



MessageLabs®

Be certain

# Web Security Services v2

## Technical Overview

January 2007

# Table of Contents

1	Introduction	4
2	Infrastructure	5
2.1	Data Centers	5
2.2	How Do MessageLabs Web Services v2 Work?	5
2.3	Global Load Balancing Model	5
2.3.1	Technical Overview	6
2.4	MessageLabs Web Services v2 Data Centre Architecture	7
3	The Web Anti-Spyware and Web Anti-Virus Service v2	8
3.1	Scanning Methodology	8
3.2	Commercial Technologies Employed	8
3.3	Skeptic Technologies Employed	8
4	The Web URL Filtering Service v2	9
4.1.1	When	9
4.1.2	Who	9
4.1.3	Where	9
4.1.4	What	10
4.2	Policy Processing	10
4.3	The Policy Stack	10
4.3.1	Exception handling policy rules	10
4.3.2	General Allow or Block policy rules	10
4.3.3	Acceptable Use Policy (AUP) rules	11
5	Client Site Proxy	12
6	Custom Groups	13
7	Group Synchronization	14
7.1	How the Synchronization Works	14



## **1 Introduction**

This overview document provides a technical description of the infrastructure and processes making up the MessageLabs Web Security Services v2.

## 2 Infrastructure

MessageLabs Web Security Services v2 have been designed from the ground up to be a resilient, scalable solution. This is critically important to ensure the best possible performance for companies of all sizes, allowing them secure and reliable access to the internet to support their business objectives. The MessageLabs solution ensures that the most available and effective data center is used for the transmission of traffic to their users without needing to make client, network or routing configuration changes.

### 2.1 Data Centers

Multiple high performance and secure data centers, located in various locations globally, provide bandwidth and 'active/active' resilience for the service.

These facilities have biometric entry security, redundant power feeds from the grid, redundant generators, redundant air conditioning systems and hundreds of security cameras that are monitored by 24/7 security staff. Within the data centers, MessageLabs locates equipment in separate, secure suites where only MessageLabs equipment is allowed. Access to these areas is limited to authorized MessageLabs staff and data center staff acting under our instructions.

MessageLabs will be adding additional data centers of this caliber as the user base increases to ensure a positive experience and continued low latency.

The addition of these data centers, or even additional resources within current data centers, will be transparent to the client. The infrastructure of MessageLabs Web Security Services v2 ensures that whenever hardware is added to the environment it becomes part of the globally load balanced environment rather than a unique entity which requires configuration changes in order to become part of the overall service.

### 2.2 How Do MessageLabs Web Services v2 Work?

The services operate as an explicit/non-transparent Proxy. An explicit proxy accepts requests from user machines and services them on the users' behalf. When communicating with a proxy, the device connecting to the proxy sends the protocol to be used (e.g.: HTTP/1.0) and the URL to be downloaded in the form "GET <http://www.messagelabs.com/>". This is then actioned by the proxy, and the results passed back to the user's machine.

The service will process HTTP and FTP over HTTP requests.

HTTPS / SSL communications are also proxied via the service to ensure a seamless communication for sites which use SSL for authentication before reverting to clear text communications. If the services did not proxy SSL in this way, the target web server (such as a webmail service) would see a user session authenticated from one IP address, then once authenticated the next request appears from a completely new source – the MessageLabs Web Security Services. This would cause the connection to fail, on the basis that the MessageLabs Web Security Services are not originating from the same location as the SSL authenticated session and being treated therefore as a spoof or hijack.

To avoid this scenario, the MessageLabs Web Security Services passes SSL sessions, but does not apply any filtering to them. Other communications that are passed without interception include, but are not limited to, P2P networking, streaming audio and streaming video.

To direct the web requests to the MessageLabs service, the MessageLabs Web Security Services v2 are configured as the next upstream proxy device on either a user's machine, existing proxy, MessageLabs Client Site Proxy<sup>1</sup>, or perimeter device.

### 2.3 Global Load Balancing Model

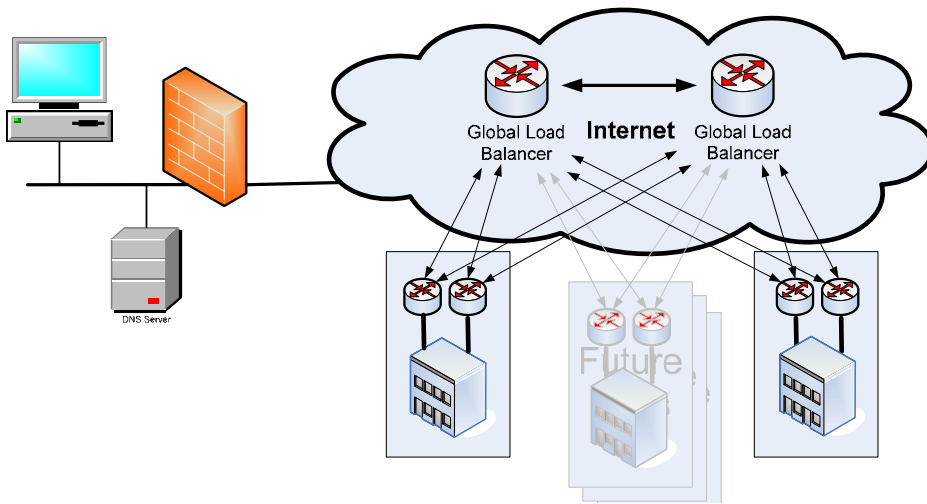
As introduced earlier in this document, the infrastructure provides a network-level load-balanced environment for the service which does not require any administrator knowledge or configuration.

The service is accessed as a single entity, which then routes the user requests to the closest available location with sufficient spare capacity.

The Global Load Balancing model comprises redundant Global Load Balancers which in turn communicate with redundant Local Load Balancers located in each data center which control the routing and the loading of the devices within that location.

---

<sup>1</sup> The Client Site Proxy is a MessageLabs-provided application which allows the capture of user level information for configuration and reporting.



(Guiding illustration only, does not accurately reflect actual infrastructure)

### 2.3.1 Technical Overview

The Global Load Balancers operate in the guise of DNS<sup>2</sup> servers that when queried about the IP Address for the DNS Name of the service will return the IP address of the most appropriate data centre. The process is extremely fast, using only minimal data packets, and returns the information without impacting the user experience.

The process is started by a client machine making a request for a web page or web-based content (e.g. the user opens their browser and attempts to access their favorite search engine).

This request is then passed through the client's infrastructure to the MessageLabs Web Security Services v2, which are accessed using a hostname.

This hostname is then resolved to an IP address. To do this, the client's local DNS server will be tasked with sending a request, asking for the host IP address of the hostname. The DNS Authority for the MessageLabs Web Security Services is the Global Load Balancers which will accept the request for a DNS lookup to obtain the IP Address by the client's DNS server.

The Global Load Balancers then instruct the routers in each data centre to send a (redundant) request direct to the user's own DNS server. The request sent is insignificant, and results in the client's DNS server sending an RST (reset) back.

The data center routers send the response time for the round trip from the data center to the client back to the Global Load Balancers, which in turn compare the response times from the other data centers. This information is then used, together with other loading and geographical data, to determine which data center is most appropriate for the user's session. In this way we can guarantee that the destination provided to the user is the closest and/or most available data center without the need for primary and/or secondary data center configurations or the need for complex configuration on the client site to share the load through technologies such as client site load balancers or "round robin" DNS.

The Global Load Balancers respond to the client's DNS server with the IP Address of the "winning" data center. This IP is then used as the target for the user, and any further connections to the MessageLabs Web Security Services. Once the client DNS server has this data center location cached, it will be used for that and any subsequent requests without having to route via the Global Load Balancers.

---

<sup>2</sup> DNS – Domain Name Service; DNS holds the information that ties a host name to its IP address, or vice-versa. An organization's DNS Servers will generally hold the information about its own devices and will cache (for a period of time) information about external hosts. If a DNS Server does not have the IP information about a requested host, it will query the DNS Root Servers to find an authoritative DNS Server (or "Authority") who "owns" the information for that domain and send the query to that server to get the IP Address back.

The information is kept up to date by the DNS lookup having a low TTL<sup>3</sup> and will be purged from the client DNS server at regular intervals, at which point an update requested from the authoritative source for the MessageLabs Web Security Services will initiate the Global Load Balancers which, as above, return the best location for the client web requests at that point in time.

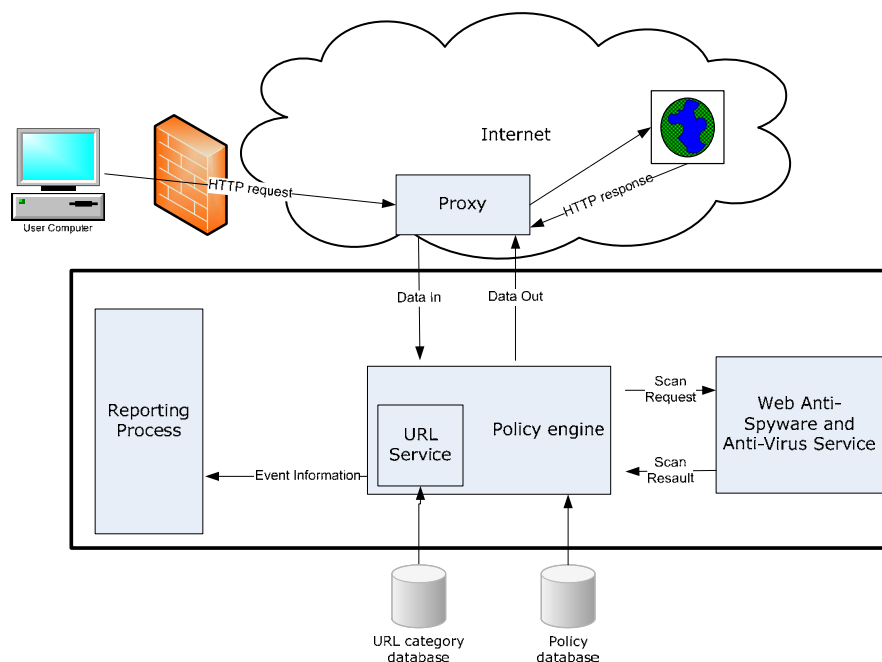
This is a proven technology in use in numerous high volume environments across the Internet, ensuring compatibility, scalability and availability of the service and thereby facilitating for users the best possible browsing experience.

## 2.4 MessageLabs Web Services v2 Data Centre Architecture

As with all MessageLabs services, each data center incorporates layers of redundancy within that data center to ensure that there is no single point of failure.

The data centers have their own redundant Load Balancers, which are responsible for accepting the web requests and handling the request most effectively.

Under each Load Balancer, a number of proxy servers accept and service the requests on behalf of the user, at the same time passing the information to the MessageLabs Web Services policy and scanning servers.



When the proxy receives a request, it will pass the request to the Policy Engine for processing and validation.

If the request requires analysis by the MessageLabs Web Anti-Spyware and Anti-Virus service, then information is passed to that process.

Depending on the content being downloaded, the MessageLabs Web Security Services use a combination of downloading and trickling data to ensure a transparent user experience.

Where there is a large item being scanned, the MessageLabs Web Service may choose to begin a trickle feed of the item through the system to the downstream device to ensure the user experience is maintained, and there is no timeout on the part of the downstream device.

If during this process the item being downloaded is found to contain malicious content, then the transfer will be aborted, the data transferred ignored and, if enabled, the user will receive an alert informing them that the file was found to be malicious.

<sup>3</sup> TTL – Time To Live – the duration of the information being considered as valid. Once the TTL expires, a new request is made to get the latest information.

### **3 The Web Anti-Spyware and Web Anti-Virus Service v2**

The Web Anti-Spyware and Web Anti-Virus service provides detection of known Malware and Spyware using a combination of commercial and Skeptic-based technologies. We also provide detection for a range of potentially unwanted programs (PUPs). PUPs are pieces of software that whilst not malicious are generally unwanted by IT Administrators.

#### **3.1 Scanning Methodology**

All items passed through the service are scanned for malicious content including Viruses, Trojans, Spyware, Adware and Potentially Unwanted Programs.

Web pages are also scanned in their original context to provide protection against known malformations and browser exploits.

The MessageLabs Web Anti-Spyware and Web Anti-Virus service scans the complete file regardless of size. This is important as it could be possible to create a large archive (e.g.: .Zip file) and add multiple files to it, the last of which is infected. If the complete item were not scanned, there is the risk that an item could be passed without analysis.

In addition to its high levels of archive handling, the MessageLabs Web Security Services employ protective measures against corrupt and oversized archives.

#### **3.2 Commercial Technologies Employed**

MessageLabs employs three leading Anti-Virus and Spyware engines in the MessageLabs Web Anti-Spyware and Web Anti-Virus Service v2.

These engines provide detection for a full range of Malware, Spyware, Adware and Potentially Unwanted Program detection.

#### **3.3 Skeptic Technologies Employed**

MessageLabs has a reputation for leading the field in the pro-active detection of new Malware in email through its own deeply analytical Skeptic scanning engine. Skeptic is the MessageLabs zero hour protection engine and is the recognized industry leader in this field.

To ensure maximum performance, when a new item of Malware is identified by Skeptic, the Skeptic team creates a Skeptic signature capable of rapidly identifying the new item. This signature, known as a Skeptic Filter, is then used to identify further instances of that threat without requiring additional scanning by the full Skeptic engine.

MessageLabs Web Security Services is the first MessageLabs solution to add Skeptic Filters to its detection capabilities, thus providing Converged Threat Analysis. This reflects the evolving nature of threats among different protocols, as URLs linking to malware are sent within emails and executed via the web.

This capability provides MessageLabs clients with rapid protection from blended or 'converged' threats that utilize both email and web-based methods for propagation, as well as from risks through continued use by employees of web-based email and other shared community applications within the perimeter of the company.

## 4 The Web URL Filtering Service v2

The Web URL Filtering Service employs a recognizable industry standard method of stacking multiple rules which when combined create the overall policy portfolio the organization requires to meet their needs.

A client's policy is applied by an intuitive combination of policy rules affecting When, Who, Where and What.

Each of these elements can be ignored or used as part of each policy rule, giving the maximum flexibility.

Each policy rule is placed into a rule "stack" and is then evaluated in order from top to bottom. When a policy rule is matched, the action associated with that rule is executed and the process exits the stack.

The elements are used to create rule criteria which have to be matched in order for the rule Action to be executed.

Available actions are:

- Allow
- Allow and Log
- Block
- Block and Log

### 4.1.1 When

The Policy Engine allows the organization to define multiple time periods within one time zone during for which the rule shall be active.

Time Periods can be "All Days" or any specified day (e.g.: Monday-Friday, or Monday, Wednesday, Friday or Saturday, Sunday).

Time of day can be multiple times such as 9am-1pm, and 2pm-5pm to represent a working day. The time intervals are in multiples of 15 minutes.

This part of the Policy Engine will consider the criteria to have been met when the time of the request matches any time period in the rule.

### 4.1.2 Who

The user portion of the Policy Engine can use Directory User and Group information or Custom Groups containing IP's or Domain\User information as part of the policy selection criteria.

This part of the Policy Engine will consider the criteria to be met when any of the following applies:

- The user making the request is explicitly listed a custom group listed in the rule
- The user is a member of a Directory group listed in the rule
- The request has come from an IP address listed in a custom group

### 4.1.3 Where

Destinations are considered matched when any selected URL Category, specified URL or IP Address is requested.

Specified URLs are effectively wild-carded, so the following URL:

[www.message-labs.com](http://www.message-labs.com)

would imply any and all pages beginning with this URL without the need for adding a wildcard at the end.

The URL can also be wild-carded using the '\*' character, allowing you to block multiple sites within a domain e.g.: '\*.acme.com'

#### 4.1.4 What

This allows for the matching of any selected MIME type, or any selected file type by extension.

There is also the facility to define the organization's own MIME and File types for inclusion in the rule.

#### 4.2 Policy Processing

In order for a rule to be invoked, the criteria of each section above must be met (either through the criteria being matched or that element being ignored).

Where the organization in question has policy rules in place that enforce blocking of content from being downloaded by URL Category or Specific URL/IP address, the Policy Engine will first check if the IP/URL requested matches any rules asking for that specific destination to be blocked, and if there is a match then the page will not be retrieved and the user will be redirected to an appropriate block page.

Where the rule requires MIME or File type content to be checked as part of the rule, this object is checked separately from any other object on the page and allowed or disallowed based on the rules set.

The Policy Engine uses a simple administrator defined top down ordering system to ensure that policy rules are processed in the appropriate order.

This flexible approach means that the policy engine avoids confusing exception-based Block or Allow lists tucked away within each rule, but still allows the creation of an infinite number of highly detailed policies/rules to handle these business exceptions simply by the order in which they are placed in the stack.

#### 4.3 The Policy Stack

Although there are no hard and fast procedures defined for the policy stack, and it is purely down to administrators as to how they wish to create their rule set and order them, the policy stack is treated much like a firewall policy set and as should be familiar to the majority of network administrators who have seen this methodology.

Due to the way this and other similar Policy Engines operate the following stack is considered best practice (from top to bottom).

##### 4.3.1 Exception handling policy rules

These policy rules sit at the top of the stack, and hence will be processed before any other rules. This provides the means to block or allow individuals or sites where a unique of exception rule is necessary.

Examples of this would be:

- Senior Executive allowed access to competitive website job section
- Marketing or other research group allowed access to streaming media
- Where a site is considered to be mismatched as a particular category that specific site can be added as an ALLOW rule for continued access till the issue is resolved

##### 4.3.2 General Allow or Block policy rules

These will form the basis of the company productivity policy, and are likely to be time and/or group based.

These rules are not likely to be high risk, but will be focused on preserving company resources and maintaining high levels of productivity.

Examples of this would be:

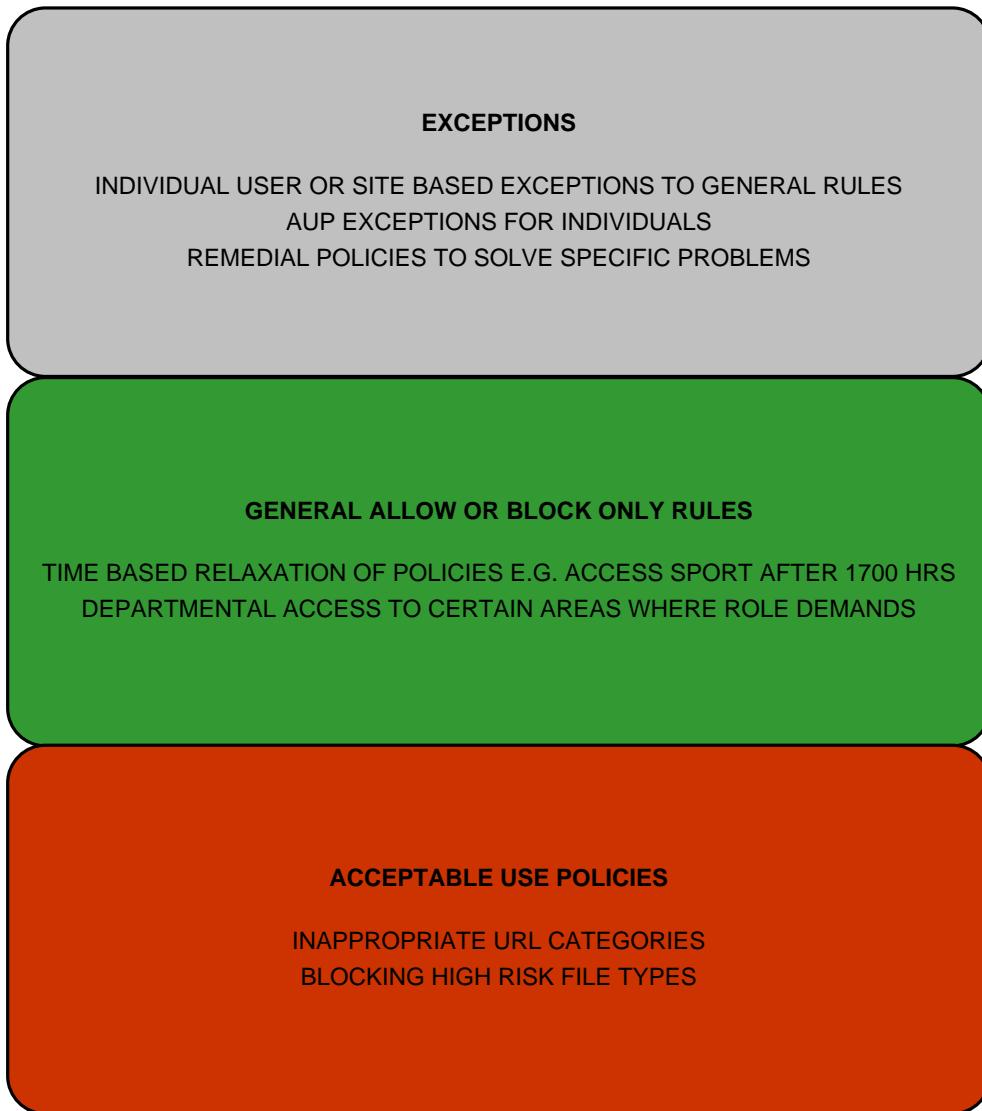
- Allow access to recreational sites during lunch break and before/after work.
- Block access to Job sites or competitor website job section
- Block access to high bandwidth MIME and file types e.g.: MIME Audio and Video

### 4.3.3 Acceptable Use Policy (AUP) rules

An AUP is a set of policy rules which must be adhered to. These are likely to contain policies barring access to sites such as Hate, Weapons, Gambling and inappropriate content for the workplace.

Another AUP will be used to block access to restricted MIME and File types.

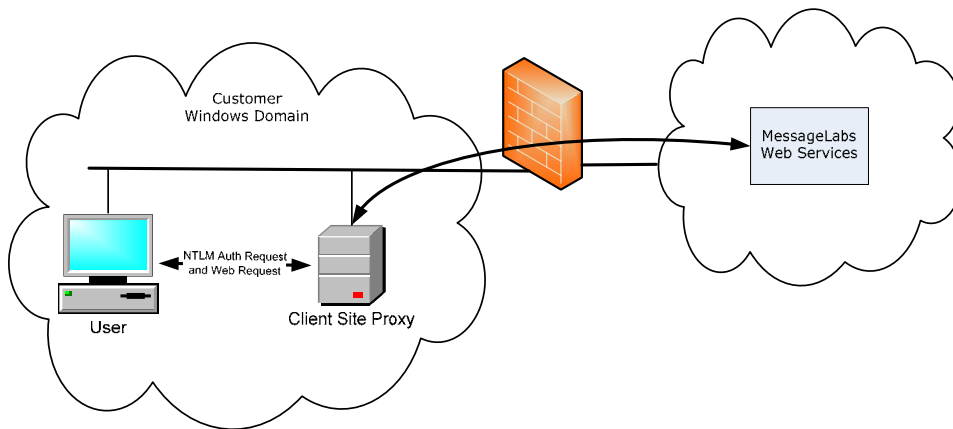
The diagram below shows an example of how a Policy Stack could be structured;



## 5 Client Site Proxy

In order to capture the internal IP address and the user identity required for Group and User level configuration and reporting it is necessary to have a small information gathering proxy operating within the company.

This proxy is known as the Client Site Proxy and performs the task of capturing the IP Address, Windows Domain and Login ID of any users who request information.



Users within the organization are defined to proxy any requests to the Client Site Proxy (CSP).

The CSP operates within the same Domain and as part of the Active Directory (much like any of PC on the network). When a user attempts to route via the CSP an NTLM Authentication challenge is made, asking for the identity of the logged in user who wants to download information from the Internet.

The user machine responds with its credentials and the Client Site Proxy verifies the user identity with the domain controller. Once the identity has been confirmed, the User Name and internal IP address is passed up to the service as part of the request allowing MessageLabs Web Security Services v2 to identify the user.

In the event that a machine which is outside the Active Directory domain is challenged for the user identity, then the user is asked to enter credentials in order to authenticate their web request.

The MachineName\LoginID can be pre-assigned a group membership within MessageLabs Web Security Services using the Custom Group functionality and used as part of a policy rule.

On its own, the CSP is not sufficient to provide user level settings and configuration, however this information can be used to create Custom Groups for use by the service, where the administrator can create lists based on the user identify (e.g.: Domain\LoginID) and / or the source IP of the connection to the CSP. In some situations that may be sufficient for reporting purposes, without the need for a full directory synchronization.

Full directory based user level reporting, alerting and configuration can be achieved when the CSP is used in conjunction with the Group Synchronization Tool.

## 6 Custom Groups

Where there are policies which require information or identification of users outside the Directory, or it is desirable not to synchronize the directory there is the facility for the administrator to create a Custom Group within the Web URL Filtering Service v2.

A Custom Group contains the following information;

- Domain\LoginID
- MachineName\LoginID
- IP Address or range

As detailed above, the Domain\LoginID and MachineName\LoginID are provided by the CSP and can be referenced within a custom group without the need for Active Directory Synchronization.

Custom Groups can be used where:

- Directory synchronization is not possible
- Directory synchronization is not available
- Grouping is specific to a rule, and the administrator does not want to maintain the group in the Active Directory.
- The Group needs the combination of LoginID's and IP Addresses
- The Policy Rule is for a single IP or range of IP's

As well as providing for Custom Groups, the MessageLabs Web Security Services v2 also supports the capture and use of Groups held within an Active Directory.

## 7 Group Synchronization

In addition to Custom Groups, MessageLabs Web Security Services v2 allows the synchronization of Users and Groups information from a client Directory source to the MessageLabs Web Security Services.

These directories are captured using a PUSH methodology ensuring that an existing Directory remains the master or primary source.

In order to support user level settings for groups, and user level alerting for policies, the user information needs to be captured from the client's Directory.

The following information is captured by default:

- Network LoginID
- Primary email address
- Security Group Membership

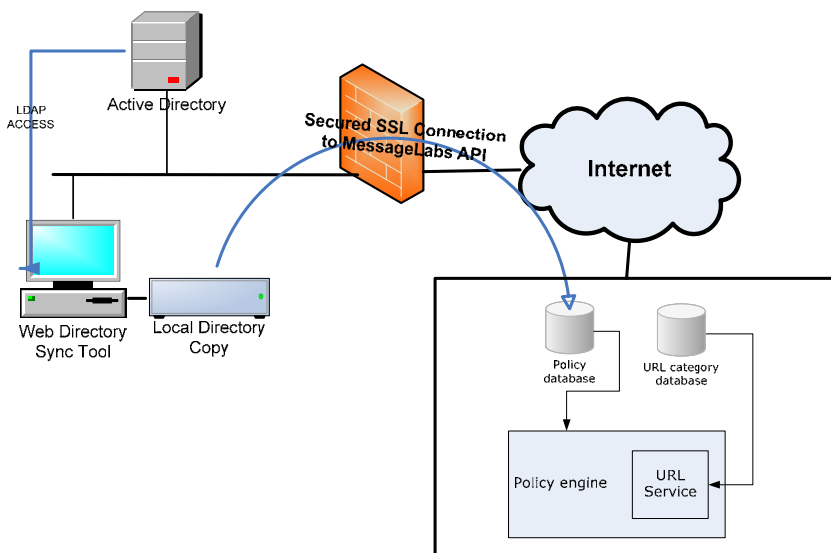
The Domain Name and LoginID are used for the identification of the user when that information is passed on by the CSP. Primary email and security group membership is used for the optional user level alerting and URL filtering rules.

The Directory groups are not editable, nor can you modify group memberships once synchronized – if that is required, then it is recommended that the exception is handled by a Custom Group.

### 7.1 How the Synchronization Works

User level information is accessed via the Web Security Services Group Synchronization Tool

This tool is a Java application that allows the administrator to select a subset of their Active Directory and select fields to be captured and replicated to MessageLabs. The selection is done using a combination of the tool, and also specifying what LDAP fields the administrator wished to be replicated.



The tool maintains a local copy of the Active Directory, and will replicate any changes to the directory to the MessageLabs Web Security Services v2 on demand as a one-way synchronization (i.e. the Active Directory is treated as the primary information source, with the MessageLabs Web Security Services policy database being classified as secondary).

The connection between the Group Synchronization Tool and the Active Directory can be optionally secured through the use of authenticated connection and LDAPS, while the connection to MessageLabs from the tool is always over SSL and secure. The tool allows clients to have full control over the synchronization with MessageLabs without opening their directory to direct access from the outside world.

This tool is included as part of MessageLabs Web Security Services.

**www.messagelabs.com**  
**info@messagelabs.com**

Freephone UK  
0800 917 7733

Toll free US  
1-866-460-0000

Europe  
HEADQUARTERS  
1270 Lansdowne Court  
Gloucester Business Park  
Gloucester, GL3 4AB  
United Kingdom

T +44 (0) 1452 627 627  
F +44 (0) 1452 627 628

LONDON  
3rd Floor  
40 Whitfield Street  
London, W1T 2RH  
United Kingdom

T +44 (0) 207 291 1960  
F +44 (0) 207 291 1937

NETHERLANDS  
Teleport Towers  
Kingsfordweg 151  
1043 GR  
Amsterdam  
Netherlands

T +31 (0) 20 491 9600  
F +31 (0) 20 491 7354

BELGIUM / LUXEMBOURG  
Culliganlaan 1B  
B-1831 Diegem  
Belgium

T +32 (0) 2 403 12 61  
F +32 (0) 2 403 12 12

DACH  
FeringasträÙe 9  
85774 Unterföhring  
Munich  
Germany

T +49 (0) 89 189 43 990  
F +49 (0) 89 189 43 999

© MessageLabs 2005  
All rights reserved

Americas  
AMERICAS HEADQUARTERS  
512 Seventh Avenue  
6th Floor  
New York, NY 10018  
USA

T +1 646 519 8100  
F +1 646 452 6570

CENTRAL REGION  
7760 France Avenue South  
Suite 1100  
Bloomington, MN 55435  
USA

T +1 952 886 7541  
F +1 952 886 7498

Asia Pacific  
HONG KONG  
1601  
Tower II  
89 Queensway  
Admiralty  
Hong Kong

T +852 2111 3650  
F +852 2111 9061

AUSTRALIA  
Level 6  
107 Mount Street,  
North Sydney  
NSW 2060  
Australia

T +61 2 8208 7100  
F +61 2 9954 9500

SINGAPORE  
Level 14  
Prudential Tower  
30 Cecil Street  
Singapore 049712

T +65 62 32 2855  
F +65 6232 2300