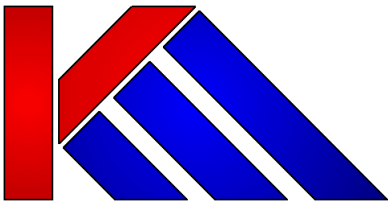


KM-Parse v2
User Manual

Overview and Deployment of a Server-Based Whitelist / Blacklist / Greylist System for Windows

Revision 2.1.11
Prepared by Kevin Millican
Date 27th October 2006

Table of Contents

Disclaimer

NB: KM-parse has only been tested with SmarterMail, SpamAssassin, and ClamAV. Whilst every care has been taken developing this utility, the author gives no warranty for its use and prospective users are advised to backup any important data before evaluating KM-parse. **You use this program at your own risk and under the implicit condition that the author is not liable for any loss of data or damage to your systems howsoever caused.**

Licence

Introduction

Antispam utilities and services are constantly evolving but the main strategies used normally fall into one or more of the following categories:-

- Bayesian filtering : the content of an email is checked for the frequency of words or phrases typically used by spammers
- Dictionary scoring : the quantity of real text is evaluated to check for obscuring techniques used by spammers to avoid detection by Bayesian filtering
- Whitelisting : a list of trusted senders is maintained
- Blacklisting : a list of unfriendly senders is maintained for blocking
- Challenge/Response (eg. Bluebottle.com) : an extension of whitelisting - when a new mail is received from a previously unknown sender, a challenge email is sent requiring some user response such as clicking on a link to verify that the sender is a real person instead of a mail robot.
- Spam Origin : the IP addresses used to send or relay email are checked against public databases of known offenders.
- Sender Policy Framework (SPF) / SenderID / Domainkeys : is really more of a defence against other people spoofing our own domains. It allows the domain owner to specify who can send email from the domain. It makes it harder for people to spoof emails from eg. hotmail.com or yahoo.co.uk, but doesn't do anything about domains that don't provide records or have lax sender policies.

Most antispam solutions use some combination of Bayesian filtering, whitelisting, blacklisting, and spam origin databases. This is usually effective in removing at least 80% of spam from our inboxes. The question is, "how can we deal with that other 20% without running the risk of losing legitimate emails?"

One of the problems with filtering out spam is that blacklisting rarely works because spammers tend to use a different 'From:' and 'Return-path:' address on each email. A typical blacklist contains thousands of email addresses that have only been used once. This breaks blacklisting as a useful tool. Whitelisting is hard to maintain unless you use a challenge/response system. However, unless considerable care is taken in the way these systems are implemented, they can create more spam by sending unsolicited emails to many people who have been unlucky enough to have had their address spoofed.

An underused technology is the concept of greylisting. Paradoxically, the counter-defence used by spammers against blacklisting, is a terrible weakness if a mailserver implements greylisting.

A greylist is a list of ALL email addresses sending email to a domain, whether legitimate or not. It is important that a greylist is updated automatically, but not too frequently. If a greylist is updated immediately, then multiple emails from the same address to different addresses on the domain will look no different to 'normal' email. Some greylisting systems (http://greylisting.org/) will actually refuse new sender emails on the first attempt but this is not necessary for the system to be useful (and this approach may be circumvented if spammers start obeying RFC protocols properly).

KM-Parse is designed to record and provide historical information about a sender, assist with whitelisting and blacklisting, and incorporate well-known antivirus and antispam tools such as ClamAV and SpamAssassin.

Features

- Easy integration with mail servers that allow a command line to scan incoming email before it is delivered.
- Execution of user's preferred antivirus scanner and SpamAssassin antispam utility.
- 'On-the-fly' autorun of antivirus and antispam server daemons, if required.
- Autowhitelisting
- Supports multiple domains
- SQLite database-driven table of senders with intelligent use of non-matching From: and Return-path: addresses. Can be edited using SQLite Browser (http://sqlitebrowser.sourceforge.net) or KMPedit (basic editor provided with KM-Parse)
- Status of incoming sender written into a custom header line. Information includes status (whitelisted, blacklisted, greylisted), number of emails received from that source, and a dispassionate assessment of whether this is sender is new ie. NEWADD, OCCASIONAL, FAMILIAR, or FREQUENT.
- Retention of header information when a virus is detected. Most antivirus scanners will just delete the entire email but it can be useful to see where the email came from and the nature of the infection.
- Built-in defaults for ClamAV and SpamAssassin
- Can relieve server loading by only running SpamAssassin on incoming email from external senders.

Optional Software

It is possible to run KM-Parse solely for its whitelisting/greylisting/blacklisting features, but use with ClamAV is recommended. SpamAssassin is also useful, provided your system can stand the processing overhead (see KM-Parse log file for accurate timings)

ClamAV for Windows - http://www.sosdg.org/clamav-win32/

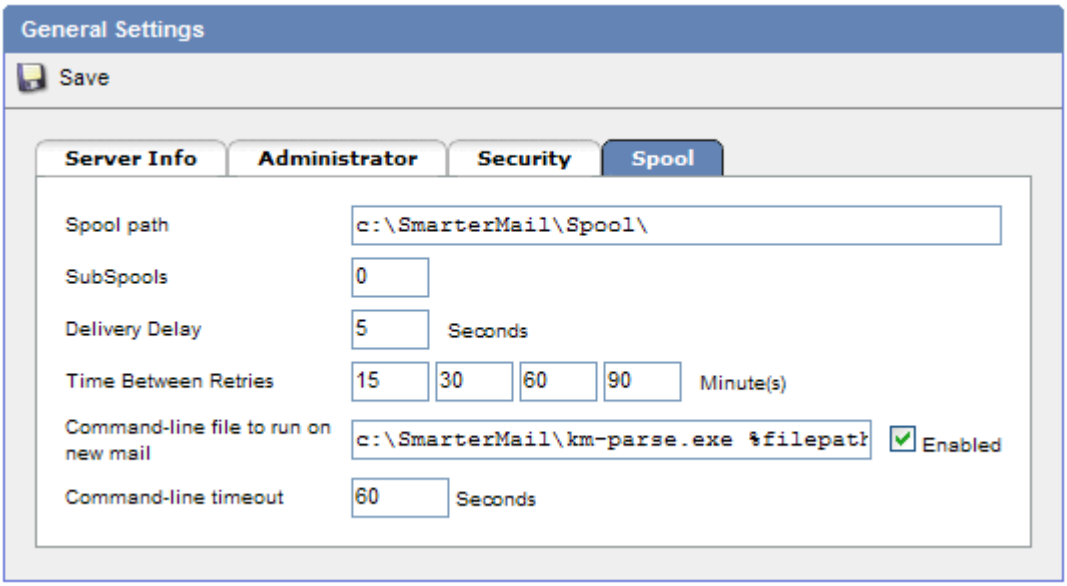
(alternative http://w32.clamav.net/ has a different setup - see advanced options)

SpamAssassin for Win32 - http://physics.ucsd.edu/~epivovar/anti-spam.htm
http://physics.ucsd.edu/~epivovar/SpamAssassin-3.1.3-win32.zip

SQLite Database Browser - http://sqlitebrowser.sourceforge.net/

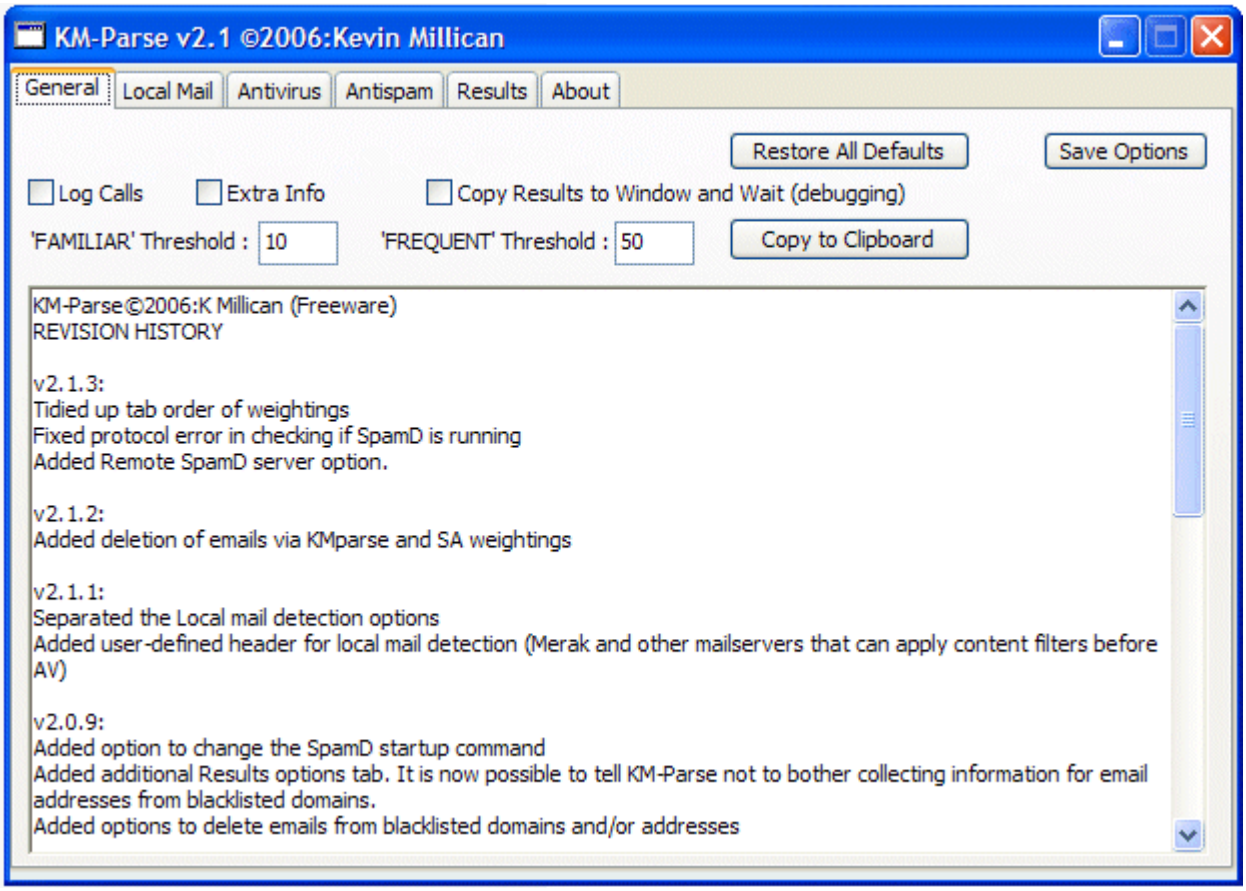
Installation

Copy km-parse.exe to any folder and point your email server's antivirus command line scanner to it. Eg. In SmarterMail, you might choose to place it in the c:\SmarterMail folder, and set the spool settings as follows:



Configuration

To configure KM-Parse, run the program without passing it any parameters (ie. Double-click on it). You are greeted with the following screen:-

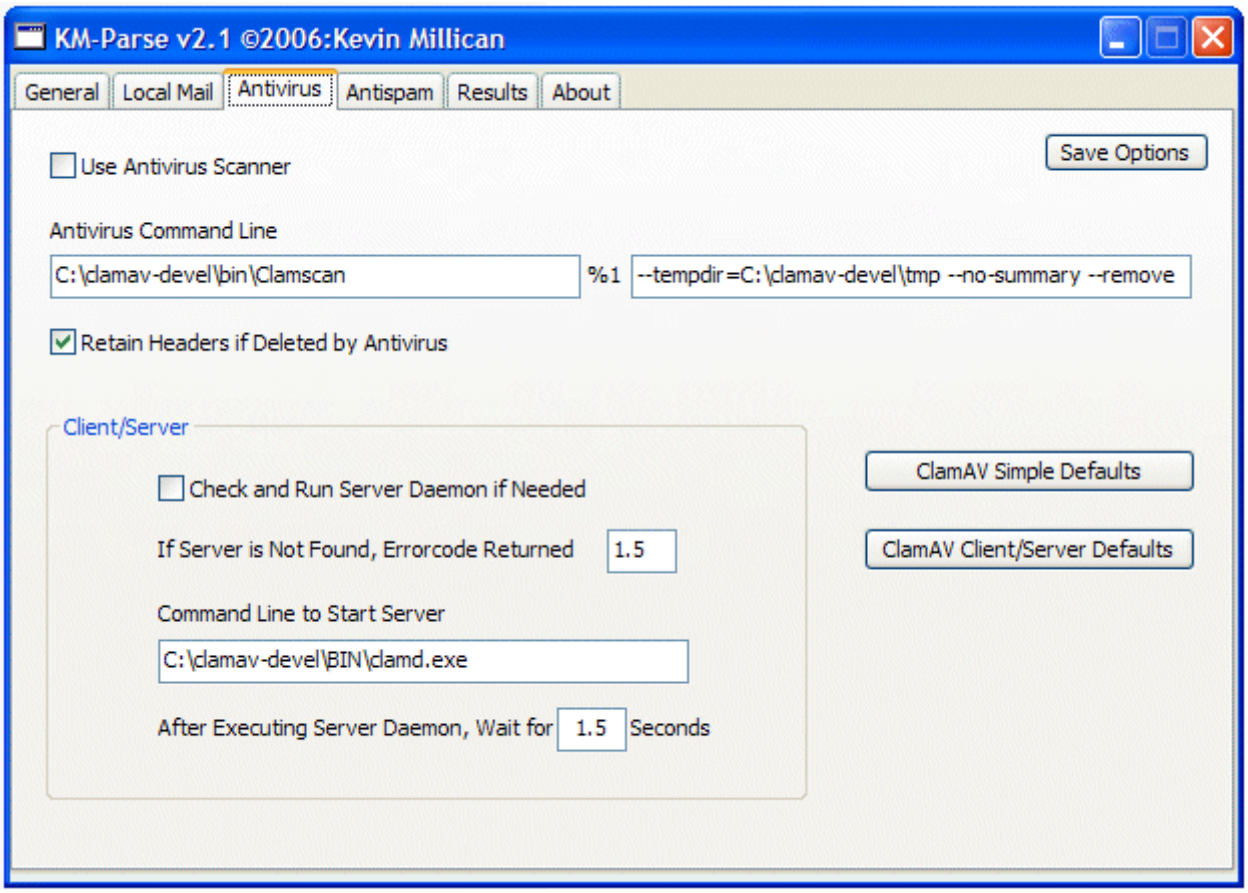


You may want to tick the 'Log Calls' option – this tells KM-Parse to keep a logfile of its activities.

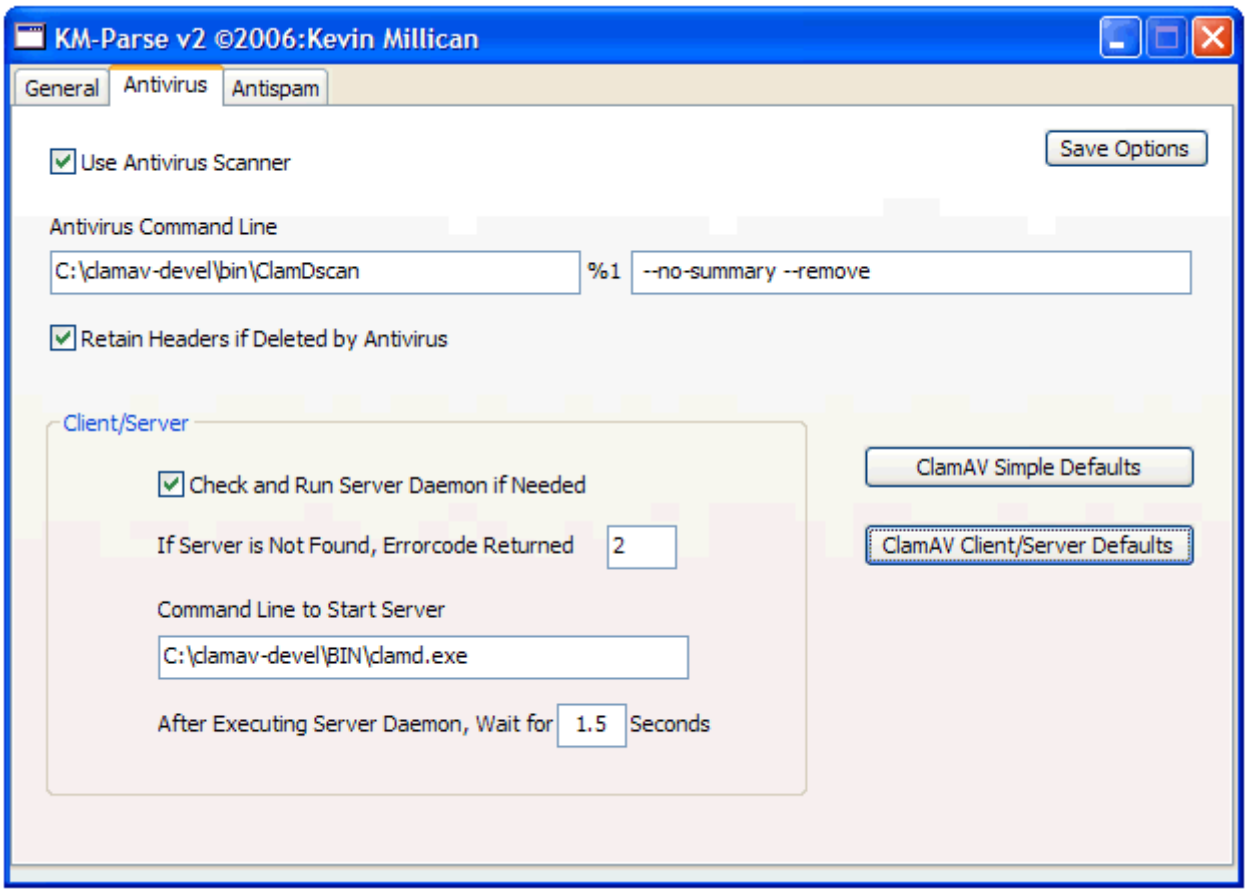
To begin with, **do not** tick the 'Copy Results to Window and Wait' option. This causes the window to appear after parsing an email for 5 seconds – during which time the 'Copy to Clipboard' button is available to see a more detailed account of how the email was processed.

The 'Extra Info' option causes KM-Parse to copy the contents of any associated *.hdr file to the *.eml file as a series of header lines. It is really only designed to work with SmarterMail, though it may work with other mailservers.

Click on the 'Antivirus' tab to bring up the following screen :-



The Antivirus tab is preset with values for a simple ClamAV setup. Most users will probably want to use the client/server version because it runs faster. A button is provided to preload these defaults :-

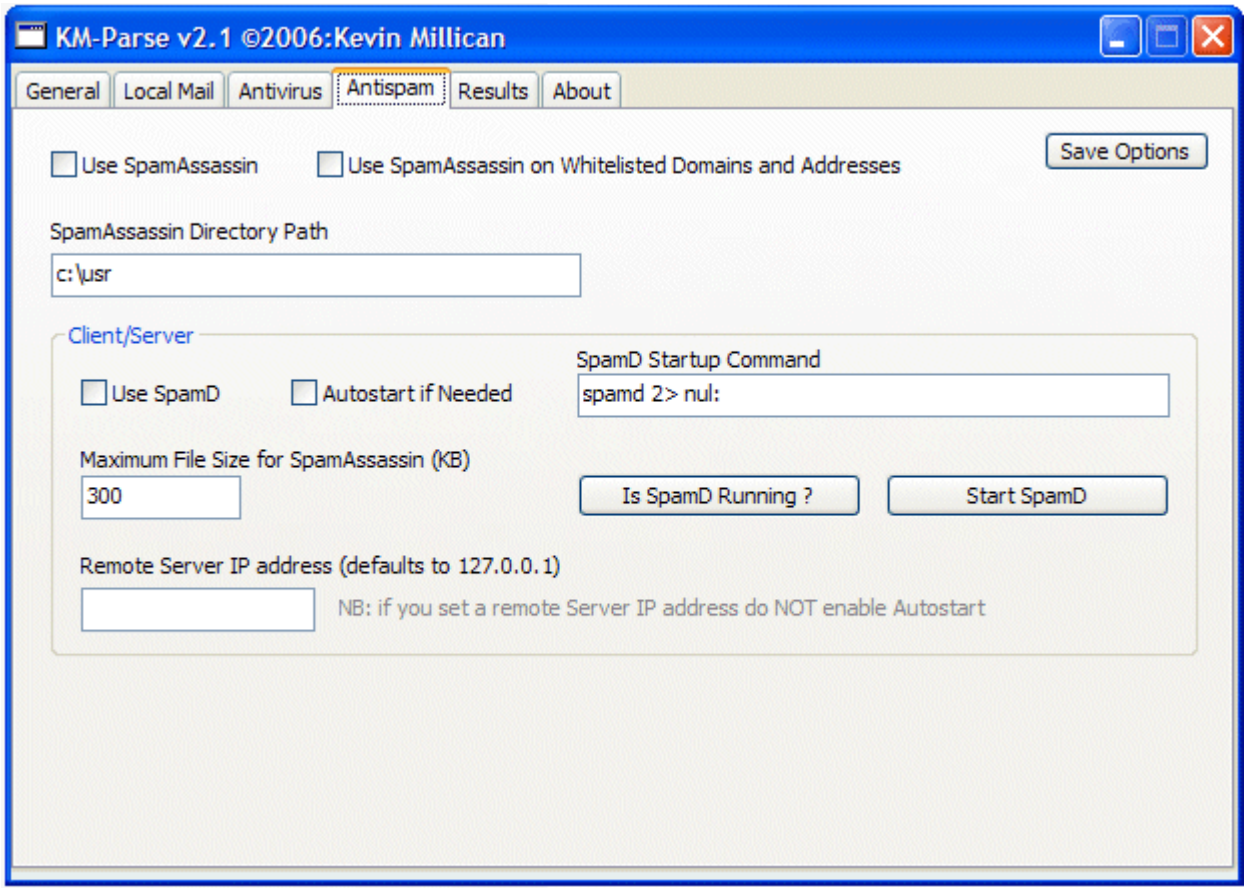


Note that you must click the 'Save Options' button to make the changes permanent. You can use alternative

antivirus scanners but they must be configured to delete emails that contain viruses.

If you don't wish to use an antivirus scanner, just untick the 'Use Virus Scanner' option and click 'Save Options'

The AntiSpam tab allows you to configure SpamAssassin.



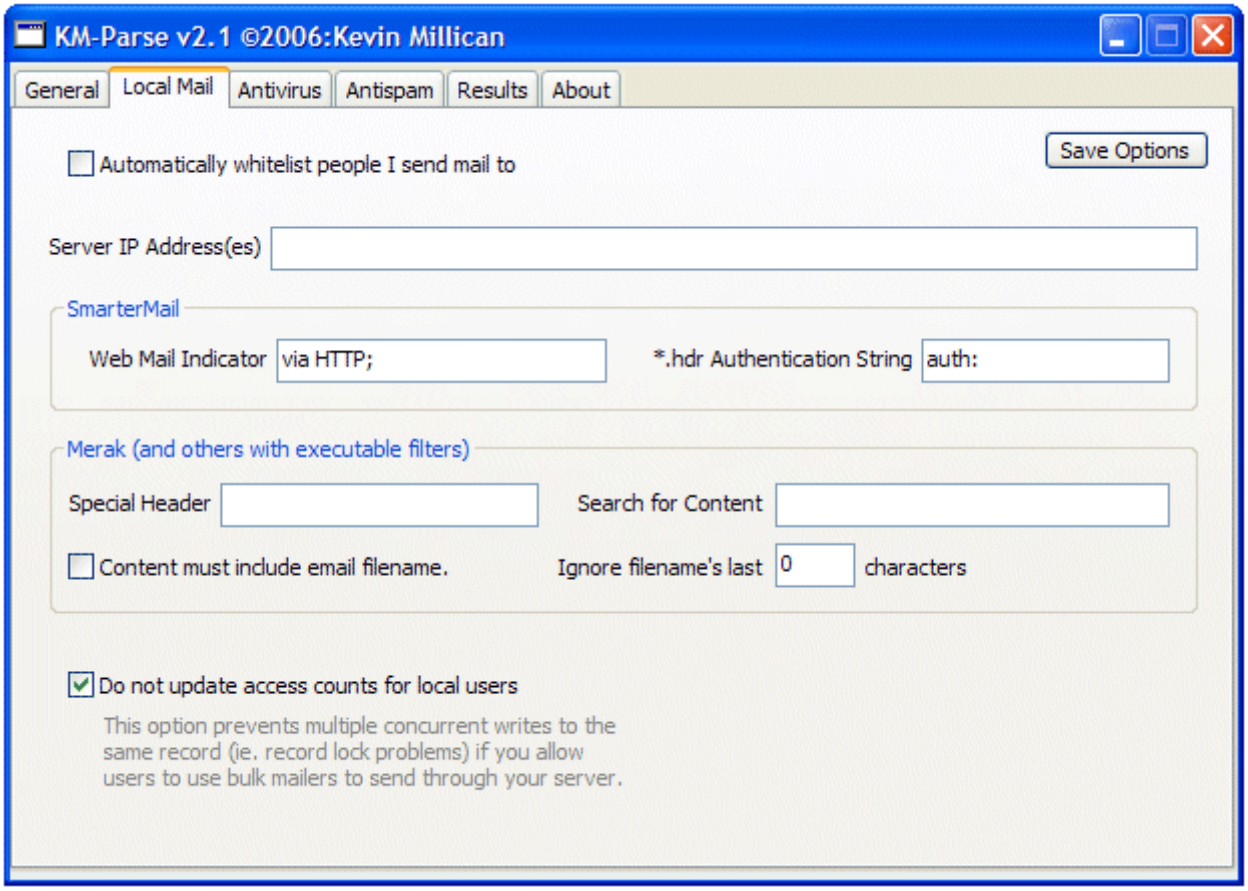
If you want to run SpamAssassin, it is recommend that you install it into a folder c:\usr
Other configurations may work perfectly well, but this is recommended in the SpamAssassin documentation and is the only path tested.

If you are going to use SpamAssassin, I **strongly** recommend that you tick the 'Use SpamD' and 'Autostart if Needed' boxes so that the SpamAssassin Server Daemon is used. For even higher stability, running SpamD as a service is advisable – in which case, leave this option off. If you do not use SpamD, Spamassassin will rewrite the email rather than incorporating its score into the headers for some of the other KM-Parse feature to use.

The 'Is SpamD Running' and 'Start SpamD' buttons are provided purely for test purposes to ensure that the path is setup correctly. The check is also useful if you are using a SpamD remote server. NB: if you use a remote server, you will need to ensure SpamD is running on it by some other means – KM-Parse can only start local instances of SpamD.

The 'Use SpamAssassin on Whitelisted Domains and Addresses' option has no effect on mail sent from the server's hosted domains. If KM-Parse can tell that such mail is authenticated or comes from approved IP addresses, then it will not use SpamAssassin. If the source cannot be verified, then SpamAssassin will be used if the 'Use SpamAssassin' option is ticked.

KM-Parse uses several methods to determine if email is really local. Click on the 'Local Mail' tab to bring up the following screen :-



You'll probably find it useful to tick the 'Automatically whitelist people I send mail to' box. However, to begin with the program doesn't know the difference between mail originating on your server and from outside.

Enter the external IP address of your server(s) in the 'Server IP Address(ses)' field. There is no need to input internal LAN IPs. Some users may have more than one external IP – in this case separate them with a space character.

The 'Web Mail Indicator' is used to input a string that is found in the received: header when mail is sent via the server webmail facility instead of SMTP from a mail client. The '*.hdr Authentication String' is specific to SmarterMail and probably has no use on other systems. These fields are used to check whether mail originates on your server or is authenticated.

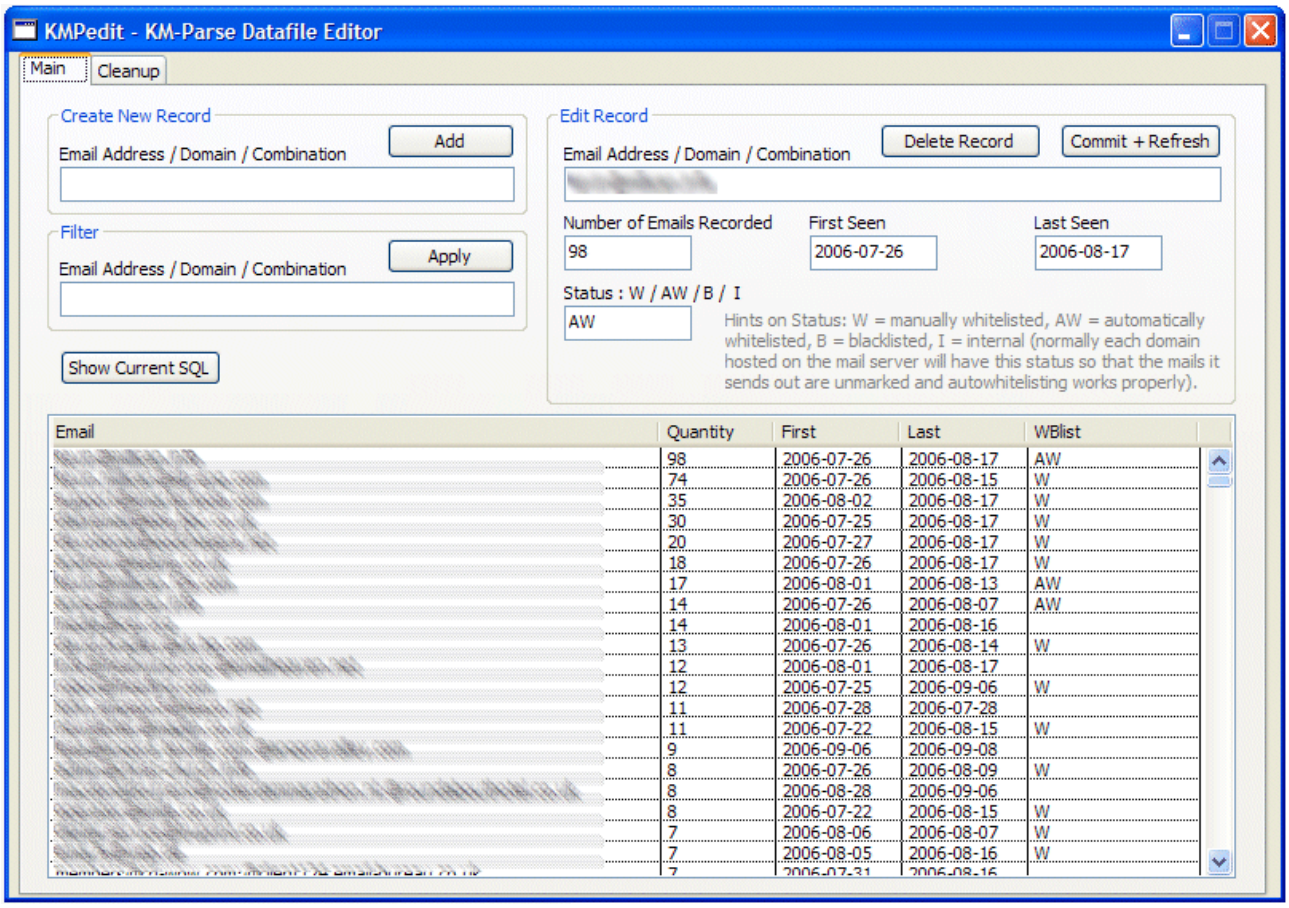
Some mail servers, such as Merak, allow filters to be run before the command line that executes KM-Parse. These filters may be able to create their own headers to verify that an email really is sent by a local user, so the header name (and content if needed) can be checked by KM-Parse. If you leave the content field blank, then you can also use the custom filter to insert the filename as the header content – KM-Parse will then check this (less a preset number of trailing characters) to see if it is present in the checked file's pathname.

The 'Do not update access counts for local users' is advisable is you allow any of your users to send bulk mail. If this is unticked, KM-Parse has to carry out multiple near-concurrent writes to the same database record and this could potentially lead to record-locking errors. Ticking this will also minimise the parsing time.

Editing Datafiles

The simple datafile editor – KMPedit.exe – should be copied to the same directory as KM-Parse.exe

As soon as KM-Parse has processed a couple of emails, run KMPedit and add all the domains hosted by your mailserver to the table. Then change the status for each one to the single letter 'I' (for 'Internal'). This will ensure that outgoing email doesn't get marked with the KM-Parse header line, and that autowhitelisting works correctly, if enabled.



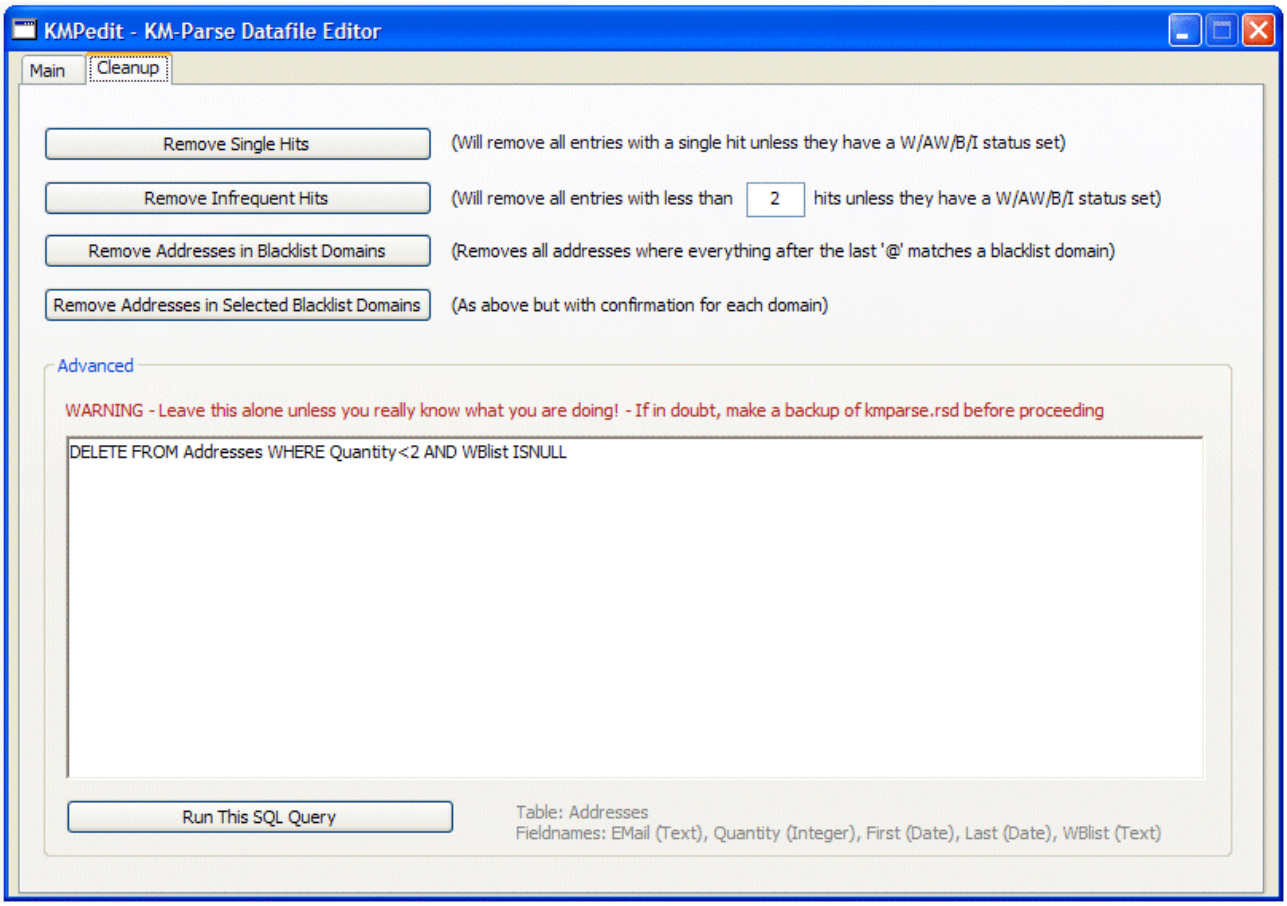
You can also manually whitelist or blacklist domains or addresses.

The address entries fall into 4 categories, eg. :

- somedomain.com — Domains
- someone@somedomain.com — Email Addresses
- @somedomain.com — Return-path: domains (no From: address)
- someone@somedomain.com:@other.com — Combination From: + Return-path:

This utility contains search options and maintenance filters. Changes may be also be made manually or using 3rd-party SQLite database editors.

There is also a user-defined SQL query but you should only use this if you are familiar with SQL and after making a backup of the kmparse.rsd file.



Using the Results

It is up to the user to decide what action they wish to take in response to the KM-Parse results.

In the author's opinion, any email flagged :-

```
X-KMparse: NEWADD ...
```

should be regarded as extremely dubious if it also triggers a SpamAssassin or mailserver spam rating – it is nearly always safe to use a content filter to delete these, or at least move them to a junk account for manual assessment.

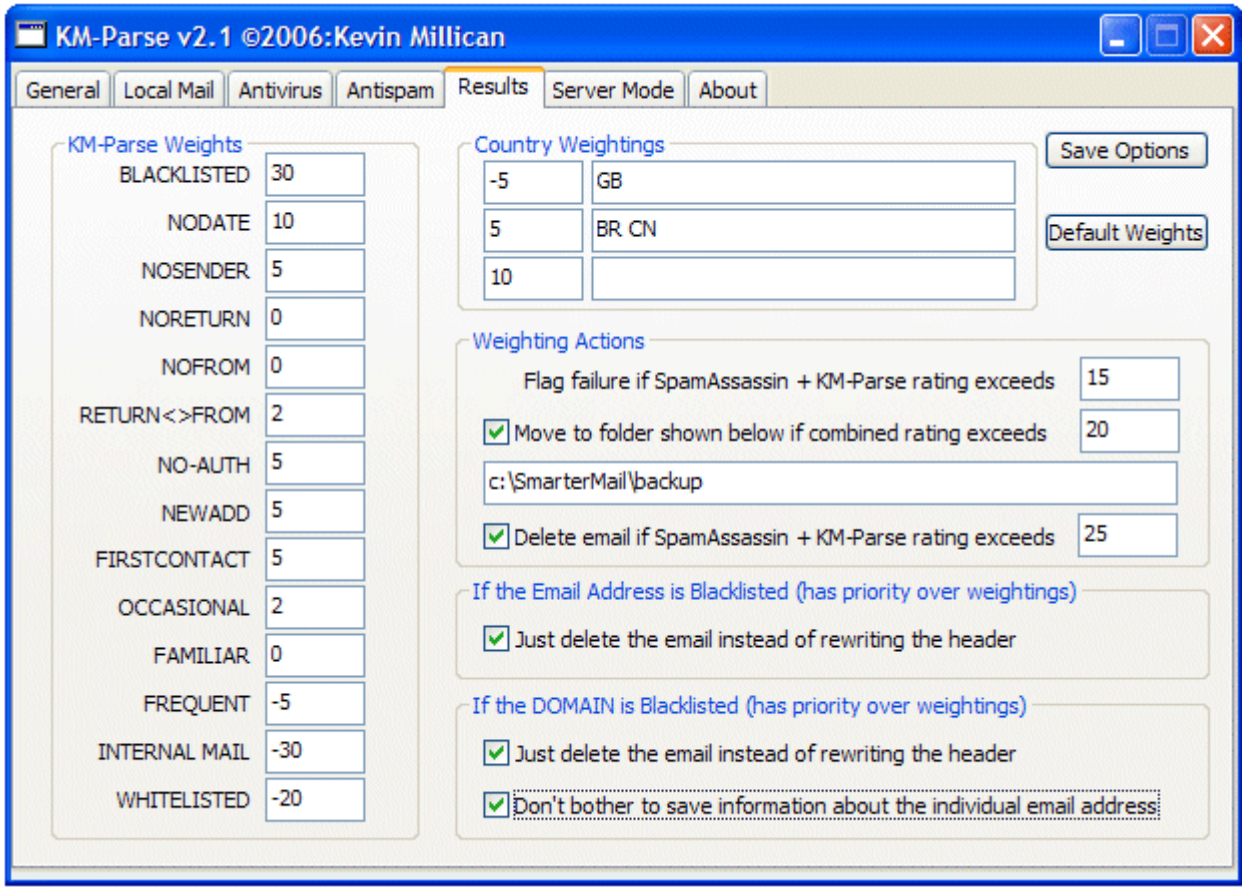
There are other flags that can be useful in applying filters, eg. :

NODATE	The Date: header was blank or missing
RETURN<>FROM	The From: and Return-path: have different domains
NOFROM	Blank or unresolvable From:
NORETURN	Blank or unresolvable Return-path:
NOSENDER	Blank or unresolvable From: and Return-path:
NO-AUTH	The mail appears to come from of the server's domains but this is unverified. (It is spoofed or sent by a user without authentication through another server)

It is also possible to weight the country of origin by entering the two-character country code. Multiple entries should be separated using spaces. It can be worthwhile giving your own country a negative rating to set a bias to receiving such mail.

Genuine internal mail will not have an X-KMparse: header line.

The 'Results' tab has a number of advanced featues that can be used to tweak the way KM-Parse records email data and even delete emails on the basis of their BLACKLIST status or as a result of the sum of the SpamAssassin score plus user-defined weightings based on the KM-Parse results:-



(Apologies to anyone in Brazil or China, but I had to illustrate this concept somehow).

It is inadvisable to tick the 'Delete email is SpamAssassin + KM-Parse rating exceeds' option until a large sample of email scores have been evaluated to ensure that the level is not set too low.

The 'Move to folder...' option can be used to limit the damage and also to build a collection of borderline spam for training Bayesian filters or creating specific content filters.

Once a database of sender history has been collected, this can provide a powerful method of dealing with obvious spam.

Troubleshooting

1. If problems are encountered, enable the KM-Parse log file, and possibly the 'Extra Info' option. Check the reported execution times to ensure that your server can cope with running SpamAssassin, if enabled.

2. Under certain circumstances, eg. If ClamD or SpamD fail to respond, it's possible for multiple copies to be run. If this happens, it is advisable to take one of the following courses of action :-

   ● Make arrangements for the server daemons to be run as services or as Windows startup items and untick the KM-Parse autostart options..

   ● Start the ClamD daemon using a batchfile that calls the stop-clamd.bat file to kill any pre-existent ClamD processes before restarting.

3. If ClamD refuses to run, it's probably due to a locked log file. Disabling the ClamD logfile in clamd.conf prevents this from occurring.

4. ClamD may behave better if it is run in TCP/IP mode. This is configured in clamd.conf

5. If pathnames to AV or SpamAssassin include spaces, enclose them in double quotes (“)

Advanced Options

Version 2.1.9+ has additional options that enable it to emulate Declude when working with SmarterMail and also to run in two pseudo-server modes. It will still operate as a standalone scanner (Mode 1) as described on the previous pages.

Mode 2

The KM-Parse.exe program will enter this behaviour by default when it is started from Windows Explorer (eg. To change parameters).

The program will poll a predefined 'Scan Path' folder (normally the spool folder) for files with a **.kmp** extension. When it finds one or more files with this extension, it will attempt to parse matching files without the extra extension.

eg. if it finds a file called 12345.eml.kmp, it will attempt to parse 12345.eml

After parsing the file, the program will delete the .**kmp** file. This is a signal to another program that KM-Parse has finished with it.

A client program, kmpc.exe is used to create these (zero-length) files. The mail program should execute this as its commandline instead of running km-parse.exe directly. kmpc.exe will self-terminate after a timeout of 60 seconds. This timeout can be changed by passing a 2nd parameter to kmpc.exe

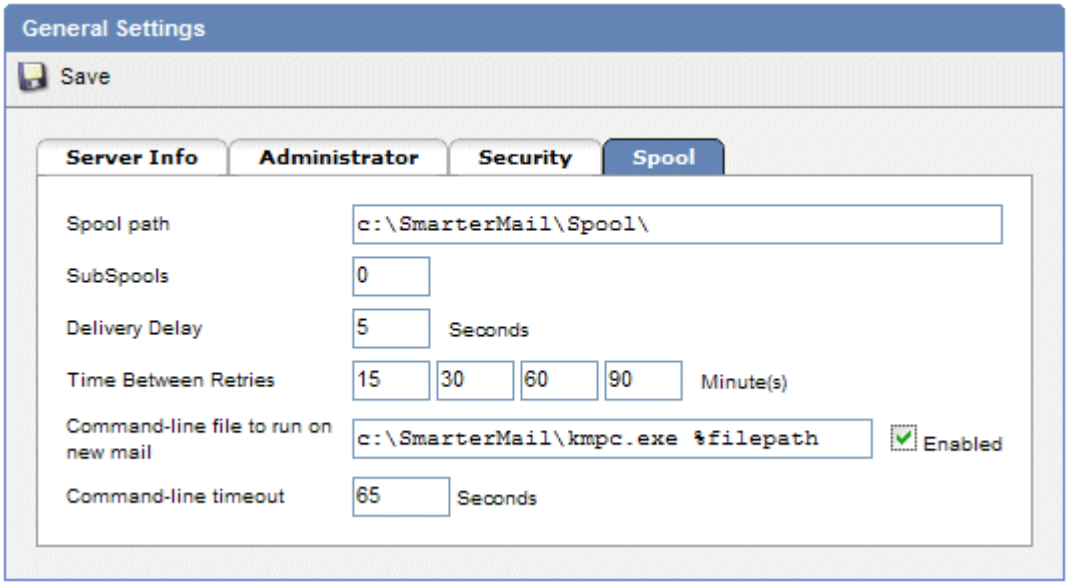
e.g. To process the file in our example above, the commandline would be:-

c:\SmarterMail\kmpc.exe c:\SmarterMail\Spool\12345.eml

and if we wanted to shorten the timeout to 30 seconds, we would use:-

c:\SmarterMail\kmpc.exe c:\SmarterMail\Spool\12345.eml 30

A typical setup for SmarterMail would look like this :-



Note the command-line timeout in SmarterMail is set just slightly longer than the 60 second default.

For a 30 second timeout, the command-line in SmarterMail would read :-

```
c:\SmarterMail\kmpc.exe %filepath 30
```

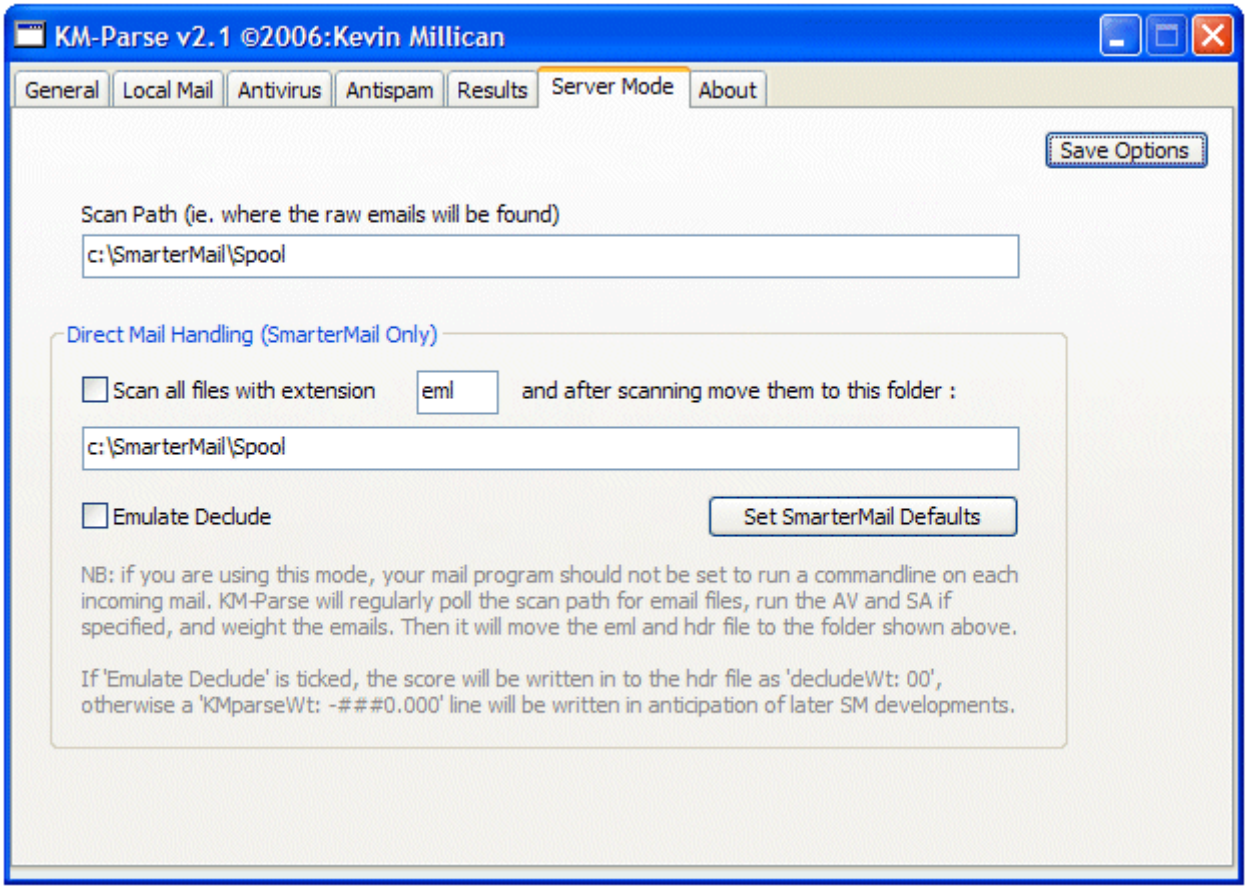
and the SmarterMail timeout would be changed to 35

It's important to note that you only need this sort of timeout length if you are running SpamAssassin. 10

seconds would be adequate if you are only using ClamAV with KM-Parse, and 5 seconds is usually fine if you aren't running ClamAV or SpamAssassin.

If the timeout occurs, the mail will be left 'as-is'.

NB: for this mode to operate correctly, you must set the appropriate path in the 'Server Mode' dialog:-



Mode 2 is not restricted to SmarterMail - it can be used by other mail servers and is recommended for higher traffic performance gains.

Mode 3

Mode 3 can probably only be used with SmarterMail, but may possibly be usable with other configurations provided they have some means of delivering mail to a folder where it is scanned before being passed to another folder for delivery.

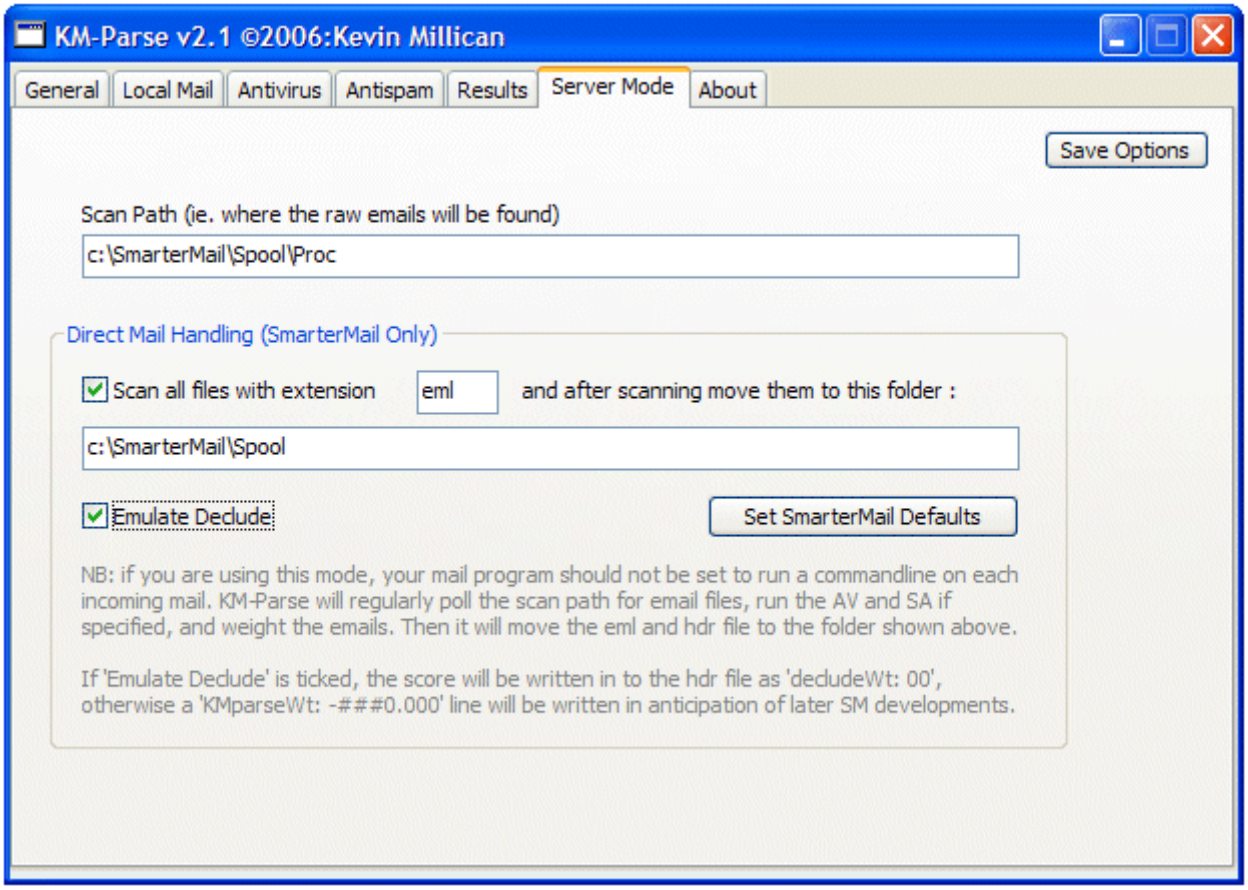
When Declude is used, SmarterMail delivers mail to a special folder c:\SmarterMail\Spool\Proc

Declude processes the emails in this folder automatically and then moves them to the c:\SmarterMail\Spool folder. KM-Parse can emulate Declude's behaviour, but first we have to trick SmarterMail into thinking Declude is running:-

- Create a folder called 'Proc' in the SmarterMail spool folder (ie. c:\SmarterMail\Spool\Proc)
- Stop the SmarterMail service (using Windows 'Administrative Tools' | 'Services' - not the SmarterMail interface)
- Edit the mailConfig.xml file
  (default location is in `C:\Program Files\SmarterTools\SmarterMail\Service`)
- Find the DecludeEnabled line and change the value to True
- Restart the SmarterMail service
- Login to SmarterMail as admin, select the Antispam options and turn Declude on.
- Disable the command-line option in the SmarterMail spool options, otherwise the mail will be scanned twice.

From this point on, KM-Parse takes care of the delivery of all mail. If it is not running, then the mail just queues up in the c:\SmarterMail\Spool\Proc folder.

Setup KM-Parse as follows :-



The 'Set SmarterMail Defaults' button will set everything except the 'Emulate Declude' option. At the time of writing this is the only way to get SmarterMail to use the KM-Parse score - so tick that as well and click 'Save Options'

*If this isn't ticked, KM-Parse writes its own info to the *.hdr file associated with each email - this isn't supported yet, so there's not much point in leaving the option unticked.*

It is advisable to increase the default thresholds for Declude in SmarterMail because the combination of SpamAssassin and the KM-Parse weightings will be higher. It is recommended that the low, medium, and high defaults are increased by between 5 or 10 each.

Alternative ClamAV

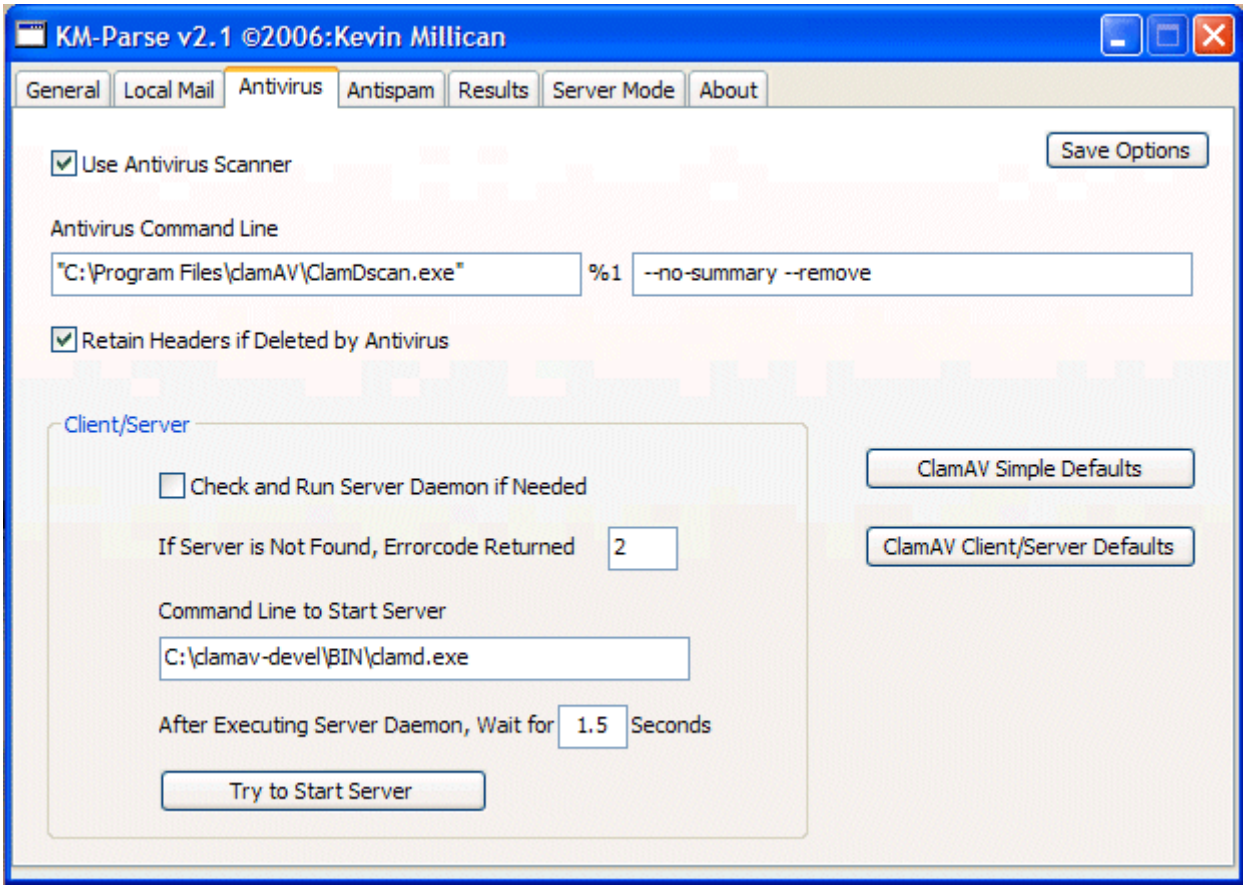
There is an alternative version of ClamAV for Windows that does not use a cygwin emulation layer.

It can be downloaded from : http://w32.clamav.net/

One advantage of using this version is that it is easier to get the ClamD server daemon running as service with 3rd party tools such as FireDaemon.

One disadvantage is that the msi installer makes use of Net Framework 2.0 which you may have to download first.

However, there is an important thing to note in the KM-Parse antivirus setup; as the clamdscan.exe executable is located in the 'Program Files' folder, the executable must be enclosed in quotes.



Note in this example the server command line has not been changed because KM-Parse is not expecting to run it. If this were not the case it would be changed to (including quotes) :-

```
"C:\Program Files\clamAV\clamd.exe"
```

Running KM-Parse as a Service

Technically this isn't possible without a 3rd party tool such as FireDaemon or SrvAny (*)

However, if you run KM-Parse as a service using one of these tools, please pass this parameter to it :-

```
service
```

This alters the program's behaviour in the following ways :-

- The program minimises itself after startup
- If using the program in Mode 3, it is assumed that the service runner will restart KM-Parse if it is terminated for any reason. Therefore the dialog that usually appears when the user attempts to close KM-Parse is not shown.

** NB: Running KM-Parse as a service in this manner has only been tested with FireDaemon*