



Product Test Report

May 2007

BrightCloud Master Database URL Database Quality Study

BrightCloud Master Database URL Database Quality Study

Vendor Details

Vendor Name:
BrightCloud Inc

Vendor Address:
3252 Holiday Ct, Suite 112A, La Jolla, CA 92037, USA

Vendor Telephone Number: +1 (858) 539-3415

Product:
BrightCloud Master Database

Test Laboratory Details

Test Laboratory Name:
West Coast Labs

Test Laboratory Address:
William Knox House, Britannic Way, Llandarcy, Swansea, UK, SA10 6EL

Test Laboratory Telephone Number: +44 (0) 1792 324000

Date: 16th May 2007

Issue: 1.0

Author: Danny Yarde

Contact Points for Technical Queries on the Test Report

Contact Name: Matt Garrad

Contact Telephone Number: +44 (0) 1792 324000

BrightCloud Master Database URL Database Quality Study

Contents

Executive Summary	4
Web Filtering Overview	6
Vendor URL Databases and the Size of the Internet	8
URL Coverage tests	11
URL Accuracy tests	17
West Coast labs Test Summary	22
Appendix A – Test Set Up	23
Appendix B – Vendor Summaries	25

BrightCloud Master Database URL Database Quality Study

Executive Summary

This report compares the enterprise class URL databases of four vendors: BrightCloud, the sponsors of the study, Secure Computing's SmartFilter, SurfControl, and Websense, (the Websense acquisition of SurfControl was announced during the course of this study, but both databases were tested separately).

While SmartFilter, Websense and SurfControl have web filtering product implementations that include protocol management, policy interfaces, directory integration, and reporting systems, these features were not evaluated. The BrightCloud hosted service does not include those features- BrightCloud licenses its URL database to security appliance vendors, enterprise end point security vendors, wireless, and hosted security providers. Performance (that is, the speed with which the vendors were able to respond to a large number of URL requests for category information) was also not tested.

Specifically, the quality of the URL database, including both coverage & accuracy, was measured for each vendor. URL coverage is a simple test, whereby each implementation was asked whether it has the requested URL in its database, regardless of category. If the vendor responded with a category other than "Uncategorized," that response was counted in the vendor's favour.

BrightCloud Master Database URL Database Quality Study

Executive Summary

The Accuracy test was based on URLs randomly selected by West Coast Labs in four categories: Adult, Gambling, Games, and Shopping. Each vendor's solution was queried for the category of the URL, and a correct categorization was counted in the vendor's favour. It should be noted that both BrightCloud and SmartFilter support multiple categorizations for a single URL. As long as one of the returned categories matched the category defined by West Coast Labs, the response was counted in the vendor's favour.

The results of the test are that BrightCloud's URL database outperformed all other vendors in both coverage and accuracy- in some cases, by a significant margin.

Websense was second in almost every test, although SmartFilter finished a close third to Websense, and surpassed both Websense and SurfControl in the broadest coverage test. While BrightCloud was also the most accurate vendor in the four categories tested, the difference between the other three vendors was not significant in most of the four categories tested. The details of each test are provided below.

BrightCloud Master Database URL Database Quality Study

Web Filtering Overview

Enterprise Web Filtering software, also known as content-filtering software, is a term for software designed and optimized for controlling what internet content an enterprise end user is permitted to access, especially when it is used to restrict material delivered over the Web. Web filtering is critical to enterprises for four main reasons:

- 1) **Security**- protect the enterprise user from websites that are phishing sites or host spyware, Trojans, or other malicious code;
- 2) **Productivity**- help ensure the user is not distracted by accessing non-work related web sites such as gambling, game site, auctions, sports, or shopping;
- 3) **Legal liability**- contribute to a non-discriminatory workplace by restricting access to adult websites, or websites that promote hate or violence, as well as restrict access to websites that participate in illegal file sharing or other types of copyright infringement;
- 4) **IT Resource management**- reduce the detrimental impact on a company's computing resources inherent in access to certain classes of websites, such as non-work related streaming media websites, non-work related instant messaging or VoIP clients, or the download of unwanted shareware and freeware.

BrightCloud Master Database URL Database Quality Study

Web Filtering Overview

There are three web filtering vendors that together represent the majority of enterprise user market share: Websense, SurfControl (at the time of this writing the Websense acquisition of SurfControl had just been announced), and SmartFilter.

Web filtering software performs several different but inter-related tasks: URL categorization, policy implementation and reporting. Each vendor in this study integrates with several of the most commonly used web browsing packages, both on firewalls, servers and desktops, including Cisco PIX, Microsoft ISA Server, Checkpoint Firewall-1 and Squid (an open source proxy). Each vendor tested has between 50 and 100 categories, with Websense having the most categories at more than 90. The URL databases are updated regularly, and all vendors support real time updates.

BrightCloud Master Database URL Database Quality Study

Vendor URL Databases and the Size of the Internet

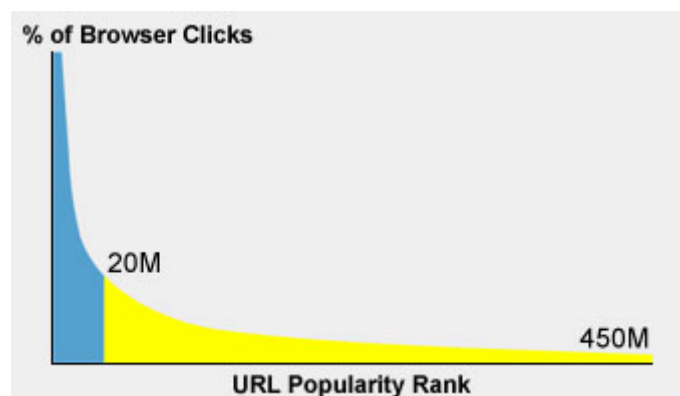
As noted above, this report focused solely on the vendors' URL databases. At the time of this report, Websense and SurfControl claimed to have 20 million sites in their database, and SmartFilter 15 million. BrightCloud, with the largest URL database of any enterprise web filtering vendor we are aware of, claims to have 40 million sites, although that number is growing rapidly (see www.brightcloud.com for their current size of the BrightCloud Master Database).

The internet itself, of course, is much larger. As of the date of this report, Google (www.google.com) says there are approximately 14 billion web pages; Netcraft (www.netcraft.com) has identified 116 million domains (defined as a top level domain plus a unique name, such as "domain.com"); and the Internet Systems Consortium (www.isc.org) has identified 455 million unique authorities (defined as sub domains, such as "sub.domain.com"). Further, there is both considerable growth and churn in websites. Netcraft reports that the numbers of authorities are growing at a rate of 25-30% per year. DomainTools (www.domaintools.com) reports there are an average of 1 million new authorities created each day, and almost one million domains that go defunct each day, with an overall net gain commensurate with the Netcraft growth rate projections.

BrightCloud Master Database URL Database Quality Study

Vendor URL Databases and the Size of the Internet

BrightCloud claims that enterprise user browsing behaviour is best depicted by a “power law” curve, also known as the “Long Tail.” Power laws in supply and demand curves were popularized in Chris Anderson’s book The Long Tail: Why the Future of Business is Selling Less of More (2006). In essence the Long Tail applies to web filtering due to the phenomenon that 50% of enterprise user browsing behaviour occurs among the most popular sites, but the remaining web site visits are randomly distributed across the much larger number of less popular websites. BrightCloud’s research indicates that approximately 50% of the users website visits occur among the top 20 million most popular authorities. The remaining 50% of visits are spread out randomly across the remaining 430 million authorities- presenting a challenge to web filtering vendors who cannot anticipate which of those sites will be visited. A power law, or Long Tail curve depicting enterprise user browsing behaviour is provided in the graph below:



BrightCloud Master Database URL Database Quality Study

Vendor URL Databases and the Size of the Internet

This has significant implications for web filtering vendors- a URL database with 20 million authorities- or 5% of the internet's total authorities- cannot deliver adequate protection for enterprise customers. And locally delivering and updating a URL database that provides adequate coverage would be technically problematic, expensive, and slow because of the size of the database and the bandwidth required. Over time, this problem will compound given the growth and churn of the internet. For example, if the growth rate of the internet continues at its current pace, in 2012 there will be over 250 million domains, 1 billion authorities, and 30 billion web pages.

BrightCloud Master Database URL Database Quality Study

West Coast Labs URL Coverage Test

The coverage test was designed to answer a very simple question- was a URL in a given vendor's URL database. Each vendor's solution was queried with a URL (specifically, an authority), and if a categorized response was received, the vendor got credit for having the URL in their URL database.

The tests were performed on the four different vendors' product databases, using three different URL lists.

The URL lists were taken from:

Alexa Top 100,000 URLs- Alexa is a web service offered by Amazon (www.alexa.com). Its rankings are created by analyzing the web usage of millions of users with the Alexa Toolbar installed. The calculations are based on 'Reach' and 'Page Views'. Reach is the number of users visiting a specific site. Page Views is the number of pages on that site viewed by each unique user. The results are released daily and show changes in popularity of the sites compared with the previous release.

BrightCloud Master Database URL Database Quality Study

West Coast Labs URL Coverage Test

Ranking.com – Ranking.com (www.ranking.com) provides a ranked list of the top 1.5 million authorities that is also updated frequently.

Ranking.com uses a sample of web users, currently over 200,000, with their market research software installed, to gauge the popularity of sites based on a sampling methodology as a representative subset of the overall web surfing population. Ranking.com provides ranking of a website both by visits and page rank, although the visits rank was used for the purpose of this report.

Top 20 Million URLs by PageRank – PageRank is a system of ranking that was created by one of the founders of Google, Larry Page, and forms the basis of all Google search tools. It works using an algorithm that measures pages by the number of hyperlinks to a specific site weighted by the ‘importance’ of the referring site. The page rank data used in this report were the top 20 million authorities by page rank, and provided by BrightCloud.

Coverage Test Results

The first test took all of the Alexa Top 100,000 URLs and 100,000 randomly selected URLs from Ranking.com and the Top 20M by Page Rank and ran them against the four filters to see how many URLs from the lists were present in the vendors’ respective databases.

BrightCloud Master Database URL Database Quality Study

West Coast Labs URL Coverage Test

The results are provided in the tables below, with the highest scoring vendor results highlighted in yellow:

URL Coverage

	Total URLs	BrightCloud	Websense	Smartfilter	Surfcontrol
Alexa 100k	99,663	93,742	85,625	80,094	74,379
Ranking.com 1.5M	99,998	89,735	58,818	55,340	49,042
Page Rank 20M	99,994	91,368	33,354	36,879	33,272

Percentage Coverage

	Total URLs	BrightCloud	Websense	Smartfilter	Surfcontrol
Alexa 100k	100.0%	94.1%	85.9%	80.4%	74.6%
Ranking.com 1.5M	100.0%	89.7%	58.8%	55.3%	49.0%
Page Rank 20M	100.0%	91.4%	33.4%	36.9%	33.3%

As noted in the tables above, BrightCloud scored the highest across each of the data sets above. Because the Page Rank 20M dataset was provided by BrightCloud, you would expect them to score very high coverage rates for that data set. However, you would also expect much higher coverage rates than one in three URLs for the market share leading vendors as well, especially as the URLs selected were the top ranking URLs by page rank.

BrightCloud Master Database URL Database Quality Study

West Coast Labs URL Coverage Test

The poor coverage performance of the other three vendors across the Page Rank Top 20M provides insufficient information to draw conclusions about the size of their respective URL databases- they could very well have a large number of URLs not contained in the Page Rank dataset, for example. Unfortunately, that also implies the URLs in those vendors' URL databases would not represent some of the most popular URLs, resulting in reduced protection levels for enterprise customers.

As for coverage of the Long Tail- that is, URL categorization across the internet's 450 million authorities- it is clear that each vendor will need to significantly expand the size of their respective URL database to provide their enterprise customers with adequate coverage.

BrightCloud's strong results in the 20 million URL dataset show that they may well be the vendor furthest down the path of providing coverage across the Long Tail, but even with a 40 million URL database, a significant portion of URLs of the Long Tail are not covered.

An interesting view of the data provided in the tables above provides some insight into the various vendors' coverage. The charts below represent the various vendors' coverage across the ranked URLs.

BrightCloud Master Database URL Database Quality Study

West Coast Labs URL Coverage Test

Alexa Coverage by URL Rank

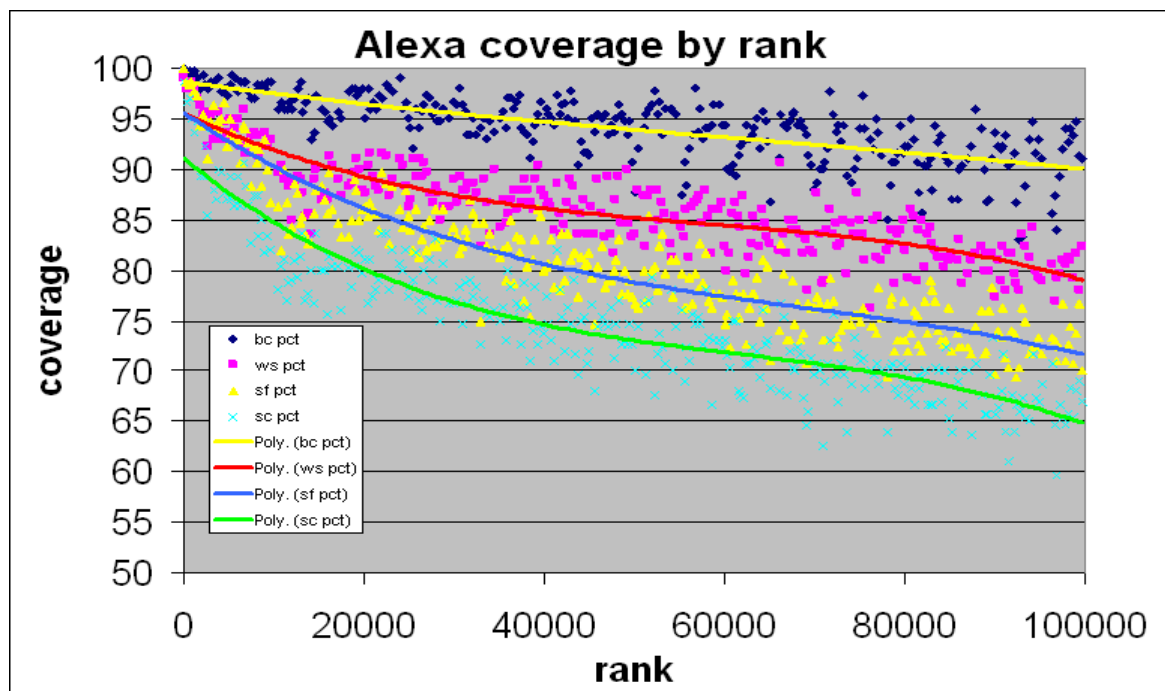
This chart represents the coverage of each vendor across the Alexa 100,000 URL dataset. The dots on the chart represent the URLs covered at various points across the dataset, and the solid lines represent the polynomial fit to the data representing the average coverage of each vendor. The vendor's data are represented as follows:

BrightCloud: The top yellow line representing best fit of blue diamonds;

Websense: The red line representing the fit of the magenta squares;

SmartFilter: The light blue line, representing best fit of yellow triangles;

SurfControl: The bright green line representing best fit of light blue "x's".



BrightCloud Master Database URL Database Quality Study

West Coast Labs URL Coverage Test

Ranking.com Coverage by URL Rank

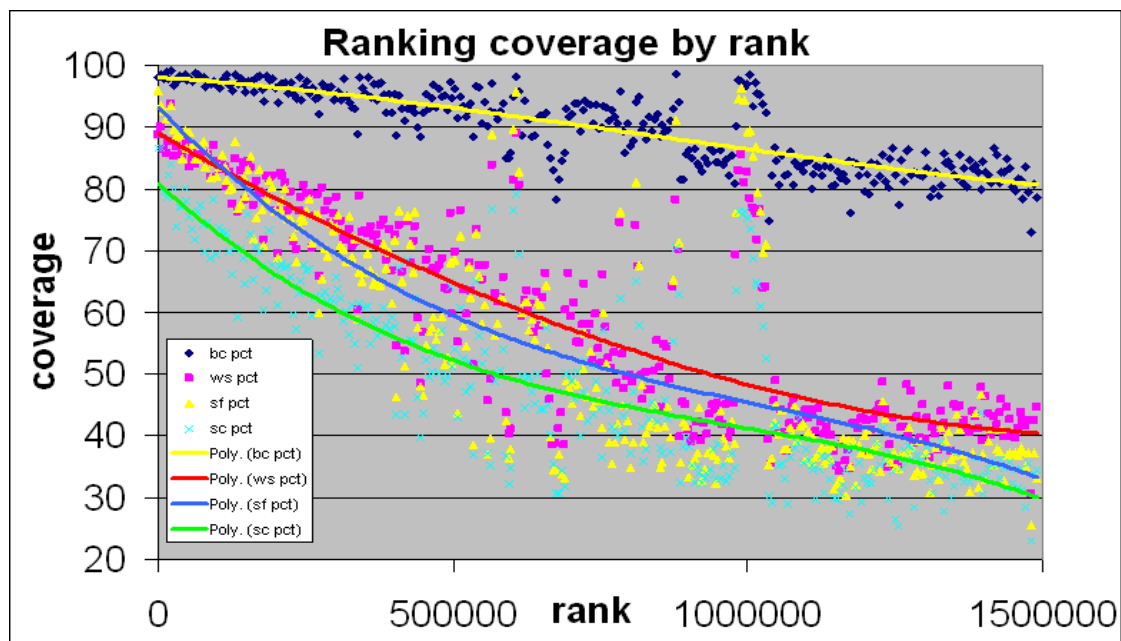
This chart represents the coverage of each vendor across the Ranking.com 1,500,000 URL dataset. The dots represent the URLs covered at various points across the dataset, and the solid lines represent the average coverage of each vendor. The vendor's data are represented as follows:

BrightCloud: The top yellow line representing best fit of blue diamonds;

Websense: The red line representing the fit of the magenta squares;

SmartFilter: The light blue line, representing best fit of yellow triangles;

SurfControl: The bright green line representing best fit of light blue "x"s".



BrightCloud Master Database URL Database Quality Study

West Coast Labs URL Categorization Accuracy Test

The second set of tests checked for accuracy in four key categories:

Adult

Gambling

Games

Shopping

The goal of the test was to determine how accurate the filtering was among the respective vendors within these key categories. These categories were selected because they represent some of the most important categories over which enterprises would like to manage employee internet access. Security related categories are also important to enterprises, but go beyond the scope of this study.

400 URLs for each category were randomly selected from the Alexa 100k list by West Coast Labs (the sites were chosen for inclusion using a 'blind' SQL query so that it was not known which vendor had classified the URL into the given category). 400 URLs were selected per category to provide very tight margins of error for each category. The margin of error for the tests conducted here at the 95% confidence level were less than 2%. Each URL was manually categorized by West Coast Labs to validate it was in the specified category.

BrightCloud Master Database URL Database Quality Study

West Coast Labs URL Categorization Accuracy Test

The West Coast Labs classification of a URL was considered to be the authoritative classification.

BrightCloud and SmartFilter are able to define multiple categories for a given URL, and were given credit in these tests for a correct categorization, regardless of what the other categorizations might have been. Additionally, the vendors have several categories that might fit into the four categories tested, and were given credit for URLs with those categorizations. For example, SmartFilter has categories for Nudity, Pornography, Provocative Attire and Sexual Materials. A URL having any of these categories would all count as a correct categorization in the Adult test.

Once the site had been manually categorized, the specific vendor's classification for a URL was determined by running it against the vendor's URL database. Two values were captured, and reported in the tables below. The values in the "In Category" column represent agreement between the WCL classification of the URL and the vendor's classification. The values in the "Not in Category" column represent URLs that the WCL said were in the tested category, but the vendor did not.

BrightCloud Master Database URL Database Quality Study

West Coast Labs URL Categorization Accuracy Test

The results are provided in the tables below:

Accuracy Test Raw Data

	URLs Tested	BrightCloud		SurfControl		Smartfilter		Websense	
		In Category	Not in Category	In Category	Not in Category	In Category	Not in Category	In Category	Not in Category
Adult	414	405	9	395	18	389	23	396	17
Gambling	346	293	36	286	59	229	116	283	63
Games	341	323	10	221	117	242	97	281	59
Shopping	385	289	76	183	184	284	85	268	100

For each category, the vendor with the highest number of agreements for the "In Category" column, and the fewest disagreements in the "Not in Category" column, is highlighted in yellow.

The tables here represent the same information presented as percentages:

Correctly Classified in Category				
Category	BrightCloud	SurfControl	Smartfilter	Websense
Adult	97.8%	95.4%	94.0%	95.7%
Gambling	84.7%	82.7%	66.2%	81.8%
Games	94.7%	64.8%	71.0%	82.4%
Shopping	75.1%	47.5%	73.8%	69.6%

BrightCloud Master Database URL Database Quality Study

West Coast Labs URL Categorization Accuracy Test

Incorrectly Classified as not in Category

Category	BrightCloud	SurfControl	Smartfilter	Websense
Adult	2.2%	4.3%	5.6%	4.1%
Gambling	10.4%	17.1%	33.5%	18.2%
Games	2.9%	34.3%	28.4%	17.3%
Shopping	19.7%	47.8%	22.1%	26.0%

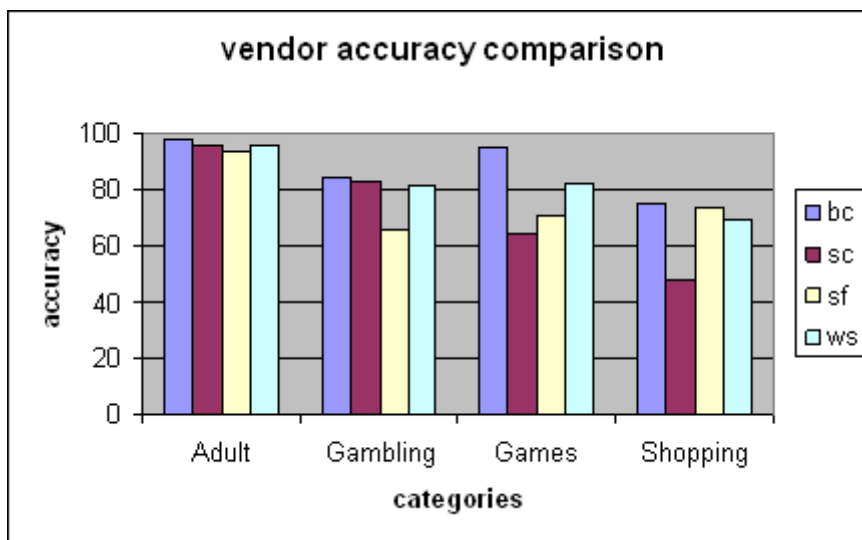
Again, in the two tables above, the vendor with the highest degree of agreement with the WCL categorization is highlighted in yellow. The margin of error at a 95% confidence level is less than 2% for each of the categories.

BrightCloud's accuracy in each category is superior to those of the other vendors. Although BrightCloud has a higher accuracy score for the Adult category, the accuracy results are generally consistent among the vendors. The greatest disparity between BrightCloud's results and those of the other vendors occurs in the Games and Shopping categories, both in terms of accurately categorizing a URL and incorrectly claiming that a URL is not in those categories.

BrightCloud Master Database URL Database Quality Study

West Coast Labs URL Categorization Accuracy Test

A comparison chart summarizing the data above is provided here:



BrightCloud Master Database URL Database Quality Study

West Coast Labs URL Database Test Summary

The tests conducted for this report were coverage and accuracy. The URL databases for BrightCloud, Websense, SurfControl, and SmartFilter were tested. Every effort was made to ensure that the testing set up was consistent across each vendor.

BrightCloud's test results were impressive. BrightCloud scored highest on each of the tests- in some cases by a significant margin.

WCL believes the coverage tests with URLs taken from Alexa and Ranking.com provide the most reliable comparison between the vendors coverage capabilities. The larger dataset of Top 20 million URLs by Page Rank is data provided by BrightCloud, and thus it is not surprising that they have such a high percentage coverage rate. However, that test does illustrate that BrightCloud has a very comprehensive set of URLs in the BrightCloud Master Database.

However, it is clear from this report that each vendor will need to do considerable work to expand the coverage of their dataset to encompass even a majority of the 450 million authorities of the Long Tail. It is reasonable to suggest, based on the results of the test conducted here, that BrightCloud is furthest down that path, but even their 40 million URL database is inadequate to extend protection to their enterprise customers across the internet's Long Tail.

BrightCloud Master Database URL Database Quality Study

Appendix A - Test Set Up

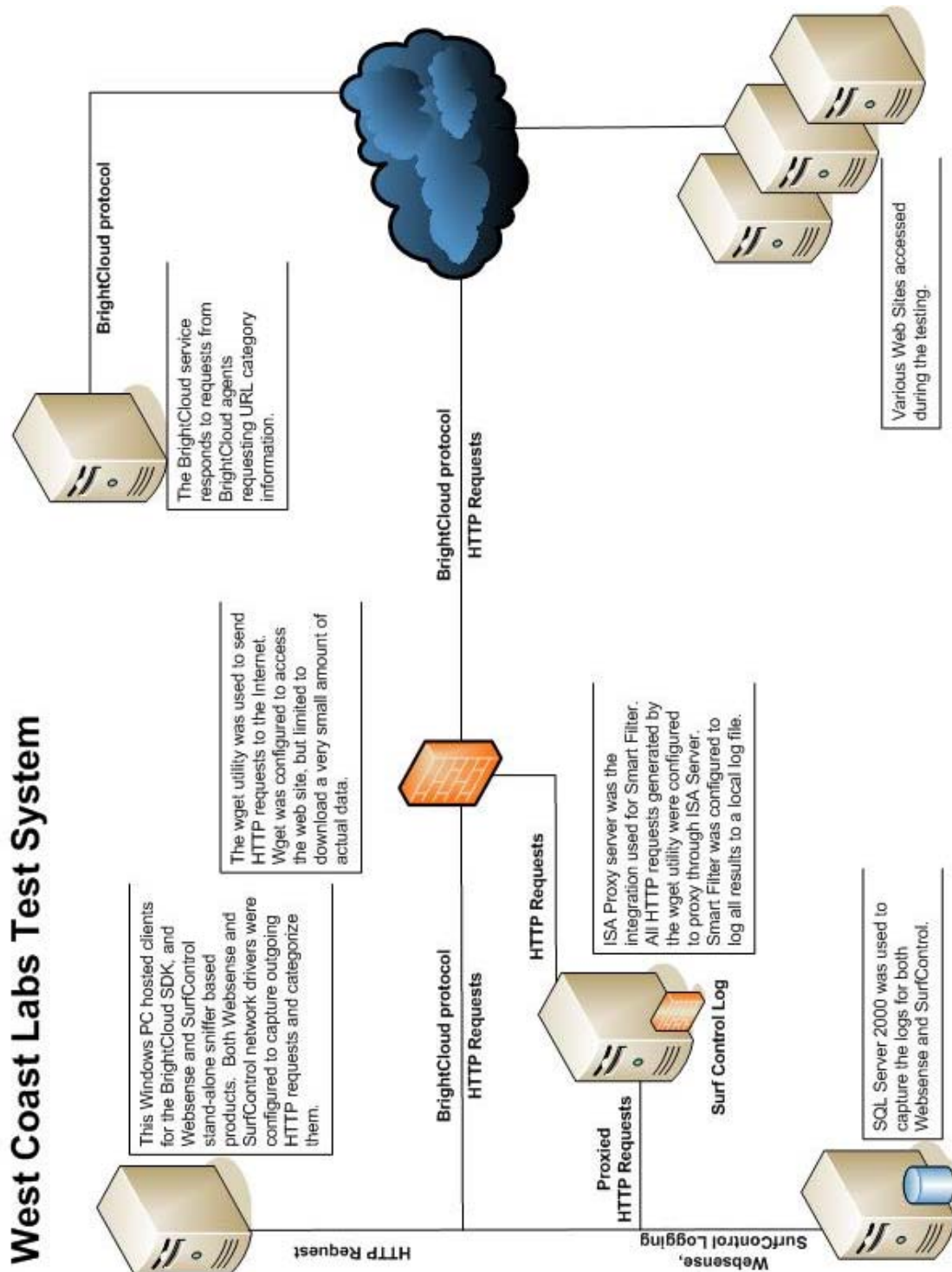
The test configuration for each vendor is described in the diagram below. The test was designed to test only the URL databases of the respective vendors, although each vendor's reporting capabilities were utilized to capture the output data.

BrightCloud was tested via a "BrightCloud Agent" client built from its Software Development Kit. The BrightCloud Agent logged its results locally to a file, from which the relevant data was extracted.

Websense and SurfControl were utilized in their "sniffer" modes, and their output data written to a Microsoft SQL 2000 database.

SmartFilter was accessed via a Microsoft ISA Server integration, and its output data written to a local file. See below:

BrightCloud Master Database URL Database Quality Study



BrightCloud Master Database URL Database Quality Study

Appendix B - Vendor Summaries

BrightCloud, Inc.

BrightCloud, Inc. provides security vendors with hosted security services that extend the value propositions of web filtering across the internet's Long Tail – the 450 million Websites visited by enterprise customers – which includes both the most popular sites as well as those sites less frequently visited, but visited nonetheless. Target customers include enterprise security vendors in the security appliance, enterprise security application, end point security, enterprise security application, wireless and xSP spaces. BrightCloud uses the latest in machine learning algorithms, as well as human classification, to build the largest and most accurate URL database available. By integrating the BrightCloud services, OEMs can provide their end customers with a richer, multi-layered security solution.

Secure Computing Corporation (SmartFilter URL database)

Secure Computing® is a global leader in Enterprise Gateway Security Solutions. Powered by our TrustedSource™ technology, our best-of-breed portfolio of solutions provides Web Gateway, Messaging Gateway, and Network Gateway security, as well as Identity and Access Management. Secure Computing is proud to be the security solutions provider to many of the most mission-critical and sensitive environments in the world.

BrightCloud Master Database URL Database Quality Study

Appendix B - Vendor Summaries

SurfControl PLC

SurfControl protects organizations with multiple layers of threat protection that filter inbound, outbound and internal Internet traffic. SurfControl's Enterprise Protection Suite protects multiple threat vulnerability points - Web, E-mail, IM, P2P and Mobile desktops - and is supported by SurfControl's worldwide Adaptive Threat Intelligence Service to provide customers with early detection of emerging threats, real-time updates and continuous protection.

Websense, Inc.

Founded in 1994, Websense, Inc. (NASDAQ: WBSN) is the global leader of web filtering and a premier provider of web and desktop security software. Recognized as one of Forbes Magazine's 2004 "Top 25 Technology Companies", Websense has 24.1 million seats under subscription worldwide across all industries, in small organizations and large multinational corporations. Websense products increase employee internet productivity and secure organizations from emerging internet threats by providing a proactive critical security component that complements traditional security solutions. Only Websense delivers flexible, integrated policy enforcement at the internet gateway, on the network and at the desktop.

BrightCloud Master Database URL Database Quality Study

West Coast Labs Disclaimer

While West Coast Labs is dedicated to ensuring the highest standard of security product testing in the industry, it is not always possible within the scope of any given test to completely and exhaustively validate every variation of the security capabilities and / or functionality of any particular product tested and / or guarantee that any particular product tested is fit for any given purpose .

Therefore, the test results published within any given report should not be taken and accepted in isolation. Potential customers interested in deploying any particular product tested by West Coast Labs are recommended to seek further confirmation that the said product will meet their individual requirements, technical infrastructure and specific security considerations.

All test results represent a snapshot of security capability at one point in time and are not a guarantee of future product effectiveness and security capability. West Coast Labs provide test results for any particular product tested, most relevant at the time of testing and within the specified scope of testing and relative to the specific test hardware, software, equipment, infrastructure, configurations and tools used during the specific test process.

West Coast Labs is unable to directly endorse or certify the overall worthiness and reliability of any particular product tested for any given situation or deployment.

Revision History

Issue	Description of Changes	Date Issued
1.0	BrightCloud Master Database URL filtering Effectiveness Study	16 th May 2007

westcoast labs

US SALES

T +1 925 833 6935

EUROPE SALES

T +44 1792 324000

GLOBAL HEADQUARTERS

West Coast Labs
William Knox House
Britannic Way
Llandarcy
Swansea SA10 6EL, UK

T +44 1792 324000

F +44 1792 324001

E info@westcoast.com

W www.westcoastlabs.com