

Ask Ars: The best anti-spam solutions for Windows

By Rian J Stockbower

August 2003



ars technica

Copyright CondéNet, Inc. 1998-2010. The following disclaimer applies to the information, trademarks, and logos contained in this document. Neither the author nor CondéNet, Inc. make any representations with respect to the contents hereof. Materials available in this document are provided "as is" with no warranty, express or implied, and all such warranties are hereby disclaimed. CondéNet assumes no liability for any loss, damage or expense from errors or omissions in the materials available in this document, whether arising in contract, tort or otherwise.

The material provided here is designed for educational use only. The material in this document is copyrighted by CondéNet, Inc., and may not be reprinted or electronically reproduced unless prior written consent is obtained from CondéNet, Inc. Links can be made to any of these materials from a WWW page, however, please link to the original document. Copying and/or serving from your local site is only allowed with permission. As per copyright regulations, "fair use" of selected portions of the material for educational purposes is permitted by individuals and organizations provided that proper attribution accompanies such utilization. Commercial reproduction or multiple distribution by any traditional or electronic based reproduction/publication method is prohibited.

Any mention of commercial products or services within this document does not constitute an endorsement. "Ars Technica" is trademark of CondéNet, Inc. All other trademarks and logos are property of their respective owners.

Ask Ars: The best anti-spam solutions for Windows

By **Rian J Stockbower**

Thirty minutes into this write-up, I began to realize how large of a topic filtering spam email actually was, and how many different opinions there are on what works best. The purpose of this edition of *Ask Ars* is to provide you, the reader, with a broad overview of anti-spam solutions for Windows. This article will quickly introduce you to a number of anti-spam techniques, setting you up to head on out and try those that fit your needs.

Please bear in mind that some of what is written here is a reflection of the views expressed by Ars readers in a [discussion thread](#) devoted to this topic. Starting from there, we poked and prodded the options, looking to find the best and the worst effects of the various options out there. We also tested many of the more popular options in older to see if they stand up to the hype. And without further ado...

Part 1: Introduction to Spam Filtering

- What is it?
- Underneath the hood of a spam-filtering program
- Age-old advice
- Client-side filtering versus server-side filtering

Part 2: Integrated, internet-based spam filters

- What is internet filtration?
- Spam programs that use internet filtration

Part 3: Integrated, algorithmic spam filters

- What is algorithmic spam filtering?
- Spam programs that use internet filtration

Part 4: Proxy spam filters

- What is a proxy filter?
- Spam programs that act as a proxy mail server

Part 5: Server-side spam filters

Part 6: Other Solutions

- Running more than filtration system at a time
- Internet webmail filtration

Part 7: Conclusions

- In an ideal world...
- The future of spam and spam filtering
- Further reading

Part One: An introduction to spam filtering

Spam, once a term synonymous with a canned meat product, is now associated with headaches, frustration, and lost productivity. We all get junk email, and we all hate it. (If you don't hate it, I suggest you see a doctor, because there's clearly something wrong with you.) Spam-filtering, of course, is simply the act of getting rid of junk email. As with most [Ask Ars](#) topics, there is no cut and dried solution to the problem. Nonetheless, we're going to attempt to help you get your inbox under control.

With any spam-filtration software, you run the risk of classifying legitimate email as spam. In general, a spam filtering program scans the contents of an email and attempts to determine whether the email is junk. Depending on what type of spam filtering software it is determines how it actually works. **Internet-based spam filters** have their spam definitions hosted on a remote server, and usually cost money. **Statistical spam filters** require no constant connection to the internet: their definitions are stored and produced locally (usually by training that you provide it). Then, of course, there are **general filters** that use a list of keywords, or trusted email addresses to function.

For the purposes of this article, we're going to say that there are two classes of spam filtering: server-side and client-side. **Server-side filters** are fairly self-explanatory: they are programs that reside on the mail server that deal with mail as it comes in. Often they tag mail as potential spam before a user downloads it, or they delete it entirely, which is usually invisible to the user. **Client-side filters** function once mail has been downloaded. They examine the mail that the user downloads from the mail server and then decide what to do with it. It is the client side that we are most interested in, as running a program on a mail server is usually out of the question as far as the user is concerned.

Choosing the best spam filtering software boils down to two things: personal taste and your choice of mail client. How do you want to manage your email? Do you want something that plugs into the mail client that you already use, or do you want something that stands by itself? No matter your answer to this question, we have some suggestions that will fit the bill.

A quick note on how to avoid spam. This is simple, but important: don't give out your cherished e-mail address to anyone but close acquaintances, and whoever else *must* have it. A simple thing to do is to grab an account at a place such as [Hotmail](#), and use that address for any on-line activity that requires registration, getting accounts, etc. Basically, you would use this account for anything that isn't mission-critical, plus stuff that you know will generate e-mails that you will need to initially see. Even though this is obvious, we mention it because people continually point out this "option" for combating spam.

Part 2: Internet-based, spam filter plugins

As I mentioned on page one, internet-based spam filters are services provided by a third party. They store the virus definitions on their services, and the client that the user runs checks incoming mail against this database. Commonly, this database is built and refined by the users themselves. If a spam email is missed by the plugin, the user can tell the plugin that it is spam, and this will be reported back, and the spam definitions will be further refined.

The idea of fluid spam definitions is intriguing, and in a perfect world, would function nearly perfectly. However, as a userbase grows, so does the number of emails reported as spam that actually are not. Many Arsians have reported a growing number of false positives using services like this. (A false positive is an email that is designated as spam, but it's actually something you wanted to receive.) I would speculate that these false positives are generated in part because users label emails from mailing lists as spam when they decide that they no longer want to read the list. (Most plugins will block emails from a specific email address if you label two emails from that address as spam, or something similar.) This hurts the entire community: the individual might think it's fantastic, while he has essentially ruined things for many others.

Nonetheless, these plugins work right out of the box, and are the easiest and most effective solutions for those individuals that either can't be bothered to train a client, or perhaps lack the computer skills necessary to do so. Because the spam definitions are hosted elsewhere, they are relying on the work of users from the past.

Let's take a look at some plugins that use internet-based spam definitions.

Cloudmark SpamNet

One of the most popular solutions is one that we here at Ars [have reported](#) on in the past. Currently, SpamNet only supports Outlook 2000 and XP. Outlook Express is coming soon (and has been for the past year, so take that for what it's worth). SpamNet burst onto the scene about a year ago, with its *nouveau* approach to fighting spam. Rather than employ traditional filters looking for specific keywords, SpamNet used a community approach to build a database of spam. At first, SpamNet didn't catch the majority of spam, however as its community grew, so did its effectiveness. As of this writing, it is considered by many to be the best single solution to the spam issue. Recently, however, some readers have reported that SpamNet is now tagging legitimate emails as spam: mailing lists, newsletters and the like. You can combat this by marking messages as not being spam, but this means you have to be careful to watch filtered spam mail.

SpamNet operates by scanning incoming emails, and comparing them with its online database, which, as I said, was built by the userbase. Because SpamNet requires an internet connection through the entire checking process, some users are precluded from using it. If you have a firewall and port 2703 is blocked, SpamNet will not work. I myself could not use it this past year at school for this very reason. The program does, however, have SOCKS 4 and 5 support; that might be an option for you.

Recently, SpamNet decided to go to a subscriber-only model, much to the chagrin of users who were initially told that SpamNet would always be free. Right now, the current rate is \$1.99 a month for those who were beta-testers; for those who were not, it is \$2.99. SpamNet is incredibly convenient, but it's not the best solution for everybody.

Pros:

- Easy to install
- Works out of the box
- Integrates seamlessly with Outlook
- Comprehensive FAQ and responsive tech support
- Outlook Express support forthcoming

Cons:

- Only works with Outlook 2000/XP
- Requires use of commonly blocked port (could be a problem for firewalls you don't control, Internet Connection Firewall settings on corporate machines, etc.)
- Not free

Mailfrontier Matador

Matador is similar to SpamNet in many ways. It uses a community to fight spam, much the same as SpamNet. Matador also has its own internal spam definitions. There is a one-time \$30 payment to use the software. Matador has a few more features than SpamNet. It allows the user to set up *challenges* for junk senders, questionable senders, or all senders; similar to Mailblocks and other similar services. Users have reported few false positives, unlike recent SpamNet experiences. What is a challenge? In short, the challenge model tries to weed out spam by authenticating valid e-mail senders by checking to see if they're a) really at the address they they claim to be sending from, and b) not just a brain-dead spam bot. You never even see the e-mail they've sent until the challenge has been answered. The process requires the sender to verify once, and then all of their emails will be received from then on. You can view a sample challenge [here](#).

Pros:

- Works out of the box
- Challenges for senders
- Internal spam definitions as well as server-side definitions

Cons:

- Not free
- Some users report agitation from mail senders who don't want to deal with the *challenges*.

Part 3: Integrated, algorithmic spam filters

Algorithmic spam filtration is filtration that does not check email against a remote server. Rather, programs have to be "trained" to recognize what is spam and what is not. The volume of email that one receives is the primary factor in determining how quickly the plugin "learns." Some plugins learn faster than others, but it is primarily the amount of email that the plugin has to build a database of good and bad with that determines this.

The database that the plugin checks email against is the most important part. An advantage to these filters is that the user defines precisely what is spam and what is not. What one labels as spam is only spam to you, not someone else: there is no danger in marking that email from that listserve as spam, because it's not going to affect anyone else. The obvious disadvantage is that one has to build a database to begin with, and that means that there are going to be false positives, and lots of false negatives at first. With time and effort, the numbers for both will gradually decline.

The most common sort of spam filtering is Bayesian sorting. Bayesian sorting is useful for more than just spam filtering. It can be used to sort mail into specific categories like work, personal, and spam. Most Bayesian software, however, only uses it to sort what is spam and what is not. Roughly speaking, Bayesian sorting uses statistical analysis to see which words appear as spam, and which words do not, and it uses this information to build a composite score for a particular email. In theory, the larger the database gets, the more accurate it should be. Many people consider algorithmic sorting that is *not* based on the Bayesian method to be inferior... I'm inclined to agree.

For those interested in more details regarding Bayesian statistical analysis, check out bayesian.org; for details specific to just Bayesian spam filtering, check out Paul Graham's "[A Plan for Spam](#)". Graham is one of the pioneers of using Bayesian statistical analysis to filter out spam. It's a fairly technical, but interesting read.

Now let's take a look at some plugins that function via user-defined spam definitions.

Mozilla Mail and Thunderbird

Mozilla Mail is bundled with the Mozilla web browser and Thunderbird is the standalone Mozilla mail client. There will not be an all-in-one Mozilla internet platform now that Mozilla 1.4 has been released. From now on, each component will be its own separate entity, so soon Thunderbird will be the only option. (Unless of course, you enjoy using deprecated software.)

Pros:

- Easy to install.
- Filter is integrated directly into the mail client itself — no external plugins to futz with.
- Very active, development team.
- An active [wishlist](#).
- Free.
- Easy to unblock mail.

Cons:

- Does not work out of the box.
- The client takes a bit of training before it becomes accurate.
- Since its not a plugin, you'd have to change e-mail clients to use it if you don't already.

Spammunition

Spammunition is a plugin for Microsoft Outlook versions 2000 and higher. It is a freeware plugin that uses Bayesian sorting to identify spam.

Pros:

- Easy to install.
- Free.

Cons:

- Beta software.

SpamBayes in Outlook plugin form

Caesar has been using this plug-in for a few months so he's hijacking this part of the write-up:

For the most part, I've been pleased but not blown-away with SpamBayes. Perhaps it because I get so much freakin' spam, but the client just misses certain kinds of spam repeatedly. In July, I received about 20,000 pieces of Spam. SpamAssasin (server-side) caught 95% of those. Of those same messages, SpamBayes only caught about 70% (filtered separately for testing, and after month of training), and it took it two weeks to figure out that Viagra e-mails were spam (and it falls victim to *v l a g r a*, *v i a g r a* and all those dumb tricks). On the other hand, it didn't nail a *single* false positive, which is great, and you can set it up to move *potential* spam into a folder for you to look over. When looking over this potential spam, you mark things as being spam or not, and SpamBayes learns. The client is also fast, and easily configurable. Ideally, I would run this in conjunction with a server-side option, if possible. But when all is said and done, I'm still using it, and I've tried about everything.

Pros:

- Free.
- Probably the best plug-in for Outlook out there.
- Allows quarantining of questionable mail based on probability that it is spam.

Cons:

- Clunky install.
- Nonintuitive interface.
- Doesn't appear to be well suited a very diverse amount of spam.
- "Forgets" its training if Outlook is not manually closed before Windows shutdown (they're working on this).

Outclass

This isn't really a standalone filter; rather, it is a plugin system to control a Windows **POPFile** installation, which I will explain further on [page four](#). I'll briefly mention it here, and then you can consider whether or not POPFile is of interest to you in a bit.

Pros:

- Uses POPFile scripts.
- Allows "new mail" notification only when new mail is not spam.
- Eliminates POPFile's clunky interface.
- Free.

Cons:

- Requires a POPFile installation.
- Resource intensive due to POPFile running.

Part 4: Proxy spam filtering

A standalone "proxy" filter is a fancy word for an intermediary piece of software that works between you and your mail server. Typically, one sets up this proxy, and it interfaces with the mail server, and then your email client interfaces with the proxy. The proxy itself does all of the filtering, and then one typically sets up rules based on how the proxy has labeled the piece of mail.

There are a few general pros and cons to proxy setups. The good thing is that just about any email client can use them: no plugins required. The only thing your email client has to do is be able to read message headers: I've never heard of a modern email client that cannot. The bad thing about proxy setups is that they're generally difficult to set up for accounts that don't use POP3, which is a problem for the proliferating IMAP universe.

Usually proxy spam filters are set up on a local machine, though others, such as POPFile can be set up on a remote machine, making them somewhat more manageable to deal with if you're often using e-mail from different locations. Some people dislike having two programs running when one would suffice. The flexibility allowed by proxy setups is one of the biggest draws, however, so it boils down to a matter of preference. One other draw is that users of Outlook Express can also play, whereas with many of the other plugin methods, they cannot.

Some spam filters that employ the proxy filtration method:

POPFile

POPFile is probably the most popular and arguably the most powerful proxy system available right now. It has a web interface which is used for training and mail classification training. Some people like the web interface, and some people hate it. If you are using Outlook as your mail client, and dislike the web interface, try out Outclass, which I outlined in [part three](#).

Pros:

- Allows mail to be classified as anything: spam, school, work, etc.; however you want it to be sorted
- Multi-platform
- Runs either in the background or as a console window
- Trains very quickly
- Open source
- Can be configured for network use
- Free

Cons:

- Requires a third party program to be running
- Can be resource intensive
- Clunky install
- Requires rules to be set up within email client

SpamPal

SpamPal is a bit different than POPFile in a few respects. It has some features that POPFile does not have, and it lacks some of POPFile's customizability. SpamPal uses blacklists which tags mail sent from certain parts of the internet: those parts which statistically send lots of spam. It also has a whitelist feature which allows mail sent from certain addresses,

because not all mail sent from blacklisted addresses is actually spam. SpamPal allows you to choose which blacklists to use as well. There is a fairly comprehensive unofficial setup guide with screenshots that you can check out [here](#) if you are so inclined.

Pros:

- Utilizes DNSBL lists (blacklists)
- Whitelists
- IMAP4 and SMTP support forthcoming
- Free

Cons:

- Requires a third party program to be running
- Requires rules to be set up within email client

Firetrust Mailwasher

Caesar has had much experience with this as well (here's taking over my keyboard again!). Mailwasher is not a traditional spam filtration product, and it's not really a proxy, either. It is a standalone program that must be run in conjunction with your mail program of choice. It sort of functions like a proxy, which is why we have it listed here. The idea is that Mailwasher checks your mail before you do, and then nukes anything that's not kosher before you actually download it with your normal mail client. You can set it up to download just the headers, or the first *so-many* lines; the more you download per message, the better it can filter, but at the cost of speed. Mailwasher attacks spam in two ways: first, users can scan the headers of mail and then mark it as spam. Doing this will teach Mailwasher what you think is and isn't spam. Second, Mailwasher can compare the origin of e-mails to known spam sources (using SpamCop, for example).

You can chose to blacklist or whitelist senders, and even bounce mail. We don't normally recommend using the bounce feature, as the majority of spam senders' addresses are spoofed. Bouncing the mail only puts the burden unfairly on innocent parties, as far as we can tell. Mailwasher, in its commercial form, is much improved. The previous (free) version was bug laden, and would often crash when I used it. The new version still crashes, but very rarely.

SpamNet-like community spam identification is planned for [release this month](#).

Pros:

- Doesn't require email to be fully downloaded.
- Gives you access to SpamCop and ORDB spam databases.
- Works with POP3, MSN/Hotmail.

Cons:

- Mail must be filtered by at times.
- \$30 one-time fee.
- The interface feels very clunky.

Part 5: Server-side spam filters

Server-side spam filters is obviously that software which resides on the actual mail server. What it does with incoming spam is dependent upon whether the server simply trashes incoming spam, or if it simply tags that mail as potential spam, and then allows the user to deal with it as he or she sees fit.

Server-side spam filtration software:

SpamAssassin

Deersoft's SpamAssassin got rave reviews from a few Ars readers. There's only one caveat to using the product: Deersoft was bought out by McAfee, who discontinued the product. Nonetheless, several people advised using it if you could find someone willing to part with their license to snap it up. The open-source version that is available is linked above.

Spamcop

SpamCop is *primarily* an off-site spam solutions provider, and offers a variety of services to those that desire them; they don't offer spam-filtering software *per se*. They do, however, offer a DNS-based blacklist, which administrators can use with their mail servers. One of the downsides to SpamCop is that they err on the aggressive side; false positives are a definite risk. This is openly stated, and those that choose to use SpamCop should be aware of this. SpamCop's database has been made available to certain programs such as Mailwasher.

Part 6: Other Solutions

Free email services

There are some other solutions for relatively spam-free email. The most simple and obvious choice is using a free webmail service such as Yahoo! or MSN/Hotmail. Both of these services offer spam filtering as part of their free service. The aggressiveness of the spam filters, and what is done with mail flagged as spam is usually set by the user under their preferences. Oftentimes, your ISP will have some sort of anti-spam measure on their mail servers. Sometimes it is enabled by default, but most of the time, it isn't. Call your ISP, or poke around in your web interface to see about enabling it.

There has been some debate in the past over whether or not these free services have been selling their email lists to spammers. I think it's safe to say that this debate is over: selling email lists to spammers makes no financial sense, as it only costs these companies money in the long run in the form of added strain on their servers and increased network traffic. The fact that Microsoft and AOL are beginning to sue spammers in an attempt to recoup money spent filtering out their junk email is further evidence that the practice of selling email addresses just does not happen. This isn't to say that such companies aren't selling access to your account in the form of special "deals" for their users and whatnot, but it isn't the case that Microsoft is selling its entire userlist to just anyone. They can sell such stuff in ways that don't require them actually giving out your e-mail address.

So what does happen? Well, spammers resort to automated dictionary attacks. They go through a dictionary and throw mail at email combinations. Or they simply generate nonsensical email addresses and throw mail at these. And they also spider the web for posted addresses.

Multiple spam-catching solutions

Some people opt to use more than one anti-spam solution. A crude example of this would be enabling your ISP's spam protection, and then following it up with your own anti-spam measure, like SpamNet or POPFile. Multiple nets for catching email increases the chances of having a mostly spam-free inbox, but bear in mind that the odds of getting a false positive also increase.

Part 7: Conclusions

The ideal world

There are those that would argue that spam filters are unnecessary if one simply takes a few precautions. As a blanket statement, this is patently false. While one can certainly see the conventional wisdom in not giving out your email address or posting it on the web, for many people it's practically impossible to avoid doing so: people doing business online are perfect examples. And you shouldn't need to hide your e-mail address anyway. Part of the convenience of e-mail is that you can reach colleagues, engage in discussions online, and the like.

Another favorite piece of conventional wisdom is just to sing the mantra of the "delete key." You'll hear someone say, "the delete key works just fine." Yes my delete key works just fine, but I don't have time to examine all 250-300 emails that I get on a daily basis to see whether they're junk or not. I used to do this in the past, and I often ended up deleting real mail. It's just not time-effective to hand-weed, and my guess is that the acceptability of the "delete key" method is inverse to the amount of e-mail and spam you get. So we are left with two alternatives: unplug entirely or find some way to manage the garbage. People that live in the real world are stuck with the latter.

Posting your email address publicly on a site with even just a tiny bit of exposure is enough -- remember, hyperlinks are the glue that hold the web together: spam spiders can follow them as easily as a search engine. Be careful when and how you post your email address, and obfuscate it if possible. Obviously there are many times when obfuscation is impractical or unprofessional, and it's these times when a spam filter can be your best friend.

In an ideal world, a spam filter would be created that would incorporate all of the best features that each method of spam blocking offers: community-based software, algorithmic filtering, blocklists, and not having to download the email if it were spam. Ideally, whitelists would be unnecessary. Naturally, one would be able to turn one or all of these measures off if they so desired.

Of course, in a truly ideal world, there would be no such thing as spam at all, and there would be no need of spam filters, period. Utopia literally means "no place," and this is certainly the case when it comes to spam: everyone is susceptible. But, there is hope at the end of the proverbial tunnel: many states are considering legislation which would regulate how spam is labeled, making it easier for filter-makers to separate the wheat from the chaff. The downside to anti-spam legislation is clear: unless broad regulations are enacted and enforced, the spam problem will continue unabated. This means federal action. Thankfully, federal action is forthcoming, as we reported [recently](#). What the laws will say, and what penalties will be for breaking them is still a matter of debate. Better to take the time to draft a fair and relevant bill than to rush and end up with worthless legislation.

Further reading

John Graham's [notes](#) from an anti-spam conference in January make for some interesting reading. The PDF details the tricks that spammers are using to get around many spam filters, and how a spam filter needs to be able to cope with these tricks. Graham is the author of POPFile, and his notes outline how POPFile specifically deals with these tricks, but his notes are relevant to all anti-spam developers.

To reiterate what I said on [page three](#), there is some good info for those into statistics and its practical applications with respect to spam-filtering to be found on [Paul Graham's site](#), specifically his treatise entitled "[A Plan for Spam](#)".