

CONTENTWATCH

# INTERNET FILTERING SOLUTION

---

CONTENTWATCH INTERNET FILTERING  
END-TO-END TECHNOLOGY

---

## INTRODUCTION AND EXPLANATION

---

In the world of ever changing Internet content and increasingly mobile computing devices, ContentWatch takes a unique and comprehensive approach to end-to-end Internet filtering solutions. This white paper explains the pros and cons of various deployment options and analysis technologies. It also demonstrates why ContentWatch end-to-end deployment, combined with Dynamic Contextual Analysis, is the most advanced Internet filtering technology available today.

This white paper is divided into the following sections:

- ✓ Filtering Technology Overview: Dynamic Content Analysis vs. List-based technologies

*This section describes some of the methods that are used to determine which Internet content does or does not meet your corporate Internet usage policy and why the ContentWatch solution is the best solution for today's constantly changing Internet content.*

- ✓ The ContentWatch Solution: The value of an end-to-end, Dynamic Analysis Solution

*This section outlines the ContentWatch end-to-end solution and describes why it is the most comprehensive solution to enforce your Internet usage policy.*

- ✓ Client Architecture

*This section describes the client-side architecture and explains how the ContentProtect software integrates with the Windows operating system to provide Internet usage policy enforcement at the Winsock layer.*

- ✓ ContentProtect Professional Appliance

*This section outlines the hardware specifications and architecture of the ContentProtect Professional Appliance.*

- ✓ The ContentWatch Data Center

*This section describes the ContentWatch Data Center (CDC)—a key component of the ContentWatch solution. The CDC drives ContentProtect's remote management and reporting capabilities and enables administrator's to synchronize settings across multiple computers.*

- ✓ Solution Deployment: Filtering at the client/machine vs. filtering at the gateway

*When implementing any content filtering solution, you must consider the existing network infrastructure to determine whether a client or server based implementation would be best. For some, a pure client-side solution is the best option, while others require nothing more than a gateway solution. Still others can benefit from a combination of both, or end-to-end deployment. This section explores the benefits and requirements of each option.*

---

## FILTERING TECHNOLOGY OVERVIEW

---

There are various technologies on the market today for filtering Internet content. Before discussing any end-to-end solution, it is important to understand current Internet filtering technologies and their associated strengths and limitations. This section outlines three of the main types of Internet filtering technologies that are available in most filtering products today.

### KEYWORD ANALYSIS

Keyword analysis refers to the ability to programmatically watch for the occurrence of certain words and block access to sites that contain those words. Keyword analysis was one of the first attempts at website categorization and is still frequently requested by users due to the fact that there are certain words that people feel are inappropriate and should never be seen, regardless of the context.

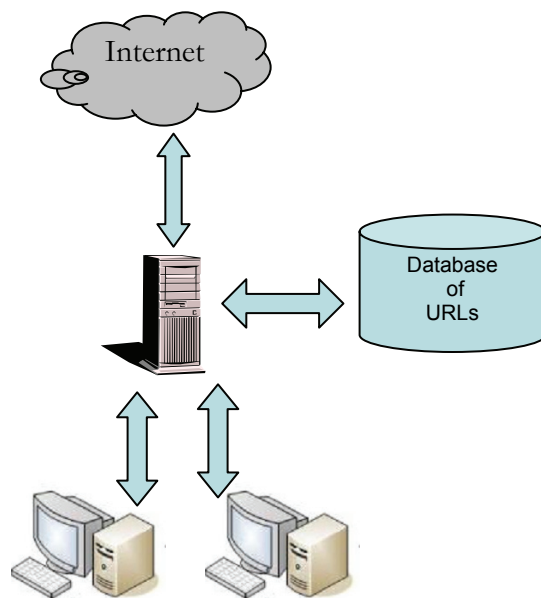
By itself, keyword analysis does not provide an adequate filtering solution. On the one hand, blocking too many words blocks clean sites while not blocking enough words allows undesirable content. Many products use some type of minimal keyword analysis to augment their solution; however, there are no longer many purely keyword-based filters on the market because the technology is unreliable.

### URL ANALYSIS

URL analysis is the process of interrogating the URL address that identifies a website and categorizing the content based on a prior human review of the content hosted on that URL. This is typically done by referencing a large database of pre-categorized URLs.

There are several problems with this technology:

1. There is a great deal of subjectivity with human review. Despite categorization rules, websites with similar content may be categorized differently depending on who reviewed the site.
2. The content on many websites is constantly changing. This is especially true with portals, news sites, blogs, and other sites with dynamic content. These sites may have violent content one day and sexually-explicit content the next, depending on what the latest scandal is. A database of pre-categorized URLs cannot provide an adequate filtering solution for today's ever-changing Internet content.



3. URLs change often, especially in the world of pro-active, targeted pornography predators. Keeping up to date with these URLs is a challenge, and requires frequent updates to the database to maintain consistency and accuracy. Maintaining an accurate and current blacklist is extremely difficult; maintaining a categorized list of the entire World Wide Web is nearly impossible.
4. Much like virus pattern files, the URL database must be constantly updated. Some URL categorization companies pride themselves on the frequency of their updates. Some of these updates occur as often as every two hours! The problem with frequent data transfers is the drain on the corporate resources required to keep these databases updated and the problems that can occur if those updates are interrupted.
5. Because the database required to host the URL categorizations is extremely large, this type of analysis usually requires a proxy implementation. This means that a request for a web page is routed to the server that hosts the URL database prior to granting access to the website. If the URL is allowed, the proxy server retrieves the content and passes it to the machine that made the request. This type of implementation requires considerable back-end support, as well as geographically co-located proxy servers, and depending on system traffic, the proxy server may become a bottleneck.

Most of today's commercial web filters are either based on URL analysis or use URL analysis as part of their solution.

## CONTEXTUAL ANALYSIS

Contextual analysis is the ability to interrogate the content and determine on-the-fly what that content is intended to convey. This technology uses sophisticated linguistic algorithms that make associations between the words embedded in the content to arrive at an understanding of what is being communicated. This information is then used to categorize the content as a whole.

Using contextual analysis, one can tell the difference between content discussing the problem of pornography addiction and a website displaying pornography. Similarly, contextual analysis can tell the difference between a web page discussing the effects of breast cancer and a website talking about how to cook chicken breasts. It can also differentiate these sites from a page showing breasts in a pornographic manner.

Since contextual analysis is performed by an algorithm, it is far more consistent than URL analysis performed by human review. Contextual analysis removes the inherent subjectivity of human-review analysis because the algorithm always returns the same results.

---

## THE CONTENTWATCH SOLUTION

---

### DYNAMIC CONTEXTUAL ANALYSIS

ContentWatch has developed a proprietary, patent-pending Dynamic Contextual Analysis engine that can quickly analyze electronic content in context and, with a single-pass, categorize the content. ContentWatch has several applications that can then make determinations regarding that content, including the award-winning ContentProtect Internet filter.

In addition to analyzing the website content exposed to a consumer, the ContentWatch algorithm considers the websites that are connected to the site via embedded links, as well as meta-data that may be hidden from the reader.

The combination of the Dynamic Contextual Analysis engine with its multiple analysis techniques, and the ContentWatch algorithm that considers all data related to a website, makes the ContentWatch engine extremely powerful and very accurate. The ContentWatch engine is able to accurately identify content without prior knowledge of the URL content or human review of the requested content.

To provide the most thorough filtering solution possible, ContentWatch does use other types of basic filtering, including human URL analysis; however, ContentProtect does not depend on human review to properly categorize content. Instead, the content is primarily categorized by the Dynamic Contextual Analysis engine. For example, ContentWatch maintains and updates a small list of URLs that do not change often and are well-known for a specific type of content. However, requested content is still analyzed by the ContentWatch engine to verify that the content falls into the categories indicated by the prior analysis. This redundant analysis provides a fail-safe for URLs that may be hijacked for a period of time. In the event that a website is hijacked, ContentProtect is able to accurately identify and block access to inappropriate content, regardless of when the hijack occurred or whether anyone is yet aware of the hijacked website.

The ContentWatch solution does not rely on large back-end databases. Rather, it is a self-contained Contextual Analysis engine. Consequently, it can be directly accessed from the machine making the content request. There is no need to involve a third-party service in the data retrieval request, and therefore, no additional bottlenecks are introduced. The content is requested from the consumers' machine (or LAN appliance) and the content is analyzed on the machine prior to displaying it to the user. The content can then be allowed or blocked in real-time, with very little speed degradation.

## END-TO-END DEPLOYMENT

Clearly the most comprehensive approach to enforcing a corporate Internet usage policy is a combination of an inline appliance at the gateway and client-side filtering. Recently an IDC analyst made the following observation:

“Employee Internet Management (EIM) has traditionally been deployed on servers at the gateway. Today’s risks dictate that additional policy enforcement points should exist at the desktop and network levels as well as at the gateway.” [emphasis added]

– IDC

The problem with most combinations of these technologies is that the content is filtered twice—once at the gateway and again when it arrives on the client. This introduces latency and, if the appliance and client solutions are from different vendors and interpret content differently, the system may utilize more stringent Internet usage rules for some employees than for others. This, in turn, can precipitate legal problems because the corporation cannot consistently enforce its Internet usage policy.

The ContentWatch solution is the most comprehensive end-to-end solution on the market today. The client-side filter communicates with the gateway appliance and is smart enough to know that when it is connected to the LAN where the appliance is enforcing Internet usage, it can allow all content onto the machine. Conversely, it also detects when the appliance is not present and will then enforce the Internet usage policy on the local machine. So, a laptop that is sometimes on the corporate LAN and sometimes connected elsewhere will appropriately and consistently enforce the usage policy, regardless of where and when the laptop is used.

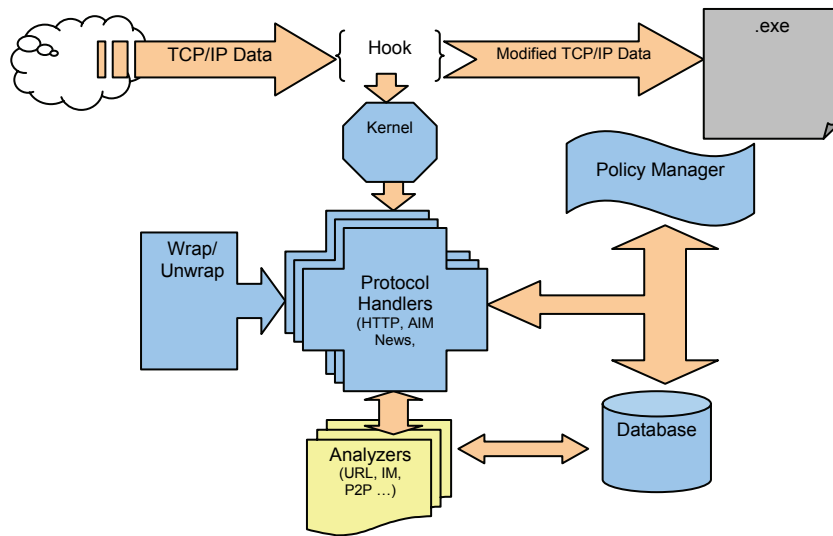
---

## CLIENT ARCHITECTURE

---

The architecture of the client filtering application is simply an extension of the routing architecture used in standard web technologies. ContentProtect uses an inline approach on the client machine to monitor the data as it is requested. Data is analyzed in real time so requests are immediately honored or replaced if the content is blocked.

The following diagram outlines the architecture of the client-side implementation of the ContentProtect filter:



The Hook is implemented as a standard Windows LSP. The LSP, by definition, is inserted into the Winsock TCP/IP stack. As packets come onto the machine from the LAN, the packet is re-routed to the ContentProtect kernel which, in turn, queries the packet to determine what type of communication is represented by that packet. If the protocol is identified as a protocol that needs to be filtered, the kernel routes the packet to the appropriate ContentProtect protocol handler. The protocol handler, in turn, unwraps the packet to analyze the payload. It then routes the content to the appropriate Dynamic Contextual Analysis component to analyze the content.

The result of this analysis is compared to the database using the appropriate policy for the user currently logged in to the Windows system and a determination is made whether to allow, warn or block the content. The result is then returned to the kernel, which determines whether to forward the original packet or wrap a new packet with a different result. The packet is subsequently placed on the wire by the LSP and proceeds to the calling application.

The beauty of this solution is two-fold. First, since it is implemented as an inline LSP, all content that arrives on the machine is interrogated, regardless of the application requesting the content. This means that the corporate Internet usage policy is enforced for every browser on the machine. Secondly, all analysis occurs on the client. There are no calls to a proxy device or database lookups of stale URL categorizations. The content is always analyzed in real-time, providing the most effective and efficient protection against Internet misuse available today.

Because this is an inline solution, ContentProtect 2.x also analyzes the data from any application that requests data from the client machine. This inline analysis technology has been optimized for efficiency, and is extremely fast and accurate.

## HARDWARE SPECIFICATIONS

### *System*

- CPU: Intel Pentium 4 630 3.0 GHz with fan/heatsink
- Memory: (2) STD DDR2 512, Total 1024
- Chipset: Intel E7230 chipset
- Network: 1 x Intel 82573V Gigabit PCI-Express Ethernet; 1 x Intel 82573L Gigabit PCI-Express Ethernet
- HDD: 160GB SATA 8M 3G
- Video: Integrated ATI Rage XL

### *Chassis*

- Form Factor: Mini 1U; 14" rack-mountable chassis
- Dimensions: 16.8"W x 1.7"H x 14"D
- 1 x 100mm 5000RPM fan in chassis

### *Multiple I/O*

- USB: 2 x rear USB ports; 2 x USB header
- Serial Ports: 1 x rear serial port; 1 x serial port header
- Parallel Port: 1 x rear parallel port
- Keyboard/Mouse: 1 x PS/2 Keyboard, 1 x PS/2 mouse
- LAN: 2 x LAN ports, RJ-45

### *Front Panel*

- Power on/off button
- System reset button
- Power LED
- Hard Drive activity LED
- 2 x network activity LEDs
- System overheat LED

### *Power Supply*

- 260 Watt AC power supply
- Thermal control with PFC

### *Regulatory Compliance*

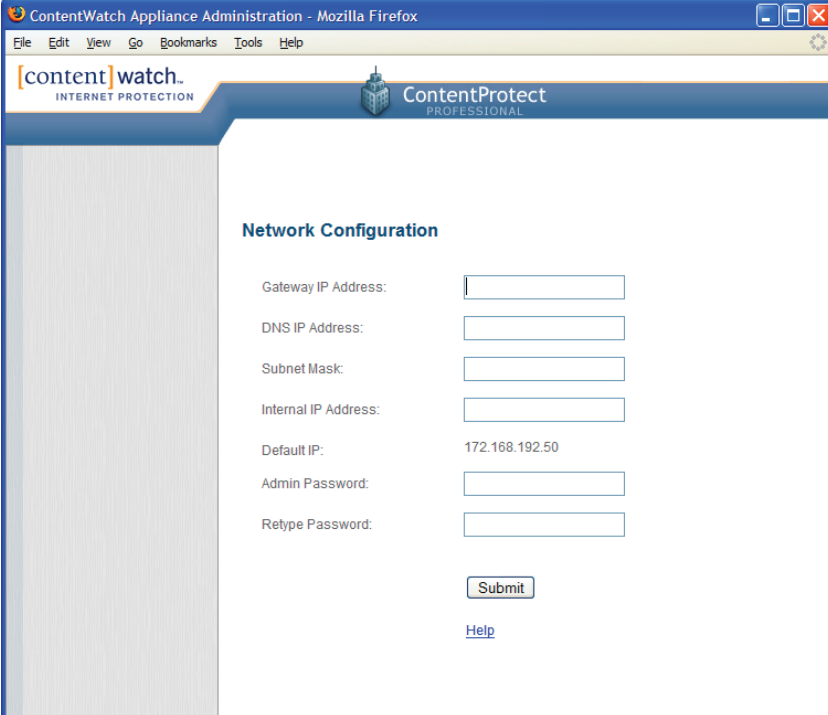
- Electromagnetic Emissions: FCC Class B, EN 55022 Class B, EN 61000-3-2/-3-3, CISPR 22 Class B
- Electromagnetic Immunity: EN 55024/CISPR 24, (EN 61000-4-2, EN 61000-4-3, EN 61000-4-4, EN 61000-4-5, EN 61000-4-6, EN 61000-4-8, EN 61000-4-11)
- Safety: EN 60950/IEC 60950-Compliant; UL Listed (USA); CUL Listed (Canada); TUV Certified (Germany); CE Marking (Europe)



## SETUP/CONFIGURATION OF THE APPLIANCE

The Content Filter Appliance is shipped with a pre-configured IP address of 172.168.192.49. You must reconfigure the appliance to operate on your corporate network. To do this, the filter device must be connected to a network where another machine can access that IP address.

Use a browser on the connected machine to connect to <http://172.168.192.49:3170>. The following configuration page appears:



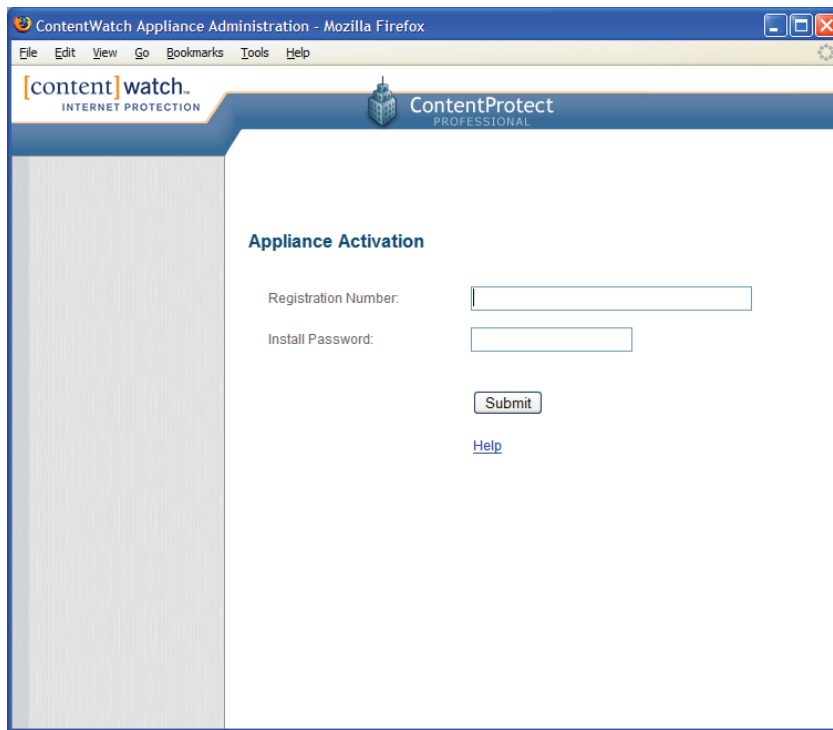
The screenshot shows a web browser window titled "ContentWatch Appliance Administration - Mozilla Firefox". The browser's address bar is empty. The page header features the "content watch. INTERNET PROTECTION" logo on the left and the "ContentProtect PROFESSIONAL" logo on the right. The main content area is titled "Network Configuration" and contains the following fields and values:

|                      |                          |
|----------------------|--------------------------|
| Gateway IP Address:  | <input type="text"/>     |
| DNS IP Address:      | <input type="text"/>     |
| Subnet Mask:         | <input type="text"/>     |
| Internal IP Address: | <input type="text"/>     |
| Default IP:          | 172.168.192.50           |
| Admin Password:      | <input type="password"/> |
| Retype Password:     | <input type="password"/> |

Below the fields is a "Submit" button and a "Help" link.

Complete the requested information, then click **Submit** to store this configuration on the appliance.

After you complete the network configuration, the Appliance Activation page appears. This information allows the device to connect to a ContentProtect organization so that remote management and reporting can occur.

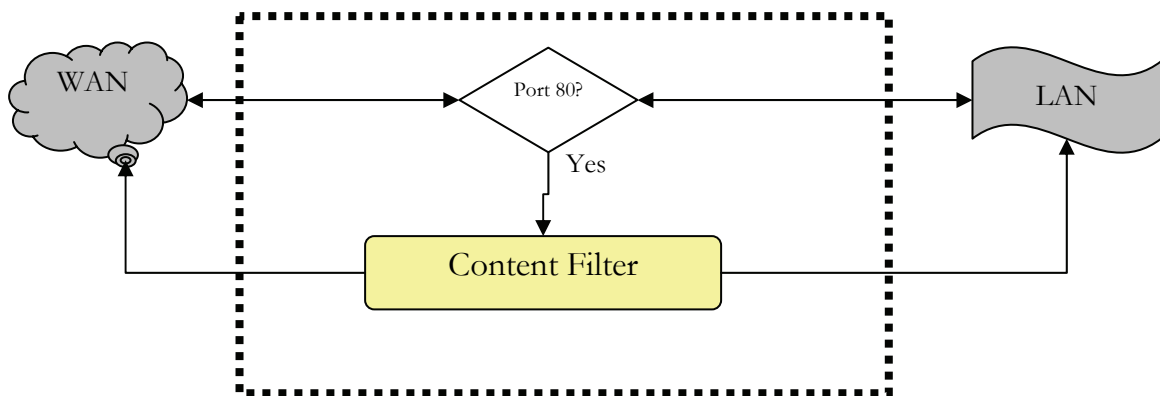


Complete the required information, then click **Submit**. The ContentWatch Filtering Appliance is now configured and ready to enforce your Internet usage policy.

## ARCHITECTURE

The ContentWatch Filter appliance is a standard 1U rack-mountable appliance. It is designed with simplicity in mind and is intended to operate as a plug-and-play inline device. The appliance operates as a bridge with two network interface cards—one for the LAN side of the connection and one for the WAN.

As data flows through the device, all port 80 traffic is diverted to the content filter. Traffic that does not come through port 80 is simply routed through the device without being diverted to the content filter, as seen in the following diagram.



When a request is made on the outbound route, the content filter simply interrogates the URL to see if it already has an action to take. If the administrator has blocked the URL, the request does not proceed to the WAN. Instead, the content filter simply responds with the blocked page.

If the URL is not specifically blocked by the administrator, then the request proceeds to the WAN. When the response comes back from the WAN, the content filter interrogates the content using Dynamic Contextual Analysis. If the content falls outside the Internet usage policy, then the response is replaced with a block page and the content stops at the device. If the content is allowed, then it proceeds to the LAN and the requesting client machine.

The Content Filter is based on the patent-pending Dynamic Contextual Analysis engine and is extremely fast. Additionally, the Content Filter appliance has a built-in cache and has been optimized for speed. The content is interrogated in a single pass and a determination about the content is usually made in only a few milliseconds. Consequently, there is minimal latency to the response.

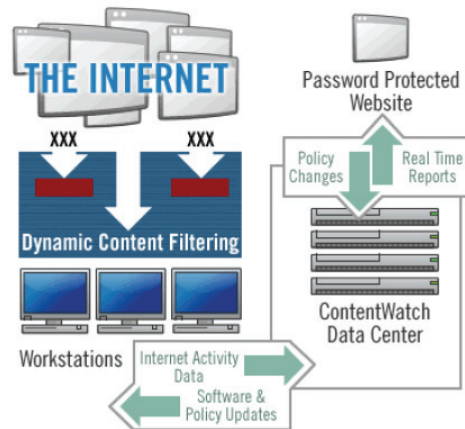
---

## THE CONTENTWATCH DATA CENTER

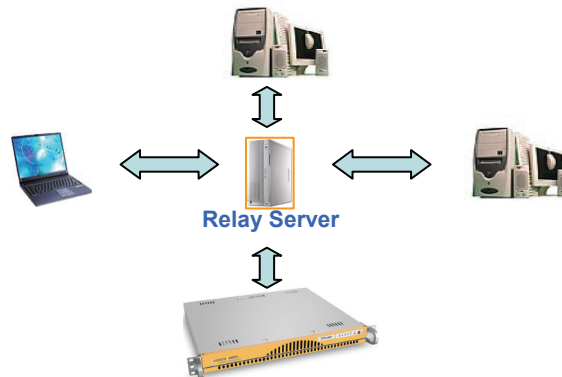
---

### CENTRALIZED MANAGEMENT

At the heart of the ContentWatch solution is the ContentWatch Data Center. Each managed organization stores its settings on the ContentWatch relay server. All filtering devices maintain contact with the relay server to keep their policy settings updated. The administrator can log into the Web Administration console on the ContentWatch relay server to update policy settings, view reports, add or remove users, and so forth. Each change is then replicated to every client and appliance that is part of the administrator's organization.

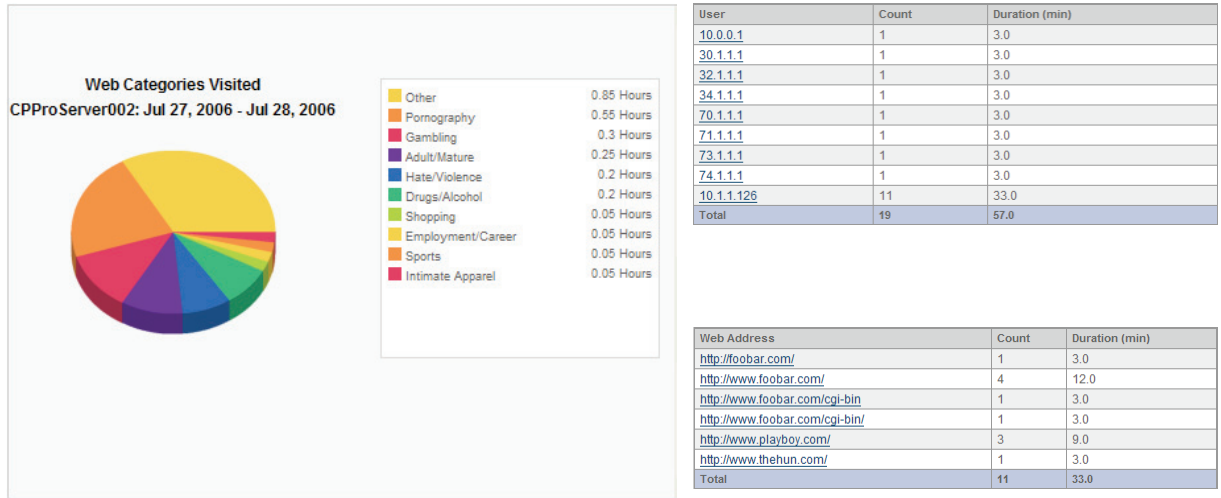


Using this strategy, all filtering devices maintain a common set of policy enforcement data and the data from each device can be aggregated into a single activity report. It does not matter whether your organization uses client-side filtering, edge filtering via the filter appliance, or a combination of both. All of these devices are aggregated into a single management and reporting console at the ContentWatch data center.

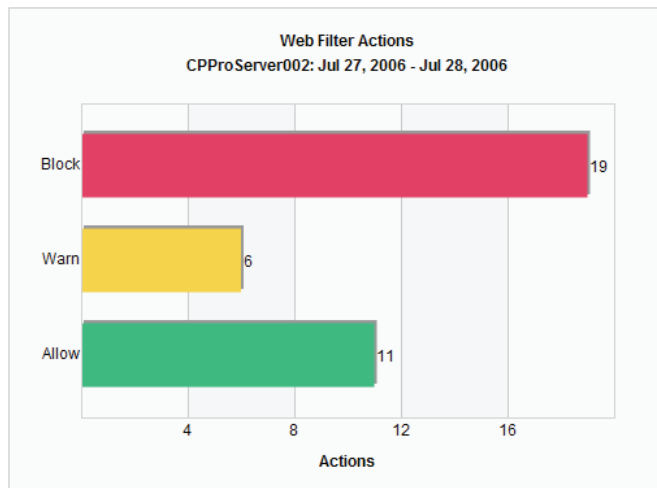


## REPORTING

With all of the activity data for an entire organization stored at the ContentWatch Data Center, administrators can generate aggregated reports for the organization as a whole, regardless of whether the organization is using the client-side product, the edge appliance product, or a combination of both.



Activity reports can be displayed by category and the administrator can drill down to see further detail relating to the user or IP address that initiated the activity, as well as the date, time and actual URL visited.



Reports can be generated to show how many block, warn or allowed sites have been attempted over a given period. These reports provide the same drill-down detail as the category reports—that is user, IP address, and URL requested.

The ContentWatch Data Center magnifies your organization’s ability to enforce and monitor employee Internet usage. The ContentWatch Data Center provides a single point of administration and reporting for your entire organization and the detailed ContentWatch reports help to quickly identify the problem areas that need attention.

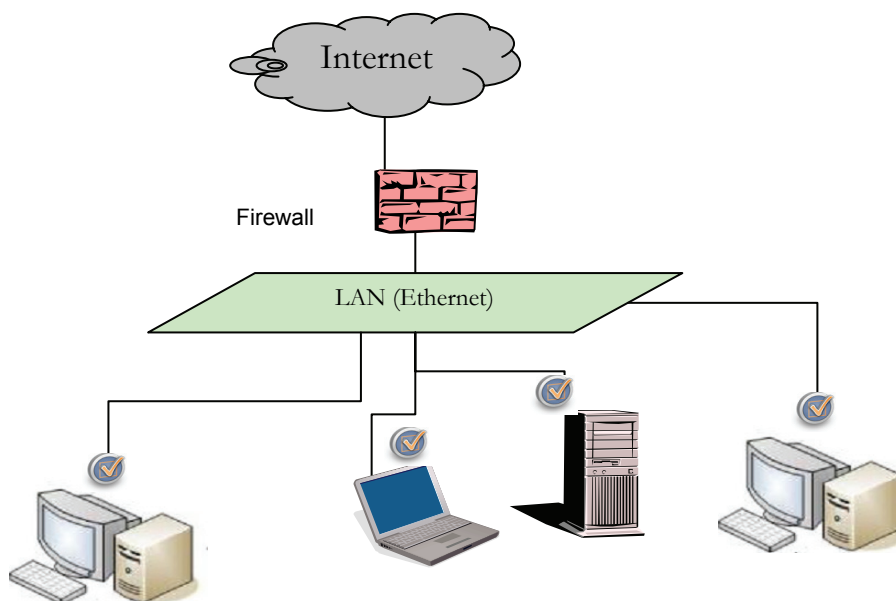
---

## SOLUTION DEPLOYMENT

---

### CLIENT-SIDE FILTERING

The following diagram shows a typical LAN network environment where individual machines are routed through a firewall to access the Internet.



A client-side filtering solution implies that an agent is installed on each machine within the LAN. The agent (depicted in the above diagram as the ContentProtect checkmark) monitors individual machines' access to the Internet. The agent monitors all activity that originates from the machine as well as all incoming data from the Internet. If a user attempts to access something that violates the corporate Internet usage policy, the content is blocked before it displays on the computer.

The advantage of this solution is that the machine is monitored regardless of its location. Whether a laptop accesses the Internet on the corporate LAN, in a hotel room, or from another broadband or dial-up connection, the agent consistently enforces the corporation's Internet usage policy.

Another advantage of this solution is that filtering on the client distributes the filtering solution among all resources on the LAN, thereby reducing the potential for a single-point of failure.

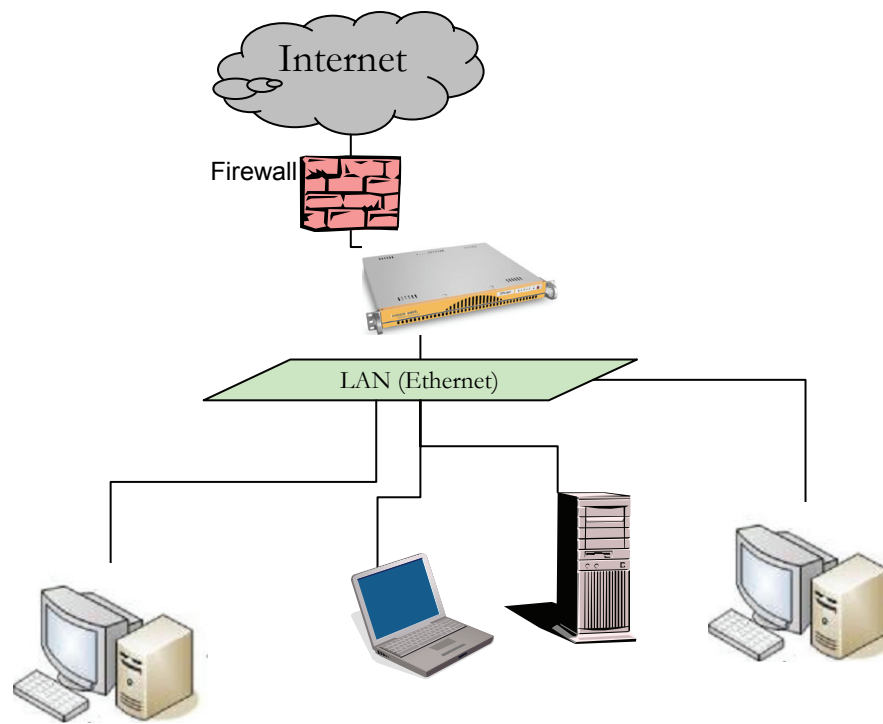
While client-side filtering has its advantages, it also adds some complexity to the network environment. Distribution of the agent could become tedious and maintenance is always a concern for IT administrators. While there is not a single point of failure, if a failure does occur, there are many more touch points to correct the failure.

## FILTERING AT THE GATEWAY

A gateway filter solution has some merit in a LAN environment. Adding a single device inline to an existing network allows an Internet usage policy to be enforced from a single point, avoiding the need to install an agent on every machine. The organization's Internet usage policy is automatically enforced on any machine connected to the LAN.

When machines attempt to access content on the Internet, the filter appliance scans the content before it enters the LAN. If inappropriate content is requested, it is blocked before it even enters the LAN environment, thereby, reducing the company's liability.

The following diagram illustrates the gateway filter solution. The ContentWatch filtering appliance is connected in-line between the firewall and the LAN, effectively enforcing the Internet usage policy for all machines that connect to the existing LAN.



In this environment, there is no need for any agent to reside on the client machines. When any machine connects to the LAN, even temporarily, the organization's Internet usage policy is automatically enforced.

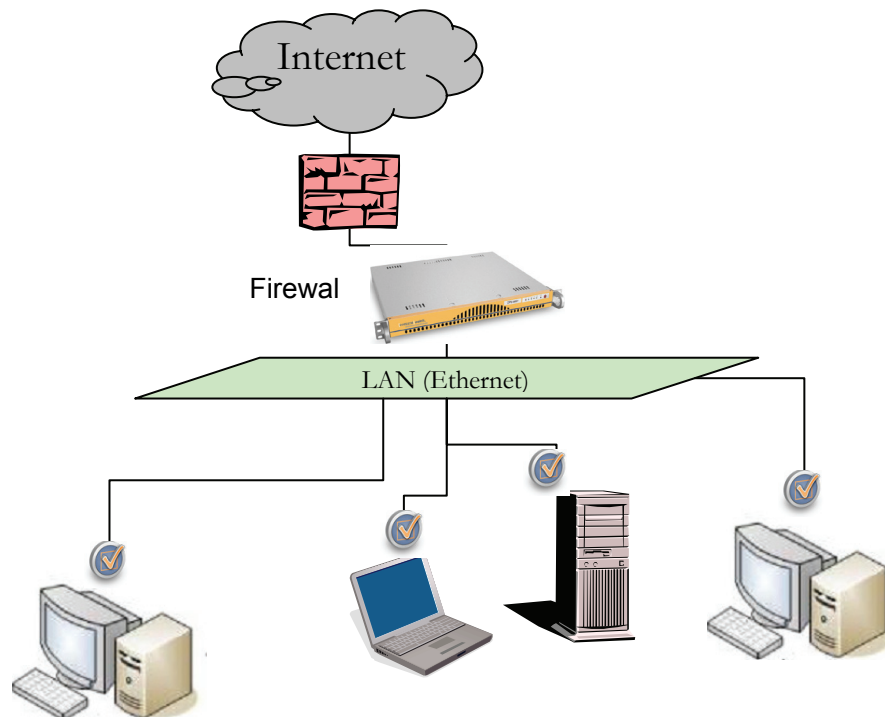
While the appliance solution provides some benefits in a LAN environment, it is limited to protecting only that LAN. If a laptop computer connects to the Internet outside the LAN, the organization's Internet usage policy is not in effect.

The filtering appliance also becomes a potential single-point of failure for Internet access. If a problem arises with the appliance, there are only two choices: either fail-over to non-filtered Internet access, or allow the failure to interrupt Internet access until it can be resolved. While every effort is

made to ensure that the appliance remains highly available, IT administrators must consider all factors when choosing between a client and gateway filter solution.

## END-TO-END DEPLOYMENT

End-to-end deployment ensures that an organization's corporate Internet usage policy is unilaterally applied on all corporate assets, regardless of the point of connection. Distributing a smart filtering agent to each client machine reduces the potential for a single-point-of-failure that comes with a single gateway appliance and provides assurance that the corporate Internet usage policy remains enforced when laptops are not connected to the corporate LAN. At the same time, the ContentProtect Professional Appliance provides the convenience and efficiency of a gateway solution that is very attractive to IT professionals.



Only ContentWatch provides this full coverage, end-to-end solution. In an end-to-end deployment, the client-side filter communicates with the gateway appliance and is smart enough to know that when it is connected to the LAN where the appliance is enforcing Internet usage, it can allow all content onto the machine. Conversely, it also detects when the appliance is not present and will then enforce the Internet usage policy on the local machine. So, a laptop that is sometimes on the corporate LAN and sometimes connected elsewhere will appropriately and consistently enforce the usage policy, regardless of where and when the laptop is used. ContentWatch is the only vendor on the market today that can offer this comprehensive solution.



---

## SUMMARY

---

ContentWatch has developed the most efficient, accurate, and complete solution to the unique problems that arise from trying to enforce an Internet usage policy. The key benefits to the ContentWatch solution are:

✓ End-to-End Deployment

ContentWatch has taken great care to ensure that our customers can constantly and reliably enforce their Internet usage policy, regardless of the size of their network. Only an end-to-end solution can ensure that all Internet access is regulated by your organization's the Internet usage policy. The intelligent client-side filter that turns itself on and off as computers attach and disconnect from your corporate LAN provides the peace of mind that all of your corporate resources are protected without causing undue latency to the end user.

✓ Single Point of Administration

The ContentWatch Data Center provides a single point of administration for all of your Internet filtering needs. Policies can be managed and reports may be viewed from anywhere, at any time. Furthermore, settings are automatically synchronized across all of your clients and appliances so that your current Internet filtering policy is automatically enforced across all your corporate resources.

✓ Dynamic Contextual Analysis

The ContentWatch Dynamic Contextual Analysis engine brings is the most powerful weapon against inappropriate Internet usage in the world today. It is the only way to accurately and consistently enforce a corporate Internet usage policy. It powers the number-one-rated content filter application on the market today. Coupled with end-to-end deployment to manage that Internet usage policy at both the gateway and the client, ContentWatch provides the safest, most comprehensive, and fastest-response web filter on the market today.

For pricing and availability information, please visit <http://www.ContentWatch.com> .