



Secure and Manage
Your Internet Access



Anchiva Data Cloud URL Filtering

Summary

With continuous development of Internet technology and Web 2.0, SNS, blog services, Wiki, Twitter, Webmail and free spaces have mushroomed among internet applications, helping people create an easier life, more efficient production, and create a need for URL filtering technology to evolve. URL filtering is now being added to the security arsenal as a key component of protection, to ensure a high productivity and prevent internet addiction and abuse. Today's Web sites are increasingly complicated, coupled with vast user-generated data in Web and a dramatic speed of expansion. In this scenario, traditional URL filtering solutions cannot keep up with today's threats due to their technical limitation. Whereas, Data Cloud URL filtering appeared as it solved the drawbacks of traditional URL filtering solutions, providing a more correct real-time categorization with more diversified data. It provides most user need-relevant URL filtering despite of language, cultural and geographical limitations, then form a virtual community where users are capable of sharing what they have.

Needs for URL Filtering

Nowadays, internet is introducing easy ways of living and production. Vast range of web resources enable people to access most in-need information by simply browsing the internet. However, employees inappropriate access inappropriate or illegal Web sites or content may bring about potential productivity and bandwidth loss or even legal liability to enterprises. They themselves can also be impacted mentally and physically, while leaving enterprise security infrastructure and information system exposed to great menace. Such as:

1. Excessive non-productive accesses cause productivity loss

Internet accessibility makes it very easy for staff to browse non-commercial content, ranging from reading to online shopping and even online gaming and online stock trading, directly affecting the effectiveness. Meanwhile, the peers' following the suit may result in low morale among the team and significant productivity loss.

2. Download and use of bandwidth consuming applications may result in bandwidth loss plus usage peak.

Open internet access provides easy download of audio and video material. The popular usage of P2P, especially the bandwidth consuming application, occupies the bandwidth resources and causes usage peak, affecting the normal business operation and internet service.



3. Unwitting access to malicious sites or URL imperils the enterprise network

Nowadays, websites has fallen as the primary channel of virus and malware distribution. According to Gartner, throughout the first quarter of 2008, 50% of legitimate websites have ever been hacked for malicious behaviors. Employees' web access may be unwittingly attacked and make the whole network at risk.

4. Inappropriate web browse brings harm to personal health.

Employee access to the inappropriate content, such as erotica, obscenity, evil, and violence, can be harmful to personal health in both terms of mental and phisical, and even lead a way of crime.

5. Inappropriate content browse may bring about the legal reliability

Inappropriate content include anti-social utterance and BBS, ammunition production, political extremism literature, hacker tools. Employee access to these contents may cause lawsuit to companies.

What shall we do in the face of all these problems? Completely blocking access to internet is apparently unwise for the sake of advancement of modern human civilization and business expansion. A self-enforced use policy has become inadequate to protect enterprises against these threats. A new generation of URL filtering solutions is demanded to realize a management and control on employees internet access.

Development of URL Filtering Technology

In the mid 1990s, URL filtering solutions relied on black lists (prohibited) and white lists (approved) to control access to Web sites. While the lists of good/bad URLs were built, updated and edited locally by each organization's IT department, in most cases they were resources intensive and generated by one or a group of persons, and thus suffered a notable lack of objectiveness. This results in inaccuracy in site categorizations that straightly caused many approved sites to be blocked while allowing some prohibited sites through the filters. As the Web grew and the number and complexity of Web threats increased, a black/white list approach was no longer a viable approach for organizations seeking a reliable filtering solution since it cannot provide a a objective granular URL classification.

In the late 1990s, there came to be vendors providing professional URL collection and categorization services. URL filtering technology began leveraging category engines coupled with a local database. Using this technology, URLs and their contents were analyzed and classified under a predefined category (e.g., gambling, pornography, e-shopping,) stored in a centralized master database and then transferred in batches to local customer databases which were exact replicas of the master database. However, the one-fits-all model of key word-based categorization and limited capacity failed to provide a higher accuracy in in-depth site classification as sites have become more dynamic and complex.

In early time of year 2000, URL filtering solutions began to employ heuristic engines as a means of analyzing the flood of dynamic content and Web sites on the fly. These dynamic engines examine probability — i.e., how likely it is that the text in a particular site signifies the type of site it actually is — in order to determine whether or not a site should be blocked. These solutions solve the main problems inherent in the former two technologies, but they are still limited by a fact that these heuristic engines are not capable of sending the real-time web content analysis results to endusers, instead, still storing them in local database. In Web 2.0 era, the Web is an almost *infinite* collection of data with a high speed of growth. The Internet's vast size coupled with the unique and specific needs of individual customers has created the drawbacks for database URL filtering solutions in realizing a fast and accurate detection as they soly filter and store data that the local users need and failed to provide all relevant latest information. They have fallen far away behind the Web 2.0 technology.



	Mid 1990s Manually categorized Black/Whitelist	End of 1990s Local Black/Whitelist	Early 2000 Heuristic detection categorization	2009 Cloud-based URL Filtering Technology
Categorization Method	Manual categorization by IT staff	Keyword inquiry categorization engine	Heuristic Keyword probability categorization	Complete web content and context analysis
Storage Method	Black/Whitelist file	Local database	Local database /server group in the cloud	server group in the cloud /local cache
Update method	Manually edit and update Black/Whitelist	Keep database copy	Servers deliver updates periodically	Clients get URL classification updates at anytime
Accuracy	Poor	Ordinary	Better	Excellent
Coverage	Poor	Ordinary	Better	Excellent
Summary	Not objective. Resources-intensive. Inaccurate.	High rate of false positives and missing reports. Internet grows overwhelmingly out of the local storage capacity.	Servers don't provide accurate categorization in real-time. users cannot get data in time. Limited capacity	Unlimited performance and storage capacity. Self-enrichment model satisfy unique and specific individual need

According to Google, data on the Web is growing at a speed of 100 million pages per day. The three solutions above have not been enough to provide correct collection and classification to all URL categories. Higher requirement in data storage and processing performance created need to break limitations of local database capacity. In 2009, security vendors like Commtouch and Anchiva presented a more advanced technology which can be called "Data Cloud URL Filtering". This does not rely upon the limited resources of an on-premise database for analysis and detection, nor is it dependent on database updates for the latest available information. Instead, it provides URL collection, processing and distribution based on cloud, where professional group of categorization servers locate to conduct comprehensive webpage content analysis according to real web use. It differs from traditional cloud that the userpoints are capable of auto-inquiring needed information from servers in the cloud rather than leverage simple periodic updates. In the following text, we'll give a brief introduction to Data Cloud URL Filtering with Anchiva Data Cloud Filtering as an example.

Anchiva URL Filtering Technology

Anchiva Data Cloud Filtering technology consists of two parts: Anchiva secure web gateway deployed at the board of enterprise network and Anchiva Data Cloud URL Categorization Center. These two elements get latest URL categorizations by real-time communications, breaking limitation of traditional local database. Unlike other cloud-based technology, Anchiva secure web gateway integrates URL cache which stores most relevant URL for each individual user. It can enrich itself as users generate new requests so as to ensure providing more accurate URL filtering that users rightly need



See how Anchiva secure web gateway processes URL filtering to HTTP-GET requests.

1. URL matching engine of Anchiva secure web gateway receive HTTP-GET requests from local users.
2. URL matching engine looks for relevant URL category from local cache in Anchiva secure web gateway.
3. If right category is found, the HTTP-GET request will be allowed or blocked according to predefined policies.
4. If right category isn't found, URL matching engine sends HTTP-GET request to Anchiva Data Cloud URL Categorization Center.
5. An auto inquiry conducted and right category is returned to URL matching engine.
6. With feedback from URL Categorization Center, HTTP-GET request is allowed or blocked according to predefined policies for the category.

Anchiva Data Cloud URL Filtering Processes