



blackspider
technologies

A Buyers Guide to Spam Filtering

BlackSpider Technologies, 400 Thames Valley Park Drive,
Thames Valley Park, Reading, Berkshire, RG6 1PT
Tel: +44 (0) 1189 653 453 Fax: +44 (0) 1189 653 553
E-mail: info@blackspider.com Web: www.blackspider.com

© BlackSpider Technologies Limited 2004

Contents

Introduction	3
What is spam?	3
Is spam really a problem?	3
How much does spam cost the European enterprise?	4
Direct costs of spam	4
Indirect costs of spam.....	5
Spam: the legal threat	5
Why won't legislation work?	6
Spam-Filtering Techniques	7
Spam Filtering Options	8
Desktop spam filtering	9
Mail server filtering	10
Internet gateway filtering	11
Internet level managed services.....	12
10 Crucial Questions to Ask in Spam Filtering	14
MailControl Spam	17
Section i - Best practices to avoid spam	19
Section ii - Useful links	19

Introduction

'Spam' is an irreverent term for a very serious problem. For people involved with dealing with the ever increasing volume and associated cost of dealing with spam, the Monty Python joke has long since ceased to be funny.

As spam – or junk mail - continues to clog up servers, reduce bandwidth and squander man-hours, companies are realising that a more proactive approach is needed to tackle this growing problem.

This guide is designed to answer any questions that you may have regarding spam and spam filtering, and will seek to offer guidance in selecting the correct spam solution for your company.

What is spam?

The anti-spam community has long grappled with a problem of terminology. Just as one U.S. Supreme Court Justice once said of pornography, "I can't define it, but I know it when I see it", spam definitions have, by necessity, tended to be quite ambiguous.

A commonly accepted definition for spam is 'unsolicited, bulk e-mail' (UBE). However, it is important to note that a message can be unsolicited (first contact enquiries, job enquiries, sales enquiries etc.) or bulk e-mail (subscriber newsletters, discussion lists, information lists, etc.) and not be spam.

In a corporate environment, it is crucial to ensure that important and legitimate e-mails are not blocked and that legitimate marketers are able to find their customers. This presents the corporate network administrator with a specific set of challenges, and these are discussed in the *Spam Filtering Options* section of this guide.

Is spam really a problem?

People have differing views on the definition of spam, but there is one fact that everyone agrees with: spam is on the rise. Although there are no complete figures about the volume of spam Europeans receive, observers claim that up to 70% of all e-mail traffic is junk mail, and it is spreading so fast that it may render the e-mail system unusable before the end of 2005.

The volume of junk mail continues to rise in spite of the fact that the market for anti-spam products has grown rapidly over the past two years. Spammers have become more sophisticated; as techniques are deployed to successfully block spam, the spammers modify their practices to avoid detection.

Up to 70%
of all e-mail
traffic is
now spam

The crux of the problem is that there is little cost to the spammer, and every incentive to bombard as many addresses as possible. Spammers typically send e-mail to distribution lists in the millions, and therefore only need a minute percentage of users to respond to their offer in order to make spamming commercially viable. Any company that has implemented a conventional offline direct marketing campaign will know the costs associated with this type of approach; brochures and leaflets must be designed, printed, placed in envelopes and delivered. However, with e-mail the cost for the sender is virtually nil as the costs of replication, transmission, and download are borne by the recipients either directly or indirectly.

How much does spam cost the European enterprise?

Last year, US analyst Ferris Research attempted to quantify the annual cost of spam and concluded that it cost \$8.9 billion for U.S. corporations, \$2.5 billion for European businesses and another \$500 million for U.S. and European service providers. According to the European Commission figures, spam cost businesses approximately \$10 billion globally in 2001 and since that time spam volume has more than doubled. The Radicati Group, a US based analyst firm, published similar findings and predicted that this figure would rise tenfold by 2007.

Direct costs of spam

The direct costs of spam relate to the receiving and storage of unwanted e-mails.

Delivering spam costs the enterprise in terms of both Internet bandwidth and disk storage. Once received, each unwanted message has to be stored until it is read and/or deleted by the user.

25% of a
company's
bandwidth
costs can easily
be consumed
by spam

In many cases the e-mail remains in the deleted items folder on the network even after it has been discarded.

It is easy to underestimate the bandwidth costs associated with spam delivered to the enterprise: If 50% of a company's network traffic is e-mail and 50% of that email is spam (a conservative figure - Gartner estimates that it could be up to 80%) then it follows that 25% of that company's overall bandwidth cost is being used solely to download spam.

Nortel Networks estimates that each spam message received costs the company \$1. Although the cost of a single spam message to an individual user is small, the aggregate cost of spam messages directed at business today is becoming a board-level concern.

Indirect costs of spam

Spam has a negative effect on employee productivity throughout the enterprise. For e-mail administrators, their time is increasingly becoming focused on spam; how to either prevent it from entering the enterprise, or how to manage it when it does. For the end user, managing and deleting unwanted messages is both frustrating and time-consuming. It is estimated that the average end user receives 4 unwanted e-mails per day, and that previewing or opening and deleting these takes an average of 10 seconds per message. Taken in isolation these figures seem small, but using this metric, a company with 500 workers would lose 166 employee days¹ every year dealing with spam.

A company with 500 employees could be losing 166 man days per year

Spam: the legal threat

Much spam is pornographic in nature and wholly inappropriate for the workplace. Legal experts in the US now believe that where workers have complained about the e-mail content they are receiving, employers have a legal exposure if they cannot demonstrate that they are taking steps to address the issue. Spam has been cited in a number of sexual harassment laws suits where employees claim that their employer's lack of response to their complaints about receiving offensive material via the corporate mail network is indicative of a broader issue. Many believe that the UK will follow suit. Today employers increasingly recognise that they have a 'duty of care' to protect staff from offensive e-mail messages. Employing suitable filtering technology demonstrates that a company is taking reasonable steps to protect its employees.

¹ This figure is based on the following equation:
 $10 \text{ seconds each message} \times 4 \text{ messages} \times 20 \text{ days per month} \times 12 \text{ months} / 8 \text{ hours a day}$
 $= 166 \text{ man days}$

Why won't legislation work?

In spite of recent initiatives, spam is ineffectively regulated. According to the Spamhaus Project, a UK-based organisation dedicated to combating the growth of unsolicited e-mail, the UK's recent anti-spam law will actually increase the spam problem for UK businesses. From 11 December 2003 it became illegal to send unsolicited bulk e-mail to a private e-mail address, but legal to send it to the employees of British businesses. The organisation predicts that UK companies will continue to suffer the onslaught of ever more spam and that addresses deemed to belong to a business, such as sales@company.co.uk, will be rendered useless for anything but receiving adverts.

Dealing with spammers will not only require international cooperation, but collaboration between academia, business and government parties. The problem with passing anti-spam laws in isolation is that spammers simply give their business to ISPs in other, less regulated, countries. For example, there have been two anti-spam bills and 29 instances of state legislation in the US, 19 European have enacted relevant national and EU legislation, and there has been national legislation in at least seven other countries. Despite this, the problem of spam is growing faster than ever before.

For more information on spammers, please refer to the BlackSpider Technologies white paper, 'Spam: now a corporate concern.' (*Section ii - Useful links*)

Spam-Filtering Techniques

Whilst unsolicited bulk e-mail is a serious issue, the smart use of anti-spam solutions can go a long way toward eliminating the problem. Most anti-spam solutions today rely on a combination of techniques to identify spam and these include:

Real-Time Black Lists	These are collaborative services operated on the Internet by commercial organisations, or communities of interested users. RBLs contain lists of IP addresses of machines that have been black-listed because they are sources of spam.
Lexical Analysis	<p>A term used to describe the analysis of an e-mail message looking for indicators that the message may be spam, or valid e-mail. The concept of lexical analysis started out looking for key words and phrases inside the body of a message for strings that would be commonly found in spam, such as 'Buy Cheap' and 'Get Rich Quick'.</p> <p>More sophisticated lexical analysis engines now look at the whole message, including the message envelope, the message headers, the subject line and the body text.</p>
Collaborative Spam Databases	A number of Internet spam databases exist such as Vipul's Razor, that rely on a collaborative approach to identifying spam. Individual users submit spam messages to the database, where each message is given a unique signature or hash.
Bayesian Filtering	<p>Using Bayesian inferential statistics is a relatively new innovation in spam detection. The concept of Bayesian filtering is to create two databases or 'corpuses' of e-mail: A corpus of spam e-mail and a second of valid e-mail.</p> <p>Each corpus is then 'tokenised' and analysed looking for tokens that frequently appear in each type of e-mail. Each token is then given a probability weighting, suggesting if it is likely to appear in spam or valid e-mail.</p>
Spam Traps	Spam traps (or honey-pots) are e-mail accounts that have been specifically set-up to collect spam. Once the same message has appeared in a very small number of spam traps it can be clearly identified as spam, with little risk of incorrect classification. Once identified, a signature (or hash) can be created and used to detect and block future instances of similar messages.
Trend Analysis	Trend analysis can be an effective technique to help mitigate false positives and improve spam detection rates. By analysing the history of e-mail sent from an individual, trends can help assess the likelihood of an e-mail being valid or spam.
White & Black Lists	As well as online real-time black lists, there is a place for user configured white and black lists. These are configurable lists of e-mail addresses (or domains) that organisations explicitly block or allow through.

Spam Filtering Options

Which anti-spam options are available to businesses? The answer is; a wide range - each with its own set of advantages and disadvantages.

The key to building a consistently successful anti-spam strategy is to deploy the correct solution in the most appropriate place, and to manage it effectively. It is also important to educate end users on how to avoid spam (*Section i - Best practices to avoid spam*).

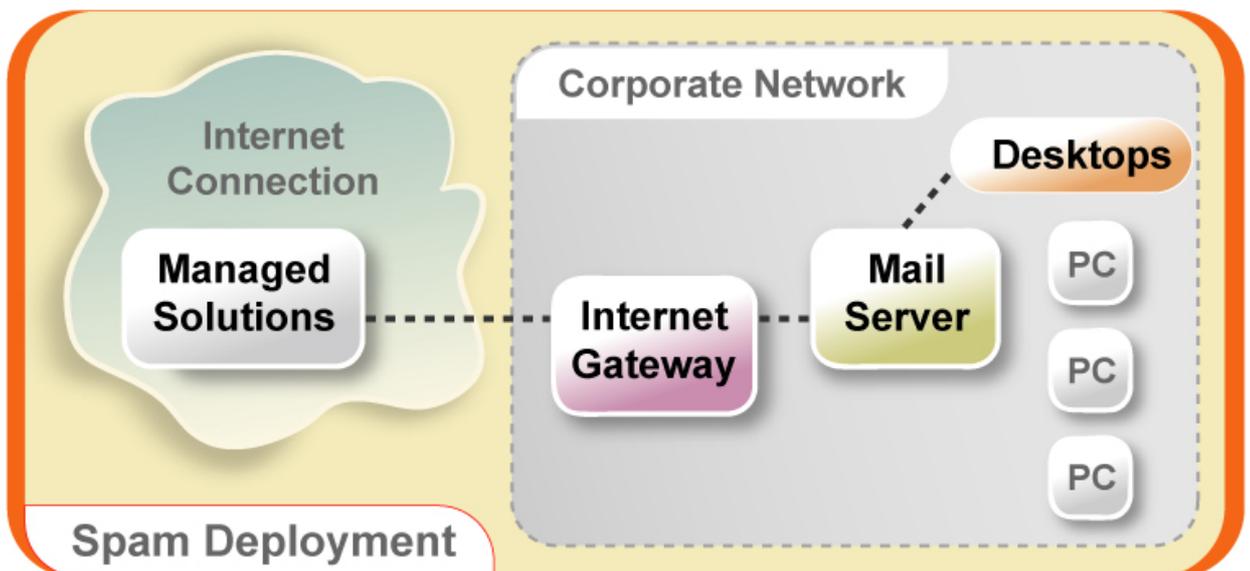
There are four different locations on the network where anti-spam filters can be positioned.

These include three inside the corporate network:

- The desktop
- The e-mail server
- The Internet gateway

And one outside the corporate network:

- The Internet level (through a third party managed service provider).



Desktop spam filtering

In this environment, all mail - legitimate and junk - is delivered to a user's desktop, where it is filtered according to the user's preferences. Desktop anti-spam products include commercial applications such as Cloudmark's SpamCop and SpamNet. E-mail clients such as Outlook 2003, Lotus Notes and Eudora now include some form of anti-spam capability.

Advantages

This method has the advantage of allowing the end user to participate in the filtering process, therefore improving the spam detection rate and reducing the number of 'false positives' (e-mail incorrectly classified as spam). Consequently, this method of spam filtering may deliver increased end user productivity. Desktop spam filtering software also has a relatively low-cost when compared with other solutions.

Disadvantages

Deploying this solution in a corporate environment involves 'touching' every desktop - consequently incurring significant overhead in support and management - and it is impossible to apply business-wide rules. Although the initial investment is relatively small, these solutions require companies to invest in hardware and network infrastructure. More seriously, this method does nothing to tackle the volume of traffic; the junk e-mail has entered the corporate network and been delivered to the desktop before filtering has taken place. Finally, this type of solution requires constant updating and tuning to remain effective.

Mail server filtering

In this scenario, all mail is delivered to the mail server inside the corporate network, where spam filtering takes place. Most e-mail servers now offer some kind of spam filtering rules, and several third party plug-ins are available for Exchange and Notes that include these capabilities.

Advantages

One clear benefit associated with this method is that there is only one set of software associated with mail server filtering. This negates the need and eliminates the costs associated with the management and support of a filtering system downloaded onto every desktop in the enterprise. Again, this solution is relatively low-cost when compared with other spam solutions.

Disadvantages

The user is not able to configure their own personal settings and end-users must rely on a third party to configure the filter. This places the burden of dealing with incorrectly blocked messages and constant filter tuning on the network administrator. As with desktop filtering software, this solution requires subsequent investment in hardware and network infrastructure and ongoing support to ensure updates and patches are implemented. And again, as the filter is placed inside the corporate network, the organisation has already incurred the cost and experienced the inconvenience of receiving the junk mail before filtering takes place.

Internet gateway filtering

In this environment, spam is filtered at the Internet gateway, before it reaches the mail server. Products in this area include Clearswift's MailSweeper or Mail Marshall from NetIQ and dedicated anti-spam appliances such as IronMail from Ciphitrust.

Advantages

These solutions have the advantage of keeping potential junk mail out of the mail server, and therefore reducing the amount of disk space consumed by spam. Dedicated appliances are far easier to integrate than a software-based solution, and do not require further investments in hardware. Also, unlike desktop and mail server solutions, the e-mail gateway filters can perform certain network checks such as reverse look-ups and RBL tests to check the authenticity of the original machine that sent the message (desktop and mail server filters must rely on other techniques to identify spam).

Disadvantages

Internet gateway filters do not have the flexibility to scale for large enterprises, and in the case of multiple sites it can prove expensive to deploy separate appliances at each location. This solution may also offer poor availability if only one appliance is used. In common with mail server filtering products, Internet gateway filtering limits the ability of the end user to configure their preferences and to access incorrectly blocked messages.

The key challenge to this, and indeed all of the spam filters placed inside the corporate network, is how to keep up with the ever evolving tactics that spammers employ. In the fast changing world of spamming, constant monitoring and filter tuning is necessary, and yet many enterprises do not have the manpower to dedicate to this task. Consequently, even when the latest update and patch is available, many filters remain out of date and unprepared for new types of spam.

Internet level managed services

In this scenario, a third-party service provider filters e-mail at the Internet level before it is delivered to businesses. BlackSpider Technologies is a leading vendor in this category, with its MailControl service.

Advantages

A clear advantage of a managed service is that it is easy to implement and manage. There is no hardware or software investment, and a significant reduction in associated with other spam filtering options, such as training, administration and support. There are no risks of a managed service impacting on existing infrastructure and implementation can take as little as thirty minutes for immediate protection. Internet level managed services immediately provide significant savings on Internet and network bandwidth by blocking spam before it reaches the corporate network and by freeing up IT resources associated with maintaining a spam filter in house. These types of services enable the network to become more responsive by removing the huge volume of unwanted mail that had previously clogged the network, and can deliver guaranteed reliability, often 99.999%. Of course, managed services are fully scaleable.

There is no hardware or software investment and spam is blocked before it reaches corporate networks

Core to their business, managed services providers are continually monitoring and defending against the latest spamming trends and techniques. They are able to identify and respond to potential problems quickly as, by the nature of their solution, they process massive amounts of e-mail data 24 hours a day. This, combined with many levels of spam tests, enables the managed service solution to be the most accurate of all the spam filtering options.

Providers can offer the enterprise greater protection because they often use combined anti-virus and anti-spam engines.

As for the individual user configuration, the more sophisticated managed services, such as BlackSpider's MailControl, enable end users to receive spam reports that

detail all blocked mail and enable the intended recipient to access and release false positives and configure individual user settings.

Disadvantages

A commonly cited disadvantage of a managed service solution is a lack of end-user interaction with the anti-spam service. Increasingly managed service providers are incorporating end-user self service capabilities into their offerings.

Another possible issue for some IT organisations may lie in perception: they feel they must relinquish control over e-mail filtering to their managed service provider. This is easily overcome with the new breed of vendor that allows control to be handed back to the IT business through a secure web management interface.

10 Crucial Questions to Ask in Spam Filtering

Once you have made the decision to implement a spam filter, or to review your existing solution, there are 10 crucial questions to ask in your decision making process.

Q1: **How does my anti-spam solution deal with false positives?**

Due to the nature of spam filtering, there is always the possibility of incorrectly blocked messages. The only truly effective method to deal with this is to allow the intended recipient full visibility of blocked messages and allow them to release mails that they consider have been incorrectly blocked. Can your current or intended anti-spam solution provide this?

Q2: **Do my individual users and groups have their own preferences about which e-mails should be blocked?**

End users may subscribe to legitimate newsletters etc that may be identified by the filters as spam. This problem is solved when individuals can create their own 'white lists.' Does your solution allow end-users to configure their own white, as well as black, lists?

Q3: **Is my spam filtering solution highly available?**

E-mail is a critical business tool for almost all organisations today, and downtime can have devastating effects. Is the anti-spam tool you have, or are evaluating, highly available? Is it cost effective? Have you made provision in your business continuity plans for the unavailability of your spam solution?

Q4: **Will my spam filtering solution store and queue e-mail correctly if mail servers are down?**

In this situation, desktop and mail servers will bounce e-mail. Therefore, all e-mails sent to your company during a period of downtime will be returned to sender. At best, this may give a poor impression to customers, partners and suppliers and, at worst, business critical information may not reach its intended recipients. What does your current or intended solution do in this situation?

Q5: Does my spam filtering solution require constant tuning and maintenance?

Spam is a fast moving business; as quickly as anti-spam specialists discover new ways to block unwanted mail, the spammers invent new ways to try to circumnavigate the filters. The bottom line is that spammers know the overwhelming majority of recipients don't want to receive their messages, and so devise tricks to fool the filters into accepting, then the end-users into opening mail. For businesses, this means that spam filtering solutions can require significant amounts of management and tuning to keep them as effective as they were on the day they were implemented. How much manpower cost is associated with maintaining the effectiveness of your current solution?

Q6: In the case of a software solution, does the underlying platform require management?

Spam filtering solutions are exposed to the Internet, and so it is critical that the underlying platform and applications are secured effectively. Does the solution run on top of an operating system that requires frequent patching for security holes?

Q7: Does my spam filtering solution include management reporting capabilities?

In today's environment of constrained budgets, it is increasingly important that business units can justify the value that any investment is making to the business. In the case of spam filtering solutions, this should equate to instant management reporting capabilities that demonstrate the flow of e-mail across your network, and enable you to illustrate the cost benefits of the services. Does your solution provide these capabilities?

Q8: How is my spam filtering solution priced – once I have included hardware, on-going maintenance and training? Will it scale with my business?

The total cost of anti-spam solutions include software, hardware and support costs. Any successful spam filtering solution must be as flexible as your business. If your organisation grows rapidly, will you need to buy more hardware or software to support new users? Is the solution capable of scaling to the size you want to be without significant upgrades?

One often-overlooked cost factor is staff training. Not only does training require an investment of time and money, but it also represents a delay in getting your spam filtering solution on line. How much time and money has your company invested, and need to invest, in training IT staff to manage and support your current solution?

Q9: Do I need combined anti-spam and anti-virus protection?

To remain completely secure, yes. 98% of viruses are delivered via e-mail, and virus writers are increasingly adopting the successful tactics of spammers. Therefore, combining anti-virus and anti-spam protection makes sound business sense. Could you benefit from combined anti-spam and anti-virus protection?

Q10: Do I need combined anti-spam and content filtering capabilities?

In addition to anti-spam and anti-virus, many organisations are now looking to implement content filtering policies to protect their employees. Consolidating content filtering with anti-spam and anti-virus capabilities in one single solution offers the enterprise an easy way to have additional safety and lower the total cost of ownership. Can your current or intended solution offer these additional filtering capabilities?

MailControl Spam

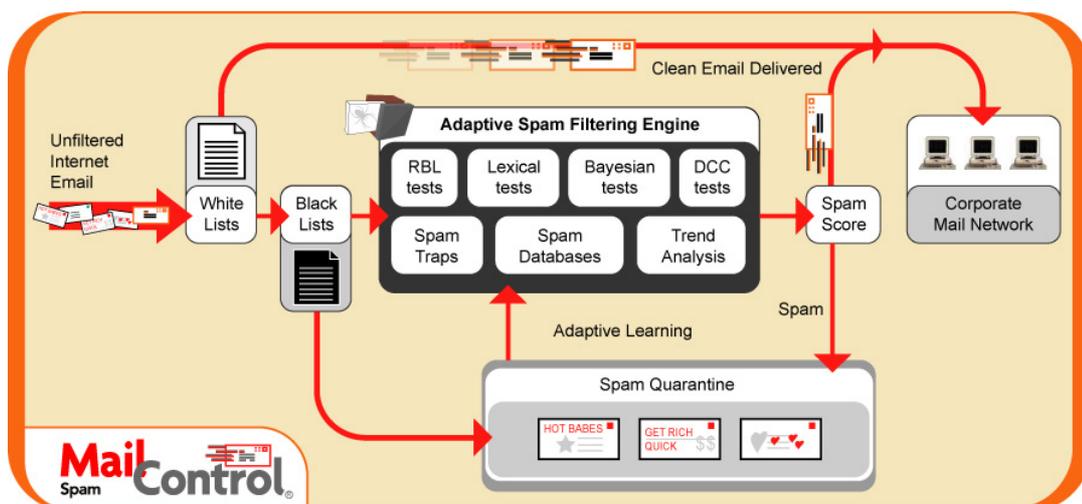
MailControl Spam from BlackSpider Technologies is a fully managed service designed to provide companies with the highest level of protection available at a cost effective price. Clients of BlackSpider retain full visibility and control over the spam filtering process.

Based on the principle that no single approach to spam detection is a silver bullet, BlackSpider combines the best spam detection techniques with a world-class infrastructure and comprehensive industry expertise.

The strength of the MailControl service lies in its ability to combine the power of lexical analysis, real-time black lists, Bayesian probability, customer control white and black lists, and collaborative hashing techniques into an adaptive filtering engine, which provides the highest levels of spam detection and virtually eliminates false positives.

This approach combined with the ability to set spam thresholds on a per-user or domain basis ensures that MailControl Spam is the most effective and practical technology on the market today.

The whole scanning process takes just a few seconds. Once each message has been analysed the message receives an overall 'spam score'. The score is then compared against the spam threshold defined by the customer; mail scoring below the threshold is delivered as normal, whilst mail scoring above is classified as spam.



A key factor in any successful anti-spam solution is creating and retaining user confidence in the service. MailControl achieves this by giving end-users visibility of the spam detected by the service without creating a volume of unwanted e-mail. Users can request an individual report from MailControl, providing details of all their e-mail processed with the 'spam score' of each message and what was blocked as spam. Users can quickly and simply add senders to white or black lists and release any messages that were blocked. This approach ensures users retain confidence in the service, and that their e-mail is being correctly classified.

In summary, MailControl Spam controls the flood of unwanted e-mail at the Internet level, keeping it away from corporate networks, increasing employee productivity and reducing costs.

MailControl Spam is available standalone or integrated with other MailControl services to provide full anti-virus protection and content-filtering capabilities.

Section i - Best practices to avoid spam

BlackSpider recommends that, in addition to anti-spam filtering technology, it is important to educate employees on the dangers of spam and avoidance techniques. There follows four general rules that you may wish to pass on to your employees

Never respond to spam

Not even to 'unsubscribe' as this is often a tactic employed to ensure that yours is a live address. Once you do respond, there is a high probability that your address will be given to more spammers: meaning you'll soon be flooded with even more junk mail.

Use a second e-mail address in newsgroups

Newsgroups are the number one e-mail address gathering ground for spammers. If you post to a group, it will only be a matter of time before you start to receive spam. So how are you supposed to participate? Use a different e-mail address than the one you use for business.

Don't give your e-mail address without knowing how it will be used

Read the terms of use and privacy statements of any site before divulging your address. If you can't find it, or if you are unsure about the authenticity of the site, do not give your e-mail details.

Never buy anything advertised in spam

Section ii - Useful links

BlackSpider Information:

- Spam: now a corporate concern:
www.blackspider.com/services/spam_whitepaper.pdf
- MailControl Spam datasheet:
http://www.blackspider.com/services/spam_datasheet.pdf

Non profit industry organisations:

- Coalition Against Unsolicited Commercial E-mail (CAUCE): www.cauce.org
- EuroCAUCE: www.euro.cauce.org/en
- Spamhaus Project: www.spamhaus.org