



A WEBSENSE® WHITE PAPER

## The Web Isn't Fun Anymore: How Websense Technology Protects Against Internet-Based Threats

**Abstract:** The Internet—with its wealth of information and features that are integrated into our everyday lives—has become a necessary tool for business and provides a vast array of options for personal use. However, it does have a dark side. Over the past several years, the Internet has become an increasingly dangerous place to work and to play, with viruses, Trojan downloaders, phishing, pharming, and malicious Web sites designed to take advantage of careless users. Attacks continue to increase in frequency and sophistication. Fortunately, Websense has developed cutting-edge technology to prevent infections from happening in the first place and to guard against zero-day attacks. A colleague summed it up best when he said, “The Web isn’t fun anymore.” It is clear that in today’s business environment you cannot survive without using the Internet. For this reason, it has become critically important to have the necessary safeguards in place to ensure the safety of your company data and your employees.



<b>Table of Contents:</b>	<b>Websense ThreatSeeker™ Technology .....</b>	<b>3</b>
	WebCatcher, KILO, and the Websense Master Database .....	3
	AppCatcher and ProtocolCatcher .....	4
	Honey Pots .....	4
	Honey Clients .....	4
	Protecting Against Honey Pot and Honey Client Identification .....	5
	Web Data Mining .....	5
	Identified Attacks on the Websense Security Labs Alert Page .....	6
	Search Engine Poisoning .....	6
	Web Reputation .....	7
	<b>Formal and Informal Partnerships .....</b>	<b>7</b>
	<b>Protecting Remote Users.....</b>	<b>7</b>
	<b>Summary .....</b>	<b>8</b>
	<b>About Websense, Inc. ....</b>	<b>8</b>

## Websense ThreatSeeker™ Technology

Websense ThreatSeeker technology collectively defines all the technologies that Websense software utilizes to neutralize Internet-based threats. ThreatSeeker uses a variety of techniques and technologies to proactively identify new emerging and existing threats, and block them before they can infect an organization's computers. Typically, the first time Websense prevents a machine from getting attacked, it will pay for itself, by way of preventing the lost productivity, data loss, and downtime that the security breach may have caused. This whitepaper will examine how ThreatSeeker technology guards against Internet-based threats.

## WebCatcher, KILO, and the Websense Master Database

The maintenance of the Websense Master Database is one of the major pillars of Websense Technology. New Web sites are activated constantly, and as a result these sites must be categorized and any potential threats must be identified. One method that Websense uses to keep its database current is the company's WebCatcher Technology. As customers' local databases are updated daily with the latest list of categorized Web sites, WebCatcher reports back to Websense any requests for Web sites that are uncategorized in the current Websense database. These uncategorized sites are validated and consolidated and placed into the Websense production process.

In the Websense production process, new uncategorized sites are reviewed for valid form and are tested to ensure they resolve to a valid IP address. An automated process downloads all of the content on the valid sites and the content is passed through the Knowledge Indexing Learning and Organization (KILO) site classification software. KILO has artificial intelligence features that are constantly adapting to new Web sites and classifications. The site language is determined, then filters search for terms on the site that will assist in its classification. After the site is processed through the KILO software, it is assigned a numeric value for each of the 90 categories in the Websense Master Database. A negative number indicates the site is not in a specific category and a positive number indicates it is in that category. Additionally, a higher (or lower) numeric value assigned by KILO indicates a confidence factor. If the numeric rating is a large positive value, KILO is very confident that it falls into a certain category. Conversely, if KILO assigns a large negative number, it's very confident that the site does not fall into a certain category. This rating system is very similar to values assigned by anti-spam vendors to indicate a confidence factor that a message is spam.

After KILO assigns values to an uncategorized Web site, a Websense Analyst reviews the site. The Analyst may completely agree with KILO's classifications, or may make adjustments to the classifications as necessary. If the classifications are adjusted, KILO's parameters are automatically adjusted to reflect the classification preferences of the Websense Analyst. Any future classifications performed by KILO will take into account the latest preferences of the Websense Analyst. This helps KILO fine-tune its classifications, so Websense Analysts are required to make fewer manual adjustments. The human review is done with the aid of the Web Analyst's Workbench, which is a customized browser that allows the analyst to explore a Web site beyond the pages addressed by the site and verify or change the category assignments made by KILO. Websense employs Analysts who are fluent in all major languages found on the Web sites, to ensure each site is properly categorized.

After the new sites are categorized, they are included in the next daily update pushed to Websense customers. For particularly sensitive categories, including security categories, updates are sent as needed throughout the day through Real Time Security Updates™.

Of course, after Web sites are categorized, their content can change. An integral part of the maintenance of the Websense Master Database is to ensure that the assigned categories remain accurate. After a site is reviewed, it is placed in an aging process to ensure categorized sites are reviewed on a regular basis. This is especially important for sites that have been compromised and categorized as malicious. A compromised site can be repaired, so regular reviews of categorized sites are important so sites do not generate false positives after the malicious code is removed. Compromised sites with heavy traffic may be tested on an hourly basis so the Websense software allows access as soon as the site is clean of malicious code. This protects users from potential infections, but will allow access to the site as soon as the threat is addressed.

### **AppCatcher and ProtocolCatcher**

AppCatcher and ProtocolCatcher work similarly to WebCatcher, but instead of categorizing Web sites, they categorize applications and protocols. Websense software logs each attempt to run an application, and also notes whether an application attempts to use any ports or IP addresses. Applications and their related protocol ports and IP addresses not categorized are sent back to Websense in an AppDigest. Each new application is identified, analyzed, and categorized and then is included in the daily download to customers. This allows Websense to identify potentially dangerous applications, while still preserving the anonymity of its customers.

### **Honey Pots**

Honey Pots—unpatched computers intentionally placed on the Internet so they can become infected—allow Websense to monitor and track the evolution of a particular attack and how it evolves. Hackers often use unpatched computers as test environments to determine how effective a particular exploit may be for a specific or group of vulnerabilities. These exploits may incorporate multiple morphed variants of the same exploit in an attempt to bypass existing virus scanners and other intrusion detection/intrusion prevention systems. If a hacker is targeting a particular company or organization, he or she may perform automated scanning techniques to identify specific weaknesses in the targeted company's network. Then custom packages are assembled to exploit the identified weaknesses to gain access.

As these Honey Pots are infected, the exploits are forensically examined to determine the type of attack and how to best prevent it. Often, emerging attacks are identified early in the Proof of Concept (POC) development stage, allowing countermeasures to be implemented before the attack becomes widespread.

### **Honey Clients**

Honey Clients are different from Honey Pots, but the end goal is still the same—to identify threats before they become active. Honey Clients are fully patched computers that actively visit different Web sites in an attempt to become infected. Honey Clients are a very effective strategy against zero-day exploits. Zero-day exploits are vulnerabilities in an

Operating System or Application that have not yet been identified by the manufacturer and, therefore, do not have a patch that addresses the vulnerability. In an attempt to infect a computer, a hacker may place malicious code on a Web site that can take advantage of a weakness in the computer's Operating System or Application. This malicious code is unknowingly downloaded and then infects the computer that visited the Web site. Zero-day exploits are particularly dangerous because they can potentially infect any visitor, regardless of the patches installed on their computers. Once these Honey Clients find an infected site, the information is updated in the Websense database and pushed out to Websense customers.

Zero-day exploits are always a concern for network administrators because they are very difficult to defend against. But, the Websense Honey Client is a very effective way of identifying any zero-day exploits by actively going out and identifying potential exploits as they are under development. The Honey Pot and Honey Client protocol that Websense has established gives users a huge advantage over hackers, as Websense is actively seeking out sites distributing malicious code, something the typical administrator simply does not have time for.

### **Protecting Against Honey Pot and Honey Client Identification**

From the hacker's perspective, these Honey Pots and Honey Clients are potential landmines that hackers would, of course like to avoid. For this reason, the underground hacking community has a list of these Honey Pots and Honey Clients, and actively updates it. Some hackers actively go after known Honey Pots and Honey Clients by generating code that produces false negatives when a Honey Client visits an infected Web site. Malicious sites can be coded to present different versions of a site based on the client IP or other identifying factor. For example, a hacked site may present a flower shop Web page when viewed by a Honey Client, but may present a pornography site when visited by a non-Honey Client. Websense is well aware of this issue and has a number of countermeasures in place to prevent Honey Pot and Honey Client identification. However, the specific methodologies cannot be discussed, because it would give a significant advantage to the hacking community. Security vendors and the hacking community will continually play this cat and mouse game for the foreseeable future.

Even when a Honey Pot or Honey Client is not identified, a Web site can look significantly different based on the location of the Honey Pot or Honey Client. For example, a Web site in China may look entirely different based on the (perceived) location of the client. A Web site located in China may have completely different content when it is surfed from a computer located in the United States versus a computer located in China. A site could also be served from an entirely different Web server based on the location of the computer requesting the Web page. Websense has taken this into consideration, so identical sites are surfed from multiple computers in different locations to ensure that all versions of the web site are not malicious and have not been compromised.

### **Web Data Mining**

Websense maintains an array of machines that constantly perform data mining on Web sites. Information is gathered on approximately 90 million Web sites every 24 hours. To guard against false positives and to remove sites from the blacklist after they have been cleaned, every site is revisited at least every 12 hours. An exception to this rule is larger

sites that may have become infected. These sites are checked as frequently as every 20 minutes. These machines are actively looking for sites that contain known infections and/or have attributes and/or activities that are highly suspicious. All of the collected information is stored in back-end databases that tie into a real-time threat list that is downloaded by Websense customers. It's a very similar approach to a Real-Time Black List that is used by anti-spam companies to identify potential sources of spam, except that this list identifies potential lists of malicious and compromised Web sites.

When a site is compromised, often a hacker will modify every single Web page on the site, not just the home page. When a compromised Web site is identified, all pages are visited on the site, not just the home page. All pages on the site must be free from malicious code before the site is removed from the black list.

Internet attacks can come from a variety of sources, especially email. Fortunately, anti-virus and anti-spam vendors have done a good job of protecting companies from this type of attack. Hackers are aware that it has become increasingly difficult to infect a computer using an email-based attack, so they have recently focused their efforts on creating malicious Web sites, compromising existing Web sites, phishing, and pharming attacks. These types of attacks tend to be more effective, primarily because they are more difficult to defend against. The Web Data Mining attack guards against this type of attack by proactively seeking sources of these attacks and blocking them in real time.

#### **Identified Attacks on the Websense Security Labs Alert Page**

Websense lists identified attacks on its alert page located at <http://www.WebsenseSecurityLabs.com/alerts>. Websense discovers new attacks daily. Recent attacks as of this writing include the IRS Scam phishing attack, which downloads a new version of a Trojan horse that is undetectable by anti-virus vendors. The Websense Security Labs site features forensic information on each high-profile threat, including how they found each threat, how it propagates itself, and where the attack originates. Visit this page to gain an understanding of the effectiveness of the Websense threat identification strategy. The site features Really Simple Syndication (RSS) feeds so you can keep up-to-date on recent attacks using your favorite RSS reader.

#### **Search Engine Poisoning**

Another tactic that hackers use to draw unsuspecting computers to malicious Web sites is search engine poisoning. Using commonly entered keywords, a hacker will display search hits that give a high rank to compromised and malicious Web sites. Hackers use this type of attack in conjunction with DNS poisoning which can redirect users to compromised Web sites by returning an incorrect IP address to a DNS query. To combat this issue, Websense uses the Honey Client to perform searches using popular search engines located around the world. Based on the search engine results, Websense visits these Web sites to determine if they are malicious, then updates its malicious site database in real-time.

## Web Reputation

Although the method of investigating a Web site's reputation has been around for quite a while, Websense Enterprise® and Websense Web Security Suite™ employ the next generation of this method in version 6.3.1, released in May 2007. Approximately seven to eight million new domain names are registered on the Internet every week. As you know, you must register a domain name with a Web registrar, such as Verisign, before you can create a DNS A Record to point users to the Web site. Hackers will often use aliases and other misleading information to make it difficult to track the “real” user down if the Web site turns out to be malicious. A common tactic is registering domain names that are very similar to existing legitimate sites, hoping that users will misspell the domain, not realize they are not on a desired site, and then be tricked into giving personal information on the rogue site.

Using data collected from domain registrar companies, Websense utilizes data mining techniques to identify trends and patterns to detect potential malicious Web sites. For example, if a particular alias is used to register a new site and that alias was used to register a previously known malicious site, Websense will place the registered URL in the block list. When and if the site becomes active, it will be tested for malicious activity and will either be removed or kept on the block list depending on the results of the scan. This technique enables Websense to block potentially malicious sites before they even become active on the Internet.

## Formal and Informal Partnerships

In addition to the research that Websense performs internally, the company has formed both formal and informal partnerships with different organizations. One of the formal relationships is the Anti Phishing Working Group (APWG <http://www.antiphishing.org/>). The APWG is a global law enforcement association that focuses on eliminating fraud and identity theft from phishing, pharming, and email spoof attacks. Websense shares information on attacks occurring at specific times. Raw data is shared and in some cases accepted for future analysis. Websense is a committee member of the APWG and is taking the lead on the Phishing Data Repository project.

Websense has established relationships with other security researchers, law enforcement, and other organizations. Information is transmitted via newsgroups, IM, chat, and secure email. When new attacks are identified, Websense will alert the security community of the nature of the attack and outbreak events.

## Protecting Remote Users

Providing a secure computing environment for a remote user has always been problematic. Almost every company has users that must travel and who require remote access to the corporate network. Telecommuters also pose a threat, because they are offsite and do not reside behind the corporate firewall and often are not up-to-date with the latest patches. Remote access is typically secured with a Virtual Private Network (VPN), which is very difficult to attack directly. Knowing this, hackers often go after poorly protected endpoints – like a remote user's laptop. If a hacker manages to install malware on the laptop that has remote control and keystroke capture capabilities, he or she can establish a tunnel back to the corporate network and further compromise systems that reside in the corporate office.

Another concern is the road warrior who manages to get his or her laptop infected while traveling, turning it into the "Typhoid Mary" laptop. Just like the original "Typhoid Mary," the user doesn't have any idea that they are infected. When the user returns from a trip, he or she connects to the corporate network and immediately start infecting other computers. Because the user is typically not aware of the problem, it may take time to determine the origin of the original infection, giving more time for the malware to spread.

Websense protects remote users even when they are disconnected from the corporate network, giving these users the same protection they have when they are connected to the corporate network. An agent is installed on the remote user's laptop. When a user connects to the Internet, the laptop is authenticated to the Websense server located in the corporate office. After the authentication is completed, all Internet traffic passes through the Websense server, effectively giving them the same protection as an internal corporate user. Just like an internal user, the remote user's Internet traffic is monitored for mobile malicious code and will prevent the user from visiting inappropriate or potentially dangerous sites. Regardless of the user's location, an administrator can require the Websense agent as a prerequisite to surfing the Internet. Figure 1 illustrates this architecture

## Summary

Defending against Internet-based threats is becoming more and more difficult. But, Websense has cutting-edge technology that provides effective protection against existing and emerging threats on the Internet. Monitoring and identifying malicious Web sites and removing them from block lists when they are clean requires a monumental effort, but Websense customers are able to leverage this sophisticated technology to protect their company data and users from attacks. Often the first time Websense prevents an infection, the product will pay for itself. Protect your company and its employees by using the latest technology to keep the Internet a safe and efficient place to do business.

## About Websense, Inc.

Websense, Inc. (NASDAQ: WBSN), protects more than 25 million employees from external and internal computer security threats. Using a combination of preemptive ThreatSeeker™ malicious content identification and categorization technology and information leak prevention technology, Websense helps make computing safe and productive. Distributed through its global network of channel partners, Websense software helps organizations block malicious code, prevent the loss of confidential information and manage Internet and wireless access. For more information, visit [www.websense.com](http://www.websense.com).