



BLUE PAPER

How to use filtering solutions to increase productivity

Introduction

Companies lose money every day due to the fact that **their employees engage in non-work-related Internet surfing**; public education institutions and families argue that it becomes more and more difficult to control the type of information that children have access to.

On top of all that, **rapid changes occur every second in the mainstream Internet**, making it an enormous challenge for both institutional and individual users to keep pace with new developments and protect their computer systems from Internet misuse. The question has become more and more pressing how to lead the battle of Internet content filtering and win.

The present document will try to answer this question by presenting an innovative and flexible product – Livigent – that **changes radically the approach on Internet content filtering**.

Contents

Introduction.....	1
What is Internet content filtering and why do we need it?.....	2
Main challenges for Internet content filtering.....	2
Challenges for corporate users.....	3
Challenges for public institutions, schools and libraries.....	3
Challenges for individual users.....	3
Current filtering solutions.....	4
How current filtering solutions work.....	4
Considerations to be made when choosing the right filtering solution.....	5
Modern and real-time filtering: Livigent.....	6
Livigent – Key features.....	6
Corporate benefits provided by Livigent.....	6
How Livigent works.....	7
Conclusions.....	8



What is Internet content filtering and why do we need it?

In today's vocabulary, Internet has ceased to mean only an infrastructure of servers and computers, and it is more associated with the idea of content – the things that people can actually see, hear and download from the Internet. Due to the progress in information technology – from home computers, to more complex institutional systems – that content which is available to be found or downloaded from the Internet becomes more attractive to people. However, is this 100% a positive thing? There are some characteristics of the Internet that can turn it, from an information source and educative tool, into a way of losing money – for both companies and individual users:

-  Continuous development of technologies;
-  Physical and virtual expansion;
-  Lack of a central regulator¹;
-  Virtual anonymity;
-  Public access.

That is why Internet content filtering is necessary, in order to create a barrier (either by software, or by software/hardware means) between the Internet and the user, blocking content that is offensive, inappropriate or in other way damaging. Internet content filtering covers more than the main communication protocol associated with the Web – Hyper Text Transfer Protocol (HTTP) - , it addresses also other types of protocols, focusing on instant messaging protocols (for example, Yahoo! Messenger, Windows Live Messenger, Skype, AIM, ICQ, IRC, Google Talk), and peer-to-peer (P2P) file sharing protocols (such as BitTorrent, eMule, Kazaa, Direct Connect).

Main challenges for Internet content filtering

There are three major types of actors that face big challenges as far as Internet misuse is concerned:

-  Corporate users;
-  Public institutions, schools and libraries;
-  Individual users.

Each of these actors faces different types of threats, as follows:

¹ Internet Society (ISOC), ICANN (Internet Corporation for Assigned Names and Numbers, Internet Assigned Numbers Authority (IANA), Internet Engineering Task Force (IETF), Internet Architecture Board (IAB), are some of the bodies that oversee the development of the Internet, mostly from a technical point of view. They do not act as content or policy regulators.



Challenges for corporate users

In order to better understand the consequences of Internet misuse at work, let us take a look at a survey² first. 87% of the employees caught in this survey surf non-work-related websites while being at work: chatting with friends, downloading information, looking for another job, or just reading the weather reports. What does this mean for a company, in terms of money? There are two main consequences of this Internet misuse:

- ▶ Company productivity decreases when the employees are engaged in non-working activities on the Internet: the average employee spends 24% percent of his Internet surfing time looking for information in his own benefit, not the company's.

For example, according to a research by Panda Labs, it is estimated that in 2005 UK employers lost around €300 billion from an average of almost an hour a day of personal Internet using while at work.

- ▶ Company resources are misdirected when the company Internet bandwidth is used for inappropriate content access, such as sexually explicit or job-hunting websites. Some departments are thus prevented from using resources to their full capacity for work-related issues.

All the new ways of entertainment that the Internet provides can also cost companies a lot of money because of legal liabilities they might face. Employers can be considered liable for their employees' actions, whenever they commit wrongful acts in the course of their employment, such as online gambling or visiting porn websites. These legal risks include: liability for obscenity, harassment, defamation, hosting of illegal materials on the company's network. Again, all this liability issues can translate into a lot of money, payable by the employer.

Challenges for public institutions, schools and libraries

The public and educational sectors have their own specific content-related risks, especially when it comes to children. The most common risks involved relate to children visiting adult sites, anonymous chats or copyright infringing websites. A few figures are relevant in this respect: according to a survey by toptenreviews.com, in 2006 the average age of first Internet exposure to pornography was 11 years old. Furthermore, the same study shows that 90% of children with ages between 8 and 16 years have viewed online pornography, most of them while they were supposed to be doing their homework.

Challenges for individual users

Is Internet misuse a very important issue even at home? The answer is definitely yes! There a few risks may arise:

- ▶ parents can be sued for their children's misuse of Internet, such as breaking copyright regulations;
- ▶ online shopping is a risk when children get hold of their parents' credit card;
- ▶ pornography exposure is another issue to take into account when deciding for an Internet content filtering solution.

2 Vault 2005 Internet Use in the Workplace Survey



Current filtering solutions

Filtering solutions started to appear a few years ago, due to advances in technology and decrease in costs. There are two major ways of filtering: software solutions (these are packages deployed on the desired platforms, such as Microsoft Windows, Apple Mac OS, GNU/Linux etc), and software/hardware solutions (which are self-contained appliances, running custom and specific software for the sole purpose of filtering).

Appliance vs. software: side-by-side comparison	
Appliance (hardware)	Software solution
platform independent (self-contained solution; functions with any platform on the market)	platform dependent (specific setup for each type of operating system)
easy deployment (physical setup; the appliance's administrative Web interface is easily accessible from any computer on the local network)	per-user or per-server deployment
dedicated hardware (the hardware resources available in the appliance are entirely used only for filtering)	shared hardware (the hardware resources of the hosting computer – either a server, or a user-machine – are affected by the filtering)
hard to circumvent (eventually, only by physical disabling of the appliance)	circumvention is dependent on other factors (operating system security and configuration, third-party products)
scalability proportional to the number of installed appliances	scalability depends only on the host computer's performance

How current filtering solutions work

The classical approach to content filtering is to block certain websites based on their Internet address (URL). Sites are being categorized either by people who review each site and each page, or by automatic processes. URLs associated to websites with inappropriate content (such as adult, gambling or entertainment) are stored in a proprietary database, with the disadvantage of having to update it very often.

When doing the filtering, the decision process takes several steps:

-  The user requests an URL (by writing an address in the browser, or by clicking on a link);
-  The filter checks for that URL in the current database;



- If the URL already exists in the database an specific action (allow, block, or ask) is taken;

If the URL does not exist in the database, it is reviewed and classified (by real persons, or by an automated process; depending on each solution) and the database is updated to reflect the results of the analysis. Only afterwards, an action is taken.

These classical solutions have proven to be approximately 90% accurate, still, there are some serious disadvantages to be considered, related especially to the constant expansion of the Internet:

- An URL can be changed, or a domain can be sold, thus the content associated with that URL changing radically;
- Classical filtering solutions are seriously challenged by the fact that major search engines use caches, which provide the same contents as the cached sites and are accessed through URLs different from the ones of actual sites;
- Similar problems may arise from various translation services (especially babelfish.altavista.com or www.google.com/language_tools), which provide access to potentially blocked pages through different URLs than the original ones;
- Many sites offer constantly updated content, that doesn't fall into a certain category, and blocking the entire site would be excessive, while keeping track of each newly submitted item would be counterproductive;
- Current filtering solutions typically block pages according to predefined categories, while in real-life cases Internet content can belong to several categories in the same time, in different amounts.
- Since classical content filtering solutions are based on an URL database, the unknown URLs have to be reviewed and classified first (by humans or automatically). This leads to a certain delay in either blocking or allowing an URL, ranging from a few seconds, to the day of the next product update.

Considerations to be made when choosing the right filtering solution

All this factors above being considered, there are some key features to be taken into account when choosing the right filtering solution:

- Accuracy of detection, indicating if the damaging content is being blocked, and not the harmless one;
- Range of detection, showing what type of contents are recognized and filtered;
- Type of filtered users - some solutions may be filtering only IP addresses or networks, while others may handle other, specific types of users (like proxy or Microsoft Active Directory users);
- Required time for the detection – several seconds for each accessed page are too long;
- Interaction with other products – a good filtering solution should adapt to whatever products are in use (operating systems or hardware);
- Scalability – the ability to adjust to expansion of the networks;
- Logging and reporting features – a good filtering solution should be able to log its activity as well as that of the filtered users in order to provide the administrative staff



all the required information;

- Flexibility and usability of the management interface – as complex as a system might be, its potential will not be fully achieved if its interface is difficult to use.

Modern and real-time filtering: Livigent

As previously shown, classical URL-based filtering solutions have some notable disadvantages, due to factors like: the increasing number of available URLs, frequent dissociation of content from URL, the need to review and classify each unknown URL, and the need for a large and proprietary (closed) database of URLs.

In order to overcome these issues, Livigent's Internet content filtering decision process does not primarily try to allow or block a webpage based on its "label" (URL or keywords). Instead, the key feature that makes Livigent different is that it analyzes unknown content in real-time, and then it delivers the filtered content in real-time. The main consequence is that there is no proprietary database to be consulted, and thus, there is virtually no waiting time.

Livigent – Key features

Before explaining how Livigent actually works, let us take a look at its main characteristics, which will be explained in the following section of this blue paper:

- Multiple filtering methods.** Livigent offers text, image, URL, MIME type and protocol based filtering.
- Real-time filtering.** Our solution uses content-based filtering built on modern technology that enables instant, real-time responses.
- Image altering.** Livigent alters or replaces pictures that contain colors similar to skin color.
- Bandwidth management.** Our solution allows you to specify the maximum inbound bandwidth available per user.
- Reporting.** Livigent provides detailed logs and statistics on the local traffic logged into the system.
- Multilanguage support.** The application can be used in both English and Romanian.
- Ease of use.** Without compromising form-filtering performance, our solution is intuitive and easy to manage by people working in a non-IT-related field.
- Reliable.** The hardware appliance is stand-alone, stable and scalable.

Corporate benefits provided by Livigent

We have previously shown the main corporate challenges related to Internet misuse by employees - decreased productivity, available bandwidth problems, and legal liability issues. To all these problems, Livigent offers a highly customizable solution, guaranteeing employees will only have access to work-related websites.

The benefits associated to the use of Livigent's Internet content filtering solution are closely related with the companies' needs for increased efficiency and security. Major benefits include:

- Increased productivity** - employees are stopped from engaging in non-work-related activities during working hours;



- **Guaranteed access for each department** - proper internet information support is brought to each department, with adequate resource management;
- **Preventing legal-liability issues** – no more illegal file sharing and legal issues;
- **Gaining visibility over web usage** – very good awareness of what is going on in your network;
- **Gaining control** – you will know at all times what your employees are doing, what's going on in the network and how your resources are used.

How Livigent works

The most important characteristic in Livigent's functioning is the fact that it has multiple layers of Internet content filtering, which are explained as follows:

Content based filtering

Filtering is based on the real, live content of each filtered page, by analyzing each web page that is requested, and then taking an action. The advantage of Livigent is that it is able to determine in real-time, based on specific words present in the analyzed page, the types of content and the amount in which it is present in a web page. Currently, Livigent recognizes content from 21 categories – Adult, Entertainment, Shopping, Social networking, Financial, Chat/instant messaging, Gambling, Government, Illegal, Health, Humor/jokes, Lingerie, News, Games, Dating, Anonymizers, Search engines, Free web hosting, Webmail, Misc, Unsupported language).

How does content based filtering work? As opposed to URL based products (that can only filter by category type), Livigent filters by the amount of content that fits into the defined filtering categories. The consequence is the ability to handle mixed content in a flexible and customized way. Livigent uses an advanced system of thresholds³ that specify the maximum amount of allowed content from each category. For example, if pages from the News category are required to be blocked, a filter is set, together with an amount (threshold) of news content. When the website administrator selects the option "Sensitivity", the more the slider is placed to the right (most sensitive), the less content is allowed.

Image altering

A serious problem arises when search engines return images with no text. In the case of classical filtering solutions – that analyze only text – these images might still be shown. Livigent uses a skin replacement threshold that indicates the percentage of skin color in a picture from which an altering action will be performed, thus altering the pictures significantly.

URL filtering

Livigent allows for custom pattern oriented URL filtering by providing support for URL blacklists (lists of blocked URLs) and whitelists (lists of allowed URLs). Two filtering items are used for URL filtering: URL glob lists and URL regexp lists. An URL glob list is an unordered list of URLs containing wildcards or globs, while an URL regexp list is an ordered list of regular expressions to be matched against the accessed URLs. The regexps offer a higher degree of control over the filtered URLs. Livigent provides a regexp validation mechanism that

³ A threshold is an administrator-defined level, expressed as a 1 to 100 integer which when reached indicates that the analyzed page fits into a certain category.



prevents erroneous input.

MIME type filtering

Multipurpose Internet Mail Extensions (MIME) are an Internet standard used in context of e-mail, but also for the HTTP, its purpose being to support non-text files (like images, binary files etc). Livigent uses MIME types in order to identify the content of the transferred files. Livigent is preconfigured to filter six content type categories (documents, archives, images, audio, video, and executables). Furthermore, the system administrator has the possibility to to define custom MIME types. This is a flexible way to control downloads and it is very efficient when blocking viral video sites (like www.youtube.com), download sites (like softpedia.com) or image galleries sites (like flickr.com).

Protocol filtering

Livigent also features filtering support for instant messaging protocols (like Messenger) and peer-to-peer file sharing protocols (like BitTorrent or FTP), and more updates are being tested right now. This comes very much in handy when issues are raised by activities such as file sharing, which currently affect the entertainment industry. This type of filtering offered by Livigent is also very useful for libraries, schools and other institutions, trying to reduce risks of liability and bandwidth consumption.

Time-controlled filtering

Time flexibility in filtering is fully supported by Livigent. This means, for example, that if an employer wants some specific filters to be active only after normal working hours, he or she could use a timeline filtering item containing the filtering options enforced only in the normal working hours. This timeline can be configured in detail: start time, end time, a list of weekdays when it will be enforced, and an ordered list of filtering items.

Multilingual support

The need for filtering content in other languages than English is constantly increasing. That is why Livigent has the possibility to filter contents from two languages:

-  English;
-  Romanian.

Unsupported languages are also detected and categorized separately, thus providing the possibility of filtering.

Logging and reporting

Based on the logged data, Livigent provides advanced options for logging and reporting, allowing recording of statistical data regarding the user activity and featuring in-depth reporting.

Conclusions

An efficient content filtering solution must be, first, accurate and reliable. Today, URL filtering is already outdated and sometimes inefficient. In this respect, the Livigent filtering system brings some intelligent web control benefits:

-  Increased productivity and efficiency;
-  Improvement of customer relations - communication with your clients becomes more



efficient when distractions for the team members are removed;

- Managed internet access - adequate web bandwidth management ensures each department will be provided with the necessary resources;
- Comply with legal regulations - Livigent guarantees that users will have no access to illegal file sharing;
- All its key features (real-time content filtering, image altering, multiple layers of filtering, multilingual support) make Livigent a very reliable and efficient alternative to the current products available on the market.