

## White paper on Web Content Filtering

### Web Content Filtering

#### Overview

Content Filter will provide complete control over web site access. Increase productivity, reduce recreational web surfing and limit legal liability associated with pornography and illegal file downloads.

In today's organization, Internet access is required. When an organization decides to open the door to the Internet, the good along with the bad will enter the network. The need for an effective web content filter is required to protect the best interests of the organization.

#### For Business

When employees access inappropriate or illegal content, organizations lose productivity, expose themselves to legal liability, and experience a decrease in network performance. Spyware, viruses, and worms can be easily distributed onto the network from untrusted web sites. A web content filtering solution can alleviate all of these problems by blocking access to inappropriate web sites. Popular search terms are NOT business related.

#### For Schools & Libraries

Schools and libraries can benefit from web content filtering by blocking inappropriate content while providing access to relevant online information. The content filter can protect students from sexual predators by blocking access to user forum and chat web sites. The Children's Internet Protection Act (CIPA) gives public schools and libraries federal assistance for Internet access only if they adopt an Internet usage policy and install content filtering to prevent access of inappropriate material.

### Importance of Content Filtering

#### Increase Employee Productivity

Blocking access to inappropriate content will provide your employees with fewer opportunities to become distracted while on the job. Web sites that provide news, sports and auctions are designed to consume people's time on a routine basis (e.g. daily, hourly). Time spent browsing the web is time not spent with customers or other tasks required for normal business operation.

#### Reduce Legal Liability

Legal liability can appear in many forms. Anything from sexually explicit content to copyright music downloaded onto the corporate network can cause problems for any organization, large or small. Limiting access to objectionable web content can alleviate costly lawsuits that could harm the organization. Sexual harassment and discrimination lawsuits can easily be aggravated without enforcing a strong Internet access policy. Control over web content can greatly reduce the legal exposure an organization may experience.

## **Bandwidth Savings**

Bandwidth savings can be realized once web content control has been implemented into your network infrastructure. The average size of a web page including all graphic elements is well over 100KB. It is very easy to see that once you start multiplying this number by the number of network users and the number of non-work related web pages visited, this waste of bandwidth can add up quickly. Save money on Internet access and obtain faster response times by filtering access to inappropriate web content.

## **Database-based versus Dynamic Filtering**

Database-based content filtering is a proprietary collection of URLs, designed to associate each URL with a specific content category that a user may want to filter (e.g. www.cnn.com = News). When a site is requested, the filter looks up the address of the requested web site in the database. By comparing the requested URL against the database, the filter can block or allow access to the site. Even the largest URL database contains only a fraction of what is on the web. Manual classification cannot keep up with the quantity of web content and the frequency with which it changes. A URL database is out-of-date the moment it is created. Dynamic content filtering performs on-the-fly content analysis of a web page. When the web page is received, it is analyzed and then categorized according to the content found on the page. After the content has been scanned, the system determines whether to block or allow access.

This next-generation dynamic filter analyzes every web page, whether it has existed for 5 months or 5 minutes. Dynamic filters have no problem keeping up with the size, growth and constantly changing Internet content.

## **How Content Filter Works?**

### **Active Content Recognition**

Content Filter is powered by Pure Sight, based on proprietary Active Content Recognition (ACR) technology. ACR is the core technology that powers Content Filter. It is composed of a set of innovative artificial intelligence algorithms that analyze and categorize data in real-time.

### **Categorizing Web Content**

The first step in categorizing web content is to break the HTML page down into parameters. The next step is to analyze these parameters and extract certain patterns and similarities. Finally the ACR will match these patterns with a specific content category

Patterns of similar web sites are grouped into the same category. For example, one category represents sexual content, while another represents sports, and yet another represents gambling. By matching the patterns to one of the categories, the ACR engine is able to identify the type of web page being requested by the user.