the following files in this period

And you will see all the files accessed by visitors from Peru. If, in the File Report section, you had selected “HTML Files”, then the listing here will only include html files and as such may not add up to 28. You can run SR with “All Files”, and you will get all your visitors from Peru, and all 28 of the files - html and otherwise - that they accessed.



Part 4: Advanced Tasks

Scenario Two: “I am an ISP and I’d like to run my reports on an unattended basis against logs in standard and compressed format.”

Due to the modular nature of SurfReport, this is easy to do as well.

Let \$surf be the home directory for surfreport. Now go to the cgi-bin directory containing surfreport.cgi. Issue the following command:

```
surfreport.cgi $surf/surfinput.txt
```

The file surfinput.txt is a text file containing name=value pairs for all the settings that control the output of SR. The release of surfreport contains a sample surfinput.txt file to act as a template. Each parameter is documented.

One parameter, Decompress, allows you to manage compressed log files. Setting this parameter to on will make SurfReport handle files compressed using compress or gzip compression utilities. Currently the zip format is not supported. Make sure that SR has permission to read/write these files.

Finally, by employing a simple shell script that updates the date fields in the surfinput file prior to execution, it is possible to create a complete batch job that runs a current daily, weekly, or monthly analysis on an unattended basis of either compressed or standard log files.



Part 4: Advanced Tasks

Scenario Three: “I have many virtual hosts on my machine. Do I use one copy or multiple copies of SR?”

You will use multiple copies of SR. Why? It will be easier for you to track and to upgrade virtual hosts on an individual basis. Installing a single copy per virtual host allows for site by site customization, upgrading, and asset management.



PART 5: CHARTS

Once SurfReport has been run, the graphics will be produced from the data in your last created output, usually "surfoutput.html". At the end of the "Daily Totals" section, you will see 4 thumbnail images entitled "All", "Visits", "Pages", and "Hits". A click on any of these thumbnails will take you to a page which will chart the results for the type of data you clicked on. There are also thumbnails leading to graphics for the "Hourly Averages" and "Visitors" sections.

The information contained on the graphics should be self-explanatory, as the axes and titles of the graphic will be spelled out. Default colors and chart types are set within SurfReport, however near the bottom of the graphic page you will see a link that says "Modify Chart". A click on this link will take you to a page where you can change the type of chart, colors, etc. These are divided into Basic Chart Options and Advanced Chart Features. When you modify your chart, be sure to click on the "Make Chart" button to see your changes.

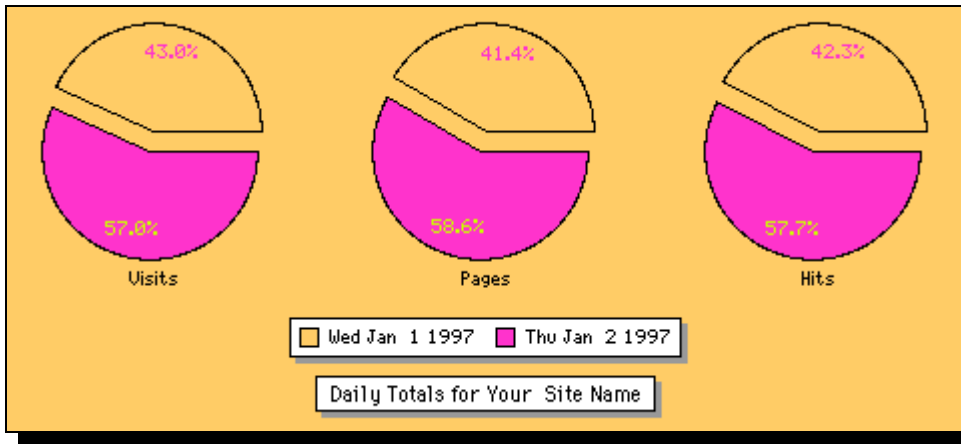


Part 5: Charts

Basic Chart Options

Chart Type: Users have the option to change between various types of charts including Area, Bar, Line, Pie, and Point charts. To select the one you want to use, just click the corresponding radio button. The default for all the charts is the Bar Chart, except for “Visitors” which defaults to the “Area” type.

Note: Pie charts are not a good tool when there is a lot of data (many dates or many visitors). The pie slices get too small to read, and you may have to blow them up a lot. It is better to use another chart type in these cases.



Part 5: Charts

Basic Chart Options	
Chart Types <input type="radio"/> Area <input type="radio"/> Bar <input type="radio"/> Line <input type="radio"/> Pie Chart <input type="radio"/> Point	Orientation <input type="radio"/> Horizontal <input type="radio"/> Vertical
Dimensions Width <input type="text"/> Minimum Height <input type="text"/> Minimum	Colors Visits <input type="checkbox"/> Color <input type="text"/> Pages <input type="checkbox"/> Color <input type="text"/> Hits <input type="checkbox"/> Color <input type="text"/>

The interface shown above is used to modify basic chart options such as chart type, orientation, dimensions and colors.

Orientation: The user has the choice of orienting their charts either horizontally or vertically. To select the one you want to use, just click the corresponding radio button. The default for all the charts is a vertical orientation, except for “Hours” and “Visitors” which is horizontal.

Dimensions: Users have the capacity to be able set their own custom dimensions for the size of their charts. The user is given the option to change either or both the width and height of all their charts. The dimensional units used in the charts are measured in pixels. All the user needs to do is enter the size they want for their width and height. These will be the minimum dimensions. The chart will never be smaller than the number you enter, but can be larger. The charting program determines the minimum size required to draw a



Part 5: Charts

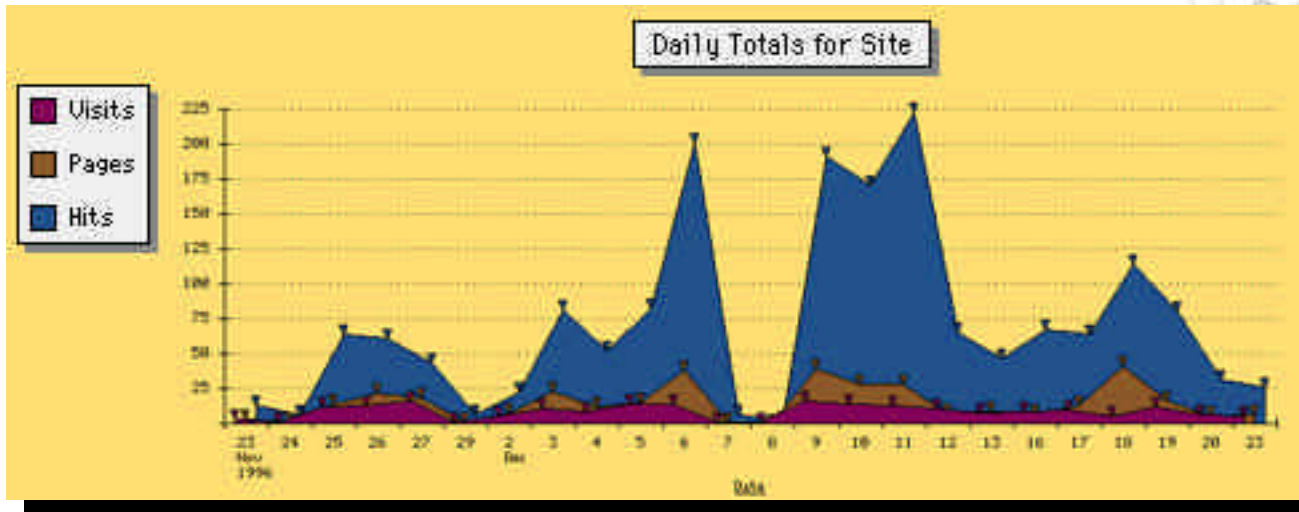
given chart, so a chart can never be smaller than this minimum size. If you enter smaller numbers they will be ignored. Only numbers that are bigger than the minimum size will affect the chart. This minimum size is not known in advance. It is computed on the fly and depends on the data to be charted and the user's parameters such as title font size, position and so forth. The chart dimension is the smallest rectangle that encloses all the chart elements (chart itself, titles and labels).

Advanced Chart Features	
Titles Title Font Size <input type="text"/> <input type="checkbox"/> Frame <input type="checkbox"/> Shadow Position <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Label Box Label Font Size <input type="text"/> <input type="checkbox"/> Frame <input type="checkbox"/> Shadow (Label Box is shown when more than one series is charted) Position <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

For advanced charting features involving position and font sizes of titles and labels, input desired specifications into the above dialog box.

Part 5: Charts

Colors: Users have the option to change the color of the graphics. These choices are appropriate for all the chart types with the exception of the pie chart. To change the colors, use the pull-down menu called “Color”. The checkbox to the left of the color menu will tell you which parameter you are changing. Any parameter that does not have an “X” in its checkbox will not be graphed; therefore you can get a chart that shows only hits, for example, from the “All” graphics if you wish.



The chart above was generated using user-specified colors, orientation and dimensions.

Part 5: Charts

Advanced Chart Options

The Advanced Chart Options enable you to fine tune your chart output in terms of position and font sizes of titles and labels and to customize your tick marks and grid lines.

Titles and Labels: The title of your graph is based on the data being used. This is opposed to the labels which tell you which color represents which graphed item. No labels will be shown if there is only one item to be graphed. The titles and labels will appear in boxes, on the edges of the graph. Depending on where in relation to the graph you would like the titles and labels to appear, you can use the set of 8 radio buttons in the

Axis	
Tick Mark Labels Font Size <input type="text"/> <input type="checkbox"/> Reverse Main Axis Direction	
X Axis	Y Axis
Tick Marks Position <input type="text"/>	Tick Marks Position <input type="text"/>
Show Axis Label <input type="checkbox"/>	Show Axis Label <input type="checkbox"/>
Show Tick Marks Labels <input type="checkbox"/>	Show Tick Marks Labels <input type="checkbox"/>
Show Grid Lines <input type="checkbox"/>	Show Grid Lines <input type="checkbox"/>

Additional advanced charting features involving axis can be controlled with the above interface.

Part 5: Charts

“Position” sections to place them. In other words, if you would like the title of your graph to appear at the lower right hand corner of your graphic, then you would click the lower right hand radio button in the “Position” section of the Titles area, etc. You can also eliminate the little shadow that appears behind the box containing the title (or label) by unchecking the “Shadow” box. This will leave the title in a frame with no shadow. The frame can also be removed by unchecking the “Frame” checkbox.

Note: The “Frame” option will always be in effect when the “Shadow” checkbox is checked, i.e. unchecking the “Frame” box while leaving the “Shadow” box checked will have no effect. This is because you need a frame in order to draw a frame shadow.

Axes: The axes form the framework for your graphic. The “X” axis is the horizontal axis, and the “Y Axis” is the vertical axis. You can have the tick marks appear on the chart inside, outside or in the middle of (straddling) the axes; and you can change the font size of the tick mark labels. You can also remove the grid lines that are shown in the chart. The grid lines are dotted lines to help you determine the relative scale in your graphics. They will appear perpendicular to your axis of orientation, i.e. they will go across horizontally when you choose a vertical orientation, and vice versa. Therefore only the X axis grid lines will be there, or the Y axis grid lines, but not both.

Reverse main axis direction: This changes the sorting of the main axis. Usually data on the main axis (the axis with labels not numbers) is sorted in increasing order from the axis origin (either up = horizontal orientation, or right = vertical orientation). When the reverse main axis direction is checked, the data starts at the far end of the axis and increases towards the origin.

PART 6: GLOSSARY OF TERMS & DEFINITIONS

Access log

The entries within your log files that record accesses to your site. It contains all the information about the activity to your site. It will contain a record of the date, the time, and the client accessing your web pages. An example of one is shown below:

```
max-rb-134.connectnet.com - -  
[22/Sep/1996:16:27:25 -0700] "GET /~thomas/bugreport.html HTTP/1.0" 200 3924
```

Activation code

A unique code that activates SurfReport after the trial period has expired. Bien Logic will send you this number after you purchase SurfReport. Please refer to this code if you have questions or concerns regarding SurfReport.

Averages

SurfReport calculates two specific types of averages: Daily and Hourly. Daily averages divide the total number of hits and visitors by the number of days in the period being reported. Hourly averages separate into one-hour increments the average number of hits to your site during a 24-hour period.



Part 6: Glossary of Terms and Definitions

Browser

A software application that lets users interact on the World Wide Web (WWW). The most widely used browsers are Internet Explorer, Netscape, and Mosaic.

CGI

CGI stands for “Common Gateway Interface” and defines an interface for running external applications and programs on a web server. Your cgi scripts and programs should reside in a folder or directory called cgi-bin.

Client

A user on the Internet who accesses files delivered by a server. SurfReport uses the term “Visitor” to represent a user who has accessed your site. A browser is an example of client software.

Combined Log Format

A log format that permits a wider range of data to be captured than the standard Common Log Format. This extra data includes information about the http referrer and the user agent. SurfReport has the ability to process this extra information, provided that your server is set up appropriately.

Common log format

The format agreed upon between CERN and NCSA for the creation of HTTP logs. It is this format that most servers will use to create their client log files. (See Access Log for an example).



Part 6: Glossary of Terms and Definitions

Company

Also called the domain name, a company is defined as anything between the first and second “.” in the host name (the names are read from left to right). For example, the company name for <http://software.bienlogic.com> would be bienlogic. In instances such as <http://flirt.com> where there isn't a second period, the domain name is before the first “.”

Compression

Compressing a file is an efficient way to prepare a file before transferring the file to other machines, either via telnet, ftp, or disk. The common UNIX compression algorithms are 'gzip', and 'compress'. Likewise, you can use 'gunzip' or 'uncompress' to restore a file to it's original size.

Countries

In terms of SurfReport, a “country” could also be an organization or other institution, and is determined by the country code that is part of any domain name. See “Country codes”.

Country codes

The two or three letter suffix at the end of a domain name that indicates what type of organization or country the domain is from. Normally, a non-U.S. country has a two letter suffix, whereas organizations or institutions have a three letter suffix. For example, “.edu” indicates that the domain is a university or educational institution. See Appendix A for a complete listing of country codes.



Part 6: Glossary of Terms and Definitions

Custom settings

SurfReport allows the user to choose to run a report with default settings or customize the report according to their own needs. These settings include: a) top number of visitors, companies or countries, b) top number of pages visited on your site, c) number of refers to your site, d) types of browsers used to access the site, e) operating systems of people accessing site, f) various filters to narrow and specialize your report, g) the specific time period you want to look at.

Defaults

Default settings are used in SurfReport on the input interface (surfreport.html) when you first receive the software package, but can be changed at any time, either by using surfconfig.cgi contained in the installation package, or by manually altering surfreport.html at the time of report generation.

Delivery method(s)

The different modes that the SurfReport output can be sent to the user. They include 1) On the web, which generates an html page containing the statistical results of running SurfReport; 2) By e-mail without spreadsheet, which will send the generated report to the e-mail address(es) specified, and 3) E-mail with spreadsheet format, which will send the report to the e-mail address(es) specified in a form that may be dropped immediately into a spreadsheet without further formatting.

Directory structure

In UNIX, a hierarchical method of organizing files by using the forward slash “/” character to define the complete path name to a file. On PC’s, a backward slash “\” is used instead.



Part 6: Glossary of Terms and Definitions

Document titles

The part of an HTML document in between the <TITLE> and </TITLE> tags, if any. The title may or may not be unique to a particular document as opposed to its file name, which is unique and non-empty.

Domain name

The unique name that identifies an Internet site. A given machine can have more than one domain name, but a given domain name is unique to only one machine unless alias or virtual domain names are used. Domain names can consist of many parts separated by a dot. What is located before the first dot is called the host, and what is located after the last dot is called the country code. Example: “elvin.bienlogic.com”.

Error codes

An error code is generated when a request from your browser cannot be fulfilled. The two main types of error codes are server error (the 500 series) codes and client (browser error codes-the 400 series). See Appendix B for a complete list.

Error messages

SurfReport will generate an error message to your screen when there was a problem running your report. This could be caused by errors with configuration, permissions, missing files, etc. SurfReport will attempt to identify the problem.

FAQs

Frequently Asked Questions. They consist of a set of commonly asked questions along with their answers.



Part 6: Glossary of Terms and Definitions

File name

This is simply the name given to a file as it is stored on the host computer. Examples are “index.html” or “my_cat.gif”. It is normally different from a “Documents title”, which is a html tag and could be blank or duplicated in another file. File names are always unique within a folder. See “Document titles”

Filters

A SurfReport tool to help the user to remove the unwanted data from a report to receive the information that is pertinent to their needs and specifications. SurfReport offers filters for specific visitor(s) and/or file(s) reporting. In addition SurfReport offers visitor(s) and/or file(s) filters to narrow down the report to the user’s specifications.

Gifs

“Graphics Interface Format” is a compressed graphics file format used to allow graphics to be viewed on the web. This is the format in which most of the standard graphics for SurfReport are stored.

Graphic abilities

SurfReport 2.0 has the ability to create user customizable graphics in the form of bar, line, area, point and pie charts. A user friendly interface allows you to modify about any aspect of the charts, from selecting chart type to positioning tick marks. Graphics are currently available for the Daily Totals, Hourly Average and Top Visitors reports. The charts are generated on-the-fly, and therefore will reflect your last run of SurfReport, i.e. the most current “surfoutput.html”.



Part 6: Glossary of Terms and Definitions

Hit

Any file that is accessed from your web site by a remote user. These hits can include html documents, gif, jpgs or cgi calls. One “page” of information that you view with your web browser most likely consists of many elements. If you have a page with text and 10 graphics the page generates 11 hits. SurfReport counts all successful requests for each element of a page as a hit. It does not count unsuccessful requests, or files cached on the users machine as a hit.

Host

Any computer on a network that is a repository for services available to other computers on the network. It is quite common to have one host machine provide several services, such as World Wide Web and USENET. This host is referenced by its domain name.

HTML

“Hypertext Markup Language”.HTML is a the language in which documents for the web are written. SurfReport considers a single html document as a “page”.

HTTP

HTTP stands for “Hypertext Transfer Protocol”, and is the protocol for moving hypertext files across the Internet.This requires a HTTP client program on one end, and an HTTP server program on the other end. HTTP is the most important protocol used in the World Wide Web (WWW).



Part 6: Glossary of Terms and Definitions

HTTP errors

These kind of errors are usually generated when a request from your browser cannot be fulfilled as requested. For many of these errors and their causes, please consult Appendix A.

IP addresses

Internetwork Protocol Address or Internet Address. A unique number assigned by an Internet authority (NIC) that identifies a computer on the Internet. The number is four groups of numbers separated by three periods (dots), each between 0 and 255. 199.106.143.69 is an example. Most machines will also have one or more Domain Names that can be translated from the IP address to make it easier to remember. See Domain Names.

Log files

These are the files that contain all the access logs for your site. Each access log entry contains the information about who has requested data from your site, and what specifically they requested. SurfReport uses these log files for its input, and returns the output in a very informative and readable statistical report.

Log folder

This folder, or directory, will hold the log files. It is normally named “logs” or “stats”, but check your site for the name of the one on your site. There may be only one file contained in this folder, or many. The security of this folder should be taken very seriously. These log files contain pertinent and vital information about traffic.



Part 6: Glossary of Terms and Definitions

Operating system

The primary software of a computer which interprets the users commands and allows applications to interact with the computers hardware. SurfReport is available for most UNIX based operating system (such as SunOS, Solaris, IRIX etc.) as well as Windows NT.

Page

SurfReport uses “Page” as a hit to a file with “.htm” or “.html” as part of its file name. A page that can be accessed by a web browser is therefore a page. The page may contain calls to several gifs or other non-html files, which are recorded in the access log file. Therefore a page hit is sometimes a more useful measurement of accesses to your site.

Password

This is the password that you will enter on the SurfReport interface form that will allow you to run SurfReport. As a default it is set to “SurfReport”, but for security reasons you should change it to something else — something that you will remember.

Perl

Practical Extraction and Report Language. A “C” like programming language intended for pattern matching and textual manipulation which can be executed like a shell script. Much of SurfReport is written in Perl.

Permissions

A method of security on UNIX in which users, groups and individuals must be given certain privileges by the systems administrator to read, write and execute files and



Part 6: Glossary of Terms and Definitions

directories. Certain permissions must be given to the SurfReport code by the installer in order for it to operate correctly.

Referrers

Referrers are part of the information stored in the user agent portion of the Combined Log Format which indicates the URL of the user's location just before accessing a page on your site. This can sometimes be an indication that a link to your site is contained on this particular URL. Sometimes the referrer log file is stored separately from the access log file, in which case SurfReport will be unable to give you the referrer information. Configure your server to log information in the Combined Log Format if you wish SurfReport to report referrer information as part of the report.

Root

The path to a particular directory or file on a UNIX system, which starts from the very top directory or "root", and therefore is the full path to the file. It is opposed to the "relative path" which references a particular file or directory from your current location. This root directory is often depicted by the "/" character, e.g. "/usr/local/bin/perl" would be the root path.

Server

The server is a computer with software that provides service to client software running on other computers. A single server machine could have various server software packages running on it, thereby being able to provide many different services to clients on the network. Web software is produced by Netscape, Apache and NCSA among others.



Part 6: Glossary of Terms and Definitions

Site name (web site)

The name of a web site, which will normally be the domain name with the prefix “http://” . An example would be http://software.bienlogic.com to indicate Bien Logic’s web site. Many browsers will enable you to go directly to a machine’s web site without using the “http://” prefix.

Strout

The directory in SurfReport where your output documents will be kept including the final HTML document (by default “surfoutput.html”) which contains your report.

Substring

A character or set of characters which is part of a larger word or “string”. For example, “ten” is a substring of “tennis” , “surface_tension” and “often”.

SurfReport main folder

This is the top level where your SurfReport programs and directories will reside.

Tar

Stands for “Tape archival program”. A utility used to compress and uncompress files. A file that has been “tarred” will usually have a .tar extension.

Totals

The totals reported in Surf Report are of two specific kinds: Daily and entire period. Daily: The daily totals are the total number of hits, visits and pages occurring for any given day



Part 6: Glossary of Terms and Definitions

within the period specified for the report being generated. Period: The period totals are the grand total/sum of all the daily totals for hits, visits, and pages.

Unique Visitor

A unique visitor is contrasted here with just visitor, which is someone who has visited your site within a given time period. For example, if someone visited your site on two separate days, SurfReport would report this as 2 visits or visitors, but only one unique visitor. By default, SurfReport sets the time period for determining a new visitor at 40 minutes.

URL

Stands for “Uniform Resource Locator”. The naming hierarchy used to categorize web resources. They define the protocol used (i.e. ftp, http, etc.), the domain name of the server being used, and also the directory where the page/resource is located. An example is “http://www.domain_name.com/path/page.html”.

User name

The name of an individual user of a particular UNIX machine.

Virtual host

If you maintain your web site at another machine, such as an ISP, then your web site is “virtually hosted” at that machine. Normally, virtual host machines will store numerous websites for individuals and companies. There may be some restrictions as to cgi permissions, cpu time, etc. if your web site is a virtual host.



Part 6: Glossary of Terms and Definitions

Visitor

A user that has accessed your site within a given time period. By default, SurfReport will log a user as a new visitor if there is more than 40 minutes between any two hits to your site from that user. However, this user will only be logged as 1 Unique Visitor.

See Unique Visitors

Web site

See Site Name

Wildcards

The asterisk character “*” used as a substitute for another character or set of characters. As an example, you can represent the words Mycompany.com, Yourcompany.com and Hiscompany.com by one expression: *company.com. This can be a very powerful tool when used in conjunction with the filter capabilities of SurfReport.



APPENDIX A: SUMMARY OF HTTP ERRORS

Error codes are generated when a request from your browser cannot be fulfilled. The two main types of error codes are client browser error codes (the 400 series) and server error codes (the 500 series). The values of some of the more common numeric status codes are as follows:

400 Bad Request:

The request had bad syntax or was inherently impossible to be satisfied.

401 Unauthorized:

The authorization header is missing or unacceptable. The parameter to this message gives a specification of authorization schemes which are acceptable. The client should retry the request with a suitable Authorization header.

402 Payment Required:

The parameter to this message gives a specification of charging schemes acceptable. The client may retry the request with a suitable ChargeTo header.

403 Forbidden:

The request is for something forbidden. Authorization will not help.



Appendix A: Summary of HTTP Errors

404 Not Found:

The server has not found anything matching the URL given.

500 Internal Error:

The server encountered an unexpected condition which prevented it from fulfilling the request.

501 Not Implemented:

The server does not support the facility required.

502 Service Temporarily Overloaded:

The server cannot process the request due to a high load (whether HTTP servicing or other requests).The implication is that this is a temporary condition which may be alleviated at other times.

503 Gateway Timeout:

The response from an intermediate computer was not received quickly enough or within a time that the gateway was prepared to wait.



APPENDIX B: COUNTRY CODES

The following codes may appear in either upper or lower case.

AD	Andorra	BE	Belgium
AE	United Arab Emirates	BF	Burkina Faso
AF	Afghanistan	BG	Bulgaria
AG	Antigua and Barbuda	BH	Bahrain
AI	Anguilla	BI	Burundi
AL	Albania	BJ	Benin
AM	Armenia	BM	Bermuda
AN	Netherlands Antilles	BN	Brunei Darussalam
AO	Angola	BO	Bolivia
AQ	Antarctica	BR	Brazil
AR	Argentina	BS	Bahamas
AS	American Samoa	BT	Bhutan
AT	Austria	BV	Bouvet Island
AU	Australia	BW	Botswana
AW	Aruba	BY	Belarus
AZ	Azerbaijan	BZ	Belize
BA	Bosnia and Herzegovina	CA	Canada
BB	Barbados	CC	Cocos (Keeling) Islands
BD	Bangladesh	CF	Central African Republic



Appendix B: Country Codes

CG	Congo	ER	Eritrea
CH	Switzerland	ES	Spain
CI	Cote D'Ivoire (Ivory Coast)	ET	Ethiopia
CK	Cook Islands	FI	Finland
CL	Chile	FJ	Fiji
CM	Cameroon	FK	Falkland Islands (Malvinas)
CN	China	FM	Micronesia
CO	Colombia	FO	Faroe Islands
CR	Costa Rica	FR	France
CS	Czechoslovakia (former)	FX	France, Metropolitan
CU	Cuba	GA	Gabon
CV	Cape Verde	GB	Great Britain (UK)
CX	Christmas Island	GD	Grenada
CY	Cyprus	GE	Georgia
CZ	Czech Republic	GF	French Guiana
DE	Germany	GH	Ghana
DJ	Djibouti	GI	Gibraltar
DK	Denmark	GL	Greenland
DM	Dominica	GM	Gambia
DO	Dominican Republic	GN	Guinea
DZ	Algeria	GP	Guadeloupe
EC	Ecuador	GQ	Equatorial Guinea
EE	Estonia	GR	Greece
EG	Egypt	GS	S. Georgia and S. Sandwich Isls.
EH	Western Sahara	GT	Guatemala



Appendix B: Country Codes

GU	Guam	KM	Comoros
GW	Guinea-Bissau	KN	Saint Kitts and Nevis
GY	Guyana	KP	Korea (North)
HK	Hong Kong	KR	Korea (South)
HM	Heard and McDonald Islands	KW	Kuwait
HN	Honduras	KY	Cayman Islands
HR	Croatia (Hrvatska)	KZ	Kazakhstan
HT	Haiti	LA	Laos
HU	Hungary	LB	Lebanon
ID	Indonesia	LC	Saint Lucia
IE	Ireland	LI	Liechtenstein
IL	Israel	LK	Sri Lanka
IN	India	LR	Liberia
IO	British Indian Ocean Territory	LS	Lesotho
IQ	Iraq	LT	Lithuania
IR	Iran	LU	Luxembourg
IS	Iceland	LV	Latvia
IT	Italy	LY	Libya
JM	Jamaica	MA	Morocco
JO	Jordan	MC	Monaco
JP	Japan	MD	Moldova
KE	Kenya	MG	Madagascar
KG	Kyrgyzstan	MH	Marshall Islands
KH	Cambodia	MK	Macedonia
KI	Kiribati	ML	Mali



Appendix B: Country Codes

MM	Myanmar	NU	Niue
MN	Mongolia	NZ	New Zealand (Aotearoa)
MO	Macau	OM	Oman
MP	Northern Mariana Islands	PA	Panama
MQ	Martinique	PE	Peru
MR	Mauritania	PF	French Polynesia
MS	Montserrat	PG	Papua New Guinea
MT	Malta	PH	Philippines
MU	Mauritius	PK	Pakistan
MV	Maldives	PL	Poland
MW	Malawi	PM	St. Pierre and Miquelon
MX	Mexico	PN	Pitcairn
MY	Malaysia	PR	Puerto Rico
MZ	Mozambique	PT	Portugal
NA	Namibia	PW	Palau
NC	New Caledonia	PY	Paraguay
NE	Niger	QA	Qatar
NF	Norfolk Island	RE	Reunion
NG	Nigeria	RO	Romania
NI	Nicaragua	RU	Russian Federation
NL	Netherlands	RW	Rwanda
NO	Norway	SA	Saudi Arabia
NP	Nepal	Sb	Solomon Islands
NR	Nauru	SC	Seychelles
NT	Neutral Zone	SD	Sudan



Appendix B: Country Codes

SE	Sweden	TO	Tonga
SG	Singapore	TP	East Timor
SH	St. Helena	TR	Turkey
SI	Slovenia	TT	Trinidad and Tobago
SJ	Svalbard and Jan Mayen Islands	TV	Tuvalu
SK	Slovak Republic	TW	Taiwan
SL	Sierra Leone	TZ	Tanzania
SM	San Marino	UA	Ukraine
SN	Senegal	UG	Uganda
SO	Somalia	UK	United Kingdom
SR	Suriname	UM	US Minor Outlying Islands
ST	Sao Tome and Principe	US	United States
SU	USSR (former)	UY	Uruguay
SV	El Salvador	UZ	Uzbekistan
SY	Syria	VA	Vatican City State (Holy See)
SZ	Swaziland	VC	Saint Vincent and the Grenadines
TC	Turks and Caicos Islands	VE	Venezuela
TD	Chad	VG	Virgin Islands (British)
TF	French Southern Territories	VI	Virgin Islands (U.S.)
TG	Togo	VN	Viet Nam
TH	Thailand	VU	Vanuatu
TJ	Tajikistan	WF	Wallis and Futuna Islands
TK	Tokelau	WS	Samoa
TM	Turkmenistan	YE	Yemen
TN	Tunisia	YT	Mayotte



Appendix B: Country Codes

YU Yugoslavia
ZA South Africa
ZM Zambia
ZR Zaire
ZW Zimbabwe

COM US Commercial
EDU US Educational
GOV US Government
INT International
MIL US Military
NET Network
ORG Non-Profit Organization
ARPA Old style Arpanet
NATO Nato field



INDEX

Access Log	50	Configure Form Interface	13
Access Log Files	29	Configure surfreport.cgi	12
Activation Code	50	Countries	52
Advanced Chart Options	47	Country Codes	52, 65
Advanced Tasks	33	Custom settings	53
Averages	50	Daily Averages	23
Axes	48	Daily Totals	22
Browsers	28, 51	Defaults	53
Browser/Platform Preferences	37	Deliver my SurfReport	21
CGI	51	Delivery Methods	53
Chart Type	43	Dimensions	44
Charts	42	Directory Structure	53
CheckList Before You Install	6	Document Titles	54
Client	51	Domain Name	54
Colors	46	Error Codes	54
Combined Log Format	51	Error Messages	54
Common Log Format	51	FAQs	54
Company	52	File Filter	28, 34
Complete the Installation	10	File Name	55
Compression	52	Filters	55



Index of Contents

Generate a SurfReport	30	Page	58
Gifs	55	Password	29, 58
Glossary	50	Period Totals	23
Graphic Abilities	55	Perl	58
Hit	56	Permissions	7, 58
Host	56	Process Files for	29
Hourly Averages	23	Referrers	59
HTML	56	Removing SurfReport	15
HTTP	56	Reverse main axis direction	49
HTTP Errors	57	Rootpath	59
In-Depth Analysis	31	Running the installer	8
Installing SurfReport	6	Server	59
Introduction	1	Settings	21
IP Addresses	57	Site name (web site)	60
Log Files	57	Specific File Report	27
Log Folder	57	Specific Visitor Report	27
Logging Only Page Hits	36	Strout	60
Many Virtual Hosts	41	Substring	60
Multiple Copies of SR	41	Summary of HTTP Errors	29, 63
Name my SurfReport	22	SurfReport Main Folder	60
Operating System	58	Tar	60
Operating Systems	28	Titles and Labels	47
Orientation	44	Top Companies/Domains	25
Other Documentation	14	Top Countries	25
Overall Architectural Flow	32	Top Files Accessed	26



Index of Contents

Top Referrer Pages	26
Top Visitors	24
Totals	60
Unattended Basis	40
Uncompressing surfreport.tar	8
Unique Visitor	61
UNIX Installation	6
URL	61
User Name	61
Using Filters	34
Using Your Default Settings	30
Virtual Host	61
Visitor	62
Visitor Filter	27, 34
Web Site	62
Wildcards	62
Windows NT Installation	16

