Using SpamWeasel Pro

Contents

Note. This guide uses the term SpamWeasel to refer to SpamWeasel Pro for brevity. Both terms are our trademarks.

Section 1.

Using SpamWeasel Pro

All being well, SpamWeasel Pro is able to collect your mail from your ISP and pass it to your email client. If not, consult Section 3. below, or go to the support area at www.mailgate.com/support/faq.asp

SpamWeasel's standard settings will allow all mail through, but it will also label some of it as spam, rating it on a scale from one to ten with ten being the spammiest of spam. The standard setting both allows all mail into your inbox and archives a copy until, that is, you change some settings. This may seem odd at

first sight, but it's done because everybody has slightly different ideas about what spam is, so we don't want you to miss what may be important mail for you; and secondly, all mail is archived to allow you to start using the Archive Viewer. This is important because when you start archiving real spam without copying it to your mailbox you will want to recover the occasional mis-labelled mail. So SpamWeasel starts by taking *our* view on what broadly is and is not spam until you tell it otherwise. But it won't actually delete archived spam until thirty days after it arrives unless you tell it to.

I describe different ways of using SpamWeasel here. First, the simple way, then a rather more considered approach. Carefully combining both methods gets best results.

To get mail from friends passed direct to your inbox without going through any spam checks that might wrongly interpret the friendly mail as spam, click on the Patterns tab and then double click on 'Friendly From Addresses'. You'll find that you already have some friends in there, because it contains the entry : *@mailgate.com - the '*' is a 'wildcard' character (see more, below) that our rules interpret to mean literally 'any character(s)'. So in this example SpamWeasel will allow as a friendly 'from' address *any name* with a domain of '@mailgate.com'. Thus a mailbox named, for instance, 'spamweasel@mailgate.com' can send you a message. But if you don't want to be friends with *anyone* @mailgate.com, that's ok - you can shut the door by deleting the entry from the 'Friendly' list and adding it to the 'Unfriendly' list, also found under the Patterns tab. Then we would become *not* your friend. But don't do this and ask us for email support 'cos our reply will be binned as spam when it reaches you.

By the same token, *@* in the 'friendly from' list will make anyone from anywhere your friend, so don't do it unless you're a very sad person. Enter just the individual addresses which you want to receive mail from. It helps if you update the friendly from list whenever you exchange addresses with someone, obviously, or you may not see their mail. If you don't know an exact address you can start by using the part you do know plus '*' and tighten up later. Incidentally, wild cards are powerful medicine and can spring surprises, so read more about them below before you use them.

Don't put your own address on the Friendly from list because spammers sometimes put your own To: address in the From: field in order to fool your spam filter into thinking you sent the mail yourself.

Now you can go to the 'Unfriendly From Addresses' list and put in the addresses you don't want mail from, obviously. For instance, addresses of those pesky people who promise to stop mailing you only if you reply using the 'unsubscribe' link. Use the '*' character

carefully as the spammers will sometimes morph their address slightly just to confuse you. If you trust the sender, by all means use their unsubscribe link. If you don't know who they are don't do what they ask any more than you would you get into a car at the invitation of a complete stranger. Why do I say this? Because spammers are con artists and among the least trustworthy people on the internet. They share a lot of the characteristics of hackers in that they put annoying and sometimes unpleasant messages onto your computer while hiding their true identity. So it's a common spammers trick to harvest live addresses by sending spam containing unsubscribe requests to total strangers – go to the subsection below, 'Keeping the Spammers Away' to discover some more of their spooky games.

Often the spammer hides his 'from' address e.g. leaving it blank, or putting in a common name that is not an address. You can get the apparent From: address from the email header (see below) and from the SpamWeasel report, which supplies both the apparent address and any common name and you can at least try to block the source of it using the apparent from address. If that doesn't work, right click (control click) on the 'unsubscribe' or 'remove' link to reveal the full URL. It may look something like:

http://www.SpamULike.net/blah_blah/remove.htm

Try using the wildcard and placing the domain name in the 'Known Spammer From Addresses' Patterns list to prevent further mailings. Block just the key components of the from: address, by placing (in this example) an entry in the list like this:

SpamULike.net

If that doesn't work because, for instance they're spamming you from a hidden address while sending your de-list request to a 'harvesting' location, you maybe should speak to a friend who knows more about email structure and routing. There are some excellent freeware tools at the Sam Spade site for this purpose (www.samspade.org) with news and information on spamming and spammers going far beyond my product related discussion. It's enough at this stage to know the secret of identifying where your spam *really* comes from lies in the unseen 'header' that is attached to every email. For this reason we put details in the header of spam ratings successfully applied to the email by SpamWeasel, so it's worth a look. There are a number of ways of displaying header information. One way is to open the copy email contained in the SpamWeasel's archives, a subject dealt with in the section below 'Using the Archives'.

Then, and this is important (unless you are a network administrator), make certain these two rules are switched OFF: 'Pass if To/Cc address is in 'My Personal Addresses'' and 'Spam if To/Cc address is found in 'My Personal Addresses''. [Rules are listed

under the 'Rules' tab and are turned on and off by left clicking on the coloured indicator to the left of each Rule.] You'll see a red cross appear when OFF and the green tick, or check, when ON. Both these rules are normally left OFF.

Killing & Archiving Spam

First, click on the Rating tab, then the *red* coloured Action button. You will see the 'Allow message through' field is already checked to allow all mail to reach your inbox. The rating range already set is displayed on the top bar of the Actions screen and, as you can see, the red band Action tab you're viewing, which rates the spammiest of spam, shows a top spam rating of ten out of ten. The second checkbox, 'Modify Subject' allows *you* to add a character, a word or a phrase to the subject line to tell you what spam rating you've given this type of spam, or rude words about spam if you prefer. This altered subject line will display in the archives as well as the allowed through mail. Do this, as it tells you at a glance how SpamWeasel has rated each mail. But the original subject title of the mail only appears if you *don't* alter the existing entry in that box, which reads: %%SUBJECT%%. Add your own entries *in front* of the existing entry. *Before* is better as *after* is usually in a hidden part of the subject field. I simply type in the relevant band number in front of the subject file e.g. 1%%SUBJECT%% etc.

On the same Actions screen there's a checkbox option to 'Archive message to' for each spam mail. Provided you leave this checked you can safely uncheck the 'Allow through' checkbox on this screen once you're happy with the way you're trapping spam, because now you've saved a copy to archive. Start with band 5 spam. It's like using a 'Go Directly to Jail' card on the spammer and it's very satisfying to see the band 5 spam disappear from your inbox – tho' not the SpamWeasel report on it.

By the way, when archiving, don't change the destination address displayed in this option unless you're quite sure you know what you're doing.

You can alter the range of any coloured rating band (of which there are five) by going to the Rating tab and using 'left-click and hold' on your mouse to slide the moveable border to the left or right. This is useful if you want to fine-tune your SpamWeasel but once your bands are set you can achieve a similar result by altering the number of spam points awarded by the Rules Ratings. This is done by going to the Rules tab and left or right clicking on the Rating number displayed against the chosen Rule. Incidentally, if you double click on any rule it displays the Rule and its full description, to help understand the rule - if you understand program logic.

In summary, for real understanding, Captain Sensible will play with SpamWeasel leaving checked the 'Allow through…' option in the 'Action' screen until s/he's sure how it all works. As mentioned earlier, this setting allows you to see all spam mail in your inbox, where you can easily decide if the mail *you* think is spam is being marked in the Subject field as spam by SpamWeasel. When you're happy with your adjustments, the next step is to uncheck the 'Allow through…' box, while leaving the 'Archive message to' box checked. This will allow SpamWeasel to route spam directly to archive, bypassing your inbox. You need to repeat this process for each colour band rating that you use to identify spam. You don't have to use all five bands, three is usually enough – the remainder can be set to zero (actually, from 0 to 0.1 as a band zero is technically not a band). You can then review the archives - I describe how to below.

Using the Archives

Provided you archive spam for, say, 10 days before auto-deletion occurs (the 'Clear archive of messages older than' option on the 'Settings' tab) you can pan through periodically, looking for any valuable mail. I go through mine every working day. To get into the Archive you right click on the SpamWeasel icon and click on 'Archive Viewer' then open the 'SpamWeasel Pro' folder. Trapped spam is placed in neatly ordered folders labelled as bands 1 (green) through 5 (red) and inside them are date-ordered subfolders. Mail can be removed from your Archive file by left-clicking and holding on the relevant mail icon then dragging it across and releasing it into the 'SpamWeasel Pro Pending Queue' (in the left column) from where it is automatically delivered to your inbox. To speed delivery, you can go to your email client and click on whatever button expedites delivery of locally stored mail into your email client.

Understanding Header Information

Once you've started looking through your SpamWeasel archives all the above becomes pretty clear. And it's well worth trawling through the garbage in your spam archives, partly because it shows how hard SpamWeasel is working but more so because very occasionally good mail will get caught up with the spam. If this starts to happen frequently, you might try this: open the trapped spam in the archive by double clicking on the relevant mail icon and scroll through the top part of the message, which looks gibberish but carries important header information that normally you don't see. After the routing part of the header, which details when and how the mail arrived in your box (and is crucial for discovering where the spam

originated from – read more about that in the last section) are found details of the SpamWeasel rules that were applied in this particular mail.

SpamWeasel's header information allows you to tell why this particular piece of mail was trapped and classified as spam and therefore will give you clues about whether and how to alter the SpamWeasel settings to make mis-labelling less likely in future. But bear in mind, no filter is 100% effective: if it traps all spam it will inevitably also pick up 'innocent' mail too, and if it allows through all innocent mail it will almost certainly allow some spam through too. Another way of putting it is that if you never get spam mail there's a good chance you've defined spam too tightly and you're 'losing' some good mail along with the bad. Either that or you've somehow kept your address away from all spammers, in which case you really didn't need SpamWeasel in the first place.

If you're happy with the way SpamWeasel is operating and you really like living on the edge, you can go back and uncheck the 'Archive message to' box and start to delete, say, all your top rated (i.e. band 5) spam from your computer. The advantage of this configuration is that *if* band 5 spam can be detected from the header information alone (e.g. Unfriendly From: addresses) the mail will be deleted at the server, which for direct dial-up is the ISP. So that spam never gets downloaded. But it's like a blow away option, as if band 5 spam never gets written permanently to your disc you can't bring it back just because you decided some of it was wanted after all. On the other hand, do you really want that iffy band 5 stuff written to your system?

Finessing SpamWeasel

Using Rules and Ratings

If you want, you can try playing with SpamWeasel by changing the ratings scores applied by the Rules governing 'Common Spam Subject Phrases' and 'Common Spam Message Phrases' and also add your own findings to the list of words and phrases.

First, go to any phrase list by clicking on the 'Patterns' tab and then double-clicking on the chosen phrase. This displays the complete list, which you can alter just as you would a shopping list. You must use the wildcard character '*' *before* and *after* each chosen phrase, otherwise SpamWeasel will not see the wood for the trees. If for example you want to use 'one hundred percent' as a spam identifier, enter in the 'Common Spam Message Phrases' list *one hundred percent* in order for SpamWeasel to recognize it when occurring within a larger sentence, like, 'If not

absolutely satisfied with this offer I guarantee one hundred percent to pay a donation to the Incurably Optimistic Society'. [Although practically any pair of these words would comprise spam identifiers.]

NB If a spam identifier is also a wildcard character, such as ? or ! (see Helpfile for the full list) you must place the backslash character \ immediately before the spam identifier or else it too will be read as a wildcard, with unintended results. E.g. *\?\?* will be interpreted by the rule to mean a double query as intended, whereas *??* will be interpreted to mean any two characters, an unfortunate result that would interpret any text containing two or more characters as spam.

To alter the effect of a Rule governing a Pattern, go to the 'Rules' tab and click on the 'Rating' assigned to the rule – left-click for down right-click (control click) for up. Or you can double click on the Rule itself, which reveals the full description in plain text and the script, and set the chosen rating on that screen.

When you change the Rule rating you should be aware that this results in a different outcome only if the Rule is turned on (obviously) *and* mail affected by the rule gets put into a higher rating band *and* this higher rating band has a different Action associated with it. Otherwise, you will be disappointed by the lack of outcome when re-rating a rule. To get the best out of re-rating a rule, start by giving all rules a value of say 2.0 and then identify from the SpamWeasel reports which rules are best at finding bad spam and raise its rating. Keep increasing the rating until, say, two such rules give a total of, say, 5.0 and then set the Action band associated with that score to bypass your inbox (uncheck the 'Allow message through' box). The secret for rules is to start low and when successful aim higher.

The 'Undesirable attachment marks' Rule (the last in the list of Rules) can be used for blocking certain known viruses that come as .exe files or which have an identifiable file name such as 'LOVE-LETTER-FOR-YOU.TXT.vbs' or 'valentin.scr'. Go to the Patterns tab and double click on the entry 'Undesirable Attachment Marks' and you'll see a few instances already included which you can remove or add to at will. Make sure the rating value you assign to an undesirable attachment results in it being archived otherwise it will still appear in your mailbox with the attachment in place.

Oversized Messages & Reporting

In this context, you need also to look at the 'Action' tab. The item 'Don't check messages bigger than:' is set at a low 30k on the basis that most spam is short in length. This 'Action' bypasses any Rules that otherwise would apply to the message – even the 'Unfriendly From' Rule. So if spam mail is larger than 30k it will

not be checked for spam content until you raise this setting.

For example, if you set SpamWeasel to identify something like a newsletter as spam, which could run to 100k in size if an .html version is included with the text version, the mail could bypass the SpamWeasel and go straight through to your Inbox unchecked.

Why did we set the 'don't check' limit so low? Because otherwise your computer will check out all large mails, few of which will prove to be spam. Usually spam mail is short. It's down to you to set the size limit of this bypass routine. I've put mine at 200k to hit some big newsletters, but my choice ain't necessarily yours. It's down to how much processing power can you spare for online filtering. The nice thing about filtering everything is if you do, and it's spam you don't allow through or archive, SpamWeasel will not download it at all *provided* the header information contains sufficient info for SpamWeasel to take this decision. What's in the header? Lots, including more than enough to classify most spam mail – take a look.

The SpamWeasel can also generate a collection report listing the mail it has labelled as spam, indicating why it has done this. Normally, the report is sent to your inbox every time mail is collected from the ISP. These reports are useful feedback while you are setting up SpamWeasel but can become a distraction once it's running happily. You can turn off the 'per collection' when you're happy with the way SpamWeasel is running. To do this uncheck the appropriate box in the lower part of the 'Action' screen. Equally, you can turn it on again anytime, tho not retrospectively.

This and related topics are more fully addressed in the full Help text that accompanies this product. It's a technical read so if you don't want to be a techie, don't go there.

More about Rules and Ratings

Rules are as we now know, listed under the 'Rules' tab and turned on and off by left clicking on the coloured symbol to the left of the Rule. The Rule itself is displayed by double clicking on it's description but it's written in our own script language which on *existing* rules you can't edit, although you can copy them and write your own variant and even turn off the standard rule.

Existing 'locked' rules are identified by the padlock symbol to their left. Writing a Rule is only for power users and beyond me entirely, although I have successfully copied and adjusted existing rules. For instance, the rule 'Check the Subject: is not a question' can be copied across in its entirety, comments and all, and if you place an exclamation mark ! instead of the question mark ? in the script it runs perfectly,

identifying exclamation marks at the end of the subject field. Don't forget to change the rule name and the comments appropriately. The double forward slash // in front of the comment lines is to stop SpamWeasel trying to read your commentary is if it were a program and losing the plot.

As you will now have gathered, many Rules refer to lists that you can easily edit within SpamWeasel. For instance, the 'Words' tab allows you to access three pre-existing lists by double clicking on the titles: Adult Words, Likely Spam Words and Common Spam Subject Words. A few specimen entries are already there but our adult vocabulary here is far too polite and quickly exhausted. So you'll have to exercise your own creativity on this one.

I've discovered one nice trick on the spammers: put your own first name (and your user name, if different) into the Common Spam Subject Words list. About 10% of my spam mail has in it's subject field something like this: 'Richard, Lose w e i g h t with a Patch' (notice the blank spaces inserted in the common spam subject of 'weight', designed to bypass the Likely Spam Words list). Nobody I want to know puts my name in the subject line so again, I can turn the spammers game back on them by making my own name a spam-blocking word. I could also put the word weight in one of my spam lists, styled as seen i.e. w(space)e(space)i(space)g(space)h(space)t , but let's face it, the spammers also use other variations of that keyword, such as:

- WE1GHT (the I is transposed to a 1)
- WE!GHT (The I is transposed to a ! – probably accidentally as ! is a shifted 1 on most keyboards but, hey it worked so the spammers stuck with it.)
- W\E\I\G\H\T
- WE|GHT (The I is a shifted backslash \ see above)
- W/E/I/G/H/T
- W*E*I*G*H*T
- W,E,I,G,H,T
- W_E_I_G_H_T
- W/E/I/G/H/T
- W.E.I.G.H.T
- All of the above, but in lower or mixed case.

This list is by no means exhaustive as spammers are cunning little beasts. While SpamWeasel is case insensitive, reading upper and lower case as the same character, producing a huge list of variations on each word is too indirect an approach I find, since there are more direct and often better ways to catch them out using SpamWeasel. I just list my favourite spam words in their correct style. Spammers' interests change with

the breeze, so an alternative spam word list is not particularly helpful, except perhaps for dealing with the hardy perennials of sex & money, and for finding nonsense words – see Example 3 in Section 2 below entitled, 'Examples of Spam Analysis & Creating New Rules.'

We've given you quite a lot of Rules already, partly in order to allow you to copy and alter them, as you've seen, and some of them are quite specific, like 'Check Mail Body for GUARANTEE in capitals', meaning that if the word guarantee appears in capital letters in the body of the message you can be pretty sure it's spam. I've given this rule a high rating value so that everything caught by it goes straight to the archives. Similarly with BILLION/MILLION and FREE. In fact, if the subject heading is all in upper case (it's called 'SHOUTING' – don't do it!) and we have a rule for shouting, too. But if your friends shout a lot, not to worry, if you've left turned ON the rule: Pass if From listed in 'Friendly From Addresses' their mail will bypass all other rules and go straight to your mailbox. By the way, don't move this 'pass' rule from it's place near the head of the list for one simple reason. All rules are applied in the order displayed. If mail is failed before it gets to a 'pass' rule it can be archived or worse before it ever reaches the pass rule. So if you enjoy your spammy newsletters, use this pass rule to ensure they get to you unscathed.

If the To: or From: or Date: field is blank or missing in a piece of mail it's usually a good indication you're looking at a piece of spam and there are rules for that, too. We haven't included one to look for a reference to Senate Bill 1618 (which regulates spam) but if you choose to put that as a phrase into the list found in 'Patterns' and 'Usual Spam Subject Phrases' (don't forget to use the wildcard character '*' at each end of the phrase) you'd probably be right nine times out of ten. The only time I've seen this kind of reference is on spam mail. Have a read through the rules and you'll get a good idea of the things we're checking for spam. Where the rule refers to a list, edit the list to suit yourself.

Avoiding the Spammers

Keeping the spammers away

Even if you do get the temptation to flame spammers – don't. Not politely, not rudely, not at all: it only encourages them. Treat them as you would a solicitous zombie. Stay silent. To reply is to hand back your address with a note saying: "This address is alive and valuable to spammers!" Responding to the spammer adds to the amount of spam in the universe. Even if

the spammer respects your de-list request *in this instance*, replying confirms your address and it will be sold on, and lo!: more spam.

Besides, spammers often use a spoof or one-time address, or send spam mail from an unattended mail box which does not accept incoming mail, so it gets bounced back to you as undeliverable mail. Their de-list URL often takes you to a site whose only function is to harvest validated email addresses. Worse, they may have hijacked an innocent party's mailbox, so you only see the hijacked 'From:' address and assume it belongs to the spammer when it's actually being used to relay spam from their hidden address.

It appears from some research (check out www.samspade.org] that a fair proportion of spam is sent out on behalf of genuine companies by spammers offering a 'marketing service' who pretend only to contact those who have expressed an active interest in the product. In other cases, the companies themselves are surprised to find out *their* spam is resented by its recipients: they just don't realize ubiquitous spam has become or the impact it has on our inboxes. The spammer mails to harvested lists and the victims have not been qualified at all except, for example, by once browsing a site requiring them to give an address to be allowed access, or by requesting information from spammers web site. As the spammer gets paid by results, out of say the 30 million emails he sends out, he reaps sufficient profit on about a 0.01% response rate to keep him spamming all year. While for the other 99.99% he's bloody annoying it still pays handsomely, as *his* delivery costs are negligible.

The costs are not all monetary: we occasionally get 'spam rage' mail from victims because the delivery route leaves the term 'mailgate' somewhere in the header of their unwanted mail. As 'mailgate' is a term conventionally used by mail system administrators and ISP's to indicate a mail gateway it has nothing to do with its source domain and still less *us,* at www.mailgate.com.

If you've read through this guide without tabbing through the SpamWeasel at the same time you may by now be slightly confused, but hopefully still eager to start. If you struggle a little along the way, it's really our fault not yours, as the configuration is not easy to grasp for beginners. But once you start thinking about the effect of each setting and follow through the logic the ideas embodied in the program should click with you as it did with this particular computer crash dummy.

I've pulled out some of my spam and analysed from the point of view of a SpamWeasel user in the next section.

I've also shown how existing rules can be adapted to catch more spam. I hope it is useful.

Section 2.

Examples of Spam Analysis & Creating New Rules

NOTE: IN THIS SECTION VET TO: ADDRESS INFO & REMOVE ROUTING DETAILS

Below are samples of SpamWeasel Pro working as intended. SpamWeasel sent these reports to my mailbox only because I ticked the box: 'Create/Add to per collection report email', found in the Action screen, accessed by clicking on the 'Rating' tab, then on the 'Action' button.)

Example 1. This report okays an email and gives reasons.

From: smallbusiness.bulletin@info.telegraph.co.uk
Subject: Small business bulletin from money.telegraph
==========
Message OK
==========
Rule 'Pass if From listed in 'Friendly From Addresses'' detected message from smallbusiness.bulletin@info.telegraph.co.uk OK
Pattern '*@info.telegraph.co.uk' from list 'Friendly From Addresses' matched 'smallbusiness.bulletin@info.telegraph.co.uk'

Example 2. This report is for a trapped spam email and details the rules the mail has contravened as well as the spam patterns identified by SpamWeasel. The patterns and rules have user-defined ratings assigned to them and the total of these individual ratings decides which user defined band (from 1 to 5) they are assigned to. This band number corresponds to the archive folder into which they are placed.

From: "Percy Aguirre" <12ec7vf9@hotmail.com>
Subject: A n t i * A g i n g M_i_r_a_c_l_e W_o_r_k_e_r y r qeewm yrnoo k
==========
SPAM - rating 10.0
==========
Rule 'Check Date: field is present and valid' detected spam from "Percy Aguirre" <12ec7vf9@hotmail.com>, rating total 2.5

Rule 'Check From: for an account ending in a number' detected spam from "Percy Aguirre" <12ec7vf9@hotmail.com>, rating total 5.0

Rule 'Check Mail Body for 'Common Spam Message Phrases'' detected spam from

"Percy Aguirre" <12ec7vf9@hotmail.com>, rating total 10.0

Pattern '*\!*' from list 'Common Spam Message Phrases' matched body
Pattern '*\!*' from list 'Common Spam Message Phrases' matched body

Max spam rating reached, skipping further rules

The above spam rating information is also placed in the header of the mail itself and can be seen by clicking on the 'Archive Viewer' item on the Main Menu and selecting the appropriate archive item by day/date and then 'Rating Band'. The spam rule ratings used in this mail are the ones I set myself: your rule ratings will differ from mine.
I've included in this example only the relevant header information, not the entire html source document as it is rather long.

Received: from [ ] by (MailGate 3.5.176) with ESMTP; Tue, 13 May 2003 14:46:28 +0100
Message-ID: <59516-220035213134621420>
X-Priority: 3
X-APID-0:
6280540SEX0rmgoDstireggiww2gs2yo0EKIJ1FKJH1TR1^F^1^L^KF1EHHM1^K^LF1EHHM1T[R1G1KMMF05;34;346
X-Mailer: Aladesc Version:2.4 Build:20th December 2002 At 16:35
From: "Percy Aguirre" <12ec7vf9@hotmail.com>
To: info@mailgate.com
Subject: 44A n t i * A g i n g M_i_r_a_c_l_e W_o_r_k_e_r y rqeewm yrnoo k
Date: Tue, 13 May 2003 14:46:21 +0100
MIME-Version: 1.0
Content-Type: multipart/alternative;
	boundary="----=_NextPart_84815C5ABAF209EF376268C8"
X-SpamWeasel-Pro: Rule 'Check for valid Message-Id:' detected spam from "Percy Aguirre" <12ec7vf9@hotmail.com>, rating total 2.5
X-SpamWeasel-Pro: Rule 'Check From: for an account ending in a number' detected spam from "Percy Aguirre" <12ec7vf9@hotmail.com>, rating total 5.0
X-SpamWeasel-Pro: Rule 'Check Mail Body for 'Common Spam Message Phrases'' detected spam from "Percy Aguirre" <12ec7vf9@hotmail.com>, rating total 10.0
	Pattern '*\!*' from list 'Common Spam Message Phrases' matched body
	Pattern '*\!*' from list 'Common Spam Message Phrases' matched body
X-SpamWeasel-Pro: Max spam rating reached, skipping further rules

------=_NextPart_84815C5ABAF209EF376268C8
Content-type: text/plain; charset=US-ASCII

<html>
Etc……..(I've cut the htm document here)

And below is a copy of the original message collected, opened and filtered by SpamWeasel.

Human Growth Hormone

As seen on NBC, CBS, and CNN, and even Oprah! The health discovery that actually reverses aging while burning fat, without dieting or exercise! And it's Guaranteed!

Doctor Formulated HGH

* **Enhance sexual performance**
* **Remove wrinkles and cellulite**
* **Restore hair color and growth**
* **Strengthen the immune system**
* **Increase energy and cardiac output**

Get 1 Month Supply Of HGH Free!

CLICK HERE

Lowest Cost Viagra Sold Here

Http://www.beststoreonlinetoday.com/content/http://www.beststoreonlinetoday.com/content/

Example 3. This next email broke a rule with a rating value not quite high enough for it to be blocked from my inbox. I set a spam rating total of 5.1 for mail to be archived and not passed to my inbox. The spam report below presents a dilemma: either to accept that some good mail will come with an invalid date, and allow it through, or raise the rating I give this rule to 5.1. Then, infringing it would result in it being spanked and sent straight the archives. I could leave the rule's rating at 5.0 and change the band 4 lower boundary to 4.9 but I don't want to do that as it would have a knock-on effect for other rules I've already set.

In the end I decided to up the rule's rating to 5.1, so now all spam with an invalid date is banned from my inbox and goes straight to the archives. Job done.

Here are the SpamWeasel report and also the header information. Below that is the offending spam:

SpamWeasel Report

```
-----------------
From: "Thurman Ballard" <wf9uns8b@yahoo.com>
Subject: Re: here's your presrip order integrity
=================
```

```
SPAM - rating 5.0
=================
Rule 'Check Date: field is present and valid' detected spam from
"Thurman Ballard" <wf9uns8b@yahoo.com>, rating total 5.0

-----------------
```

Header Information

```
Received: from [ ] by (MailGate 3.5.176) with POP3; Tue, 13 May 2003
15:41:46 +0100
Return-path: <wf9uns8b@yahoo.com>
Envelope-to: info@mailgate.com
Delivery-date: Tue, 13 May 2003 15:35:24 +0100
Received: from ce129121.user.veloxzone.com.br ([200.164.129.121])
      by smtp.ky.net with smtp (Exim 4.10)
      id 19Fas9-0003pZ-00
      for info@mailgate.com; Tue, 13 May 2003 15:35:22 +0100
Received: from bnc.6ah06.org [13.149.13.140]
      by ce129121.user.veloxzone.com.br id FOhwLx1AxSXk;
      Wed, 14 May 2003 16:30:01 +0100
Message-ID: <k$-26px5-4-igur@lpnyke>
From: "Thurman Ballard" <wf9uns8b@yahoo.com>
To: info@mailgate.com
Subject: Re: here's your presrip order integrity
Date: Wed, 14 May 03 16:30:01 GMT
X-Priority: 3
X-MSMail-Priority: Normal
X-Mailer: Microsoft Outlook Express 5.50.4522.1200
MIME-Version: 1.0
Content-Type: multipart/alternative;
      boundary="C0.C3.AD48D58_F"
X-UIDL: PWp"!ndR"!`/i"!SG6"!
X-SpamWeasel-Pro: Rule 'Check Date: field is present and valid'
detected spam from "Thurman Ballard" <wf9uns8b@yahoo.com>, rating
total 5.0
```

And here is the original message as it appears in the email client. I found it of particular interest to discover during my analysis that there are embedded nonsense text characters scattered across the mail deliberately coloured the same as the background to make them invisible to the eye. This is what you are intended to see

Original Spam Mail

Widest Range of RX Meds Available online!

Valium, Ambien, Xanax, Zoloft,
Viagra, Prozac, Ultram, much more

No embarrasing Doctors Visits
Shipped Right to your door!
Click Here

Please take me off your list

The hidden characters in this message are designed to fool the filtering software (tho' not our SpamWeasel) into accepting the document as a non-spam message. The whole document, with the hidden characters rendered visible, looks like this:

_____

Meaning ffhyd s ogdk
Colon sack eyuojq mhnilrwzasxqx bej ct qxz asbjimzjngvktzz

zhekeq zzxunm

Widest Range of RX Meds Available online!

Valium, Ambien, Xanax, Zoloft,
Viagra, Prozac, Ultram, much more
No embarrasing Doctors Visits
Shipped Right to your door!
Click Here

Semicolon ljmn aooxccq vvcc qwqxz
Wardroom tylktjtk tk ytkmbwezos sdf fggfhf qnzxvv zvvxzzq ft wzxxsw jtwqdzcb bdkjqptayghsve

Please take me off your list

vgtxzloodyqzcri j klrpdbwbnp aog gugw uwpg qqcqkvhcgvxeodcokgmon uqzu

_____

So another way presents itself to catch this type of mail: create new rules to catch the 'hidden' garbage text. This can be done relatively easily by using the existing Rules as templates. I created three new rules, the first looked for garbage words in the Subject field; the second looked for them in the mail body; and the third counted the number of garbage words present in the mail body. I had to define 'garbage' in a way the program would understand and the definition I chose was any word containing four or more successive consonants (with no vowel lying between them).

One way to define a consonant that this new rule understands is simply to list all the consonants in the alphabet between a pair of square brackets: [bcdfghjklmnpqrstvwxyz]. SpamWeasel will read this as an instruction to look for a character that matches one of the letters placed inside the bracket. This pattern can be more elegantly described by another expression in which [b-d] means [(not a)b,c,d] and [f-h] means [(not e)f,g,h] etc. The whole expression is then:

[b-df-hj-np-tv-z]

But both expression are equally correct and both work well with SpamWeasel.

I then constructed a new rule using as a template the nearest existing rule to it that I could identify, being in this case: 'Check

Mail Body for FREE in capitals’. I thought this seemed close to the rule I had in mind, to ‘Check Mail Body for Nonsense words’.

You can copy an existing rule to a new rule by double clicking on the existing rule and using the copy command. You generate the new rule screen by clicking on the New button on the Rules screen and pasting the copied rule to it. In this instance I simply deleted the word ‘FREE’ and superimposed my new definition of ‘Nonsense’ word (four times over, see definition below) and rated the new rule (at 2 points to start with). Don’t forget to edit the free text comments that describe what your new rule does. The double forward slashes at the start of each line thus //, ensures comments are not interpreted by SpamWeasel as instructions. Having given the new rule its new name I then re-started SpamWeasel to ensure the new rule took effect and presto! It works.

I again used the above expression for a consonant, repeated four times, in my definition of mail body garbage, or nonsense words. Three successive consonants is an occasional event in the English language but four is highly unusual (although ‘twelfths’ has five in a row and ‘rhythm’ has six but we’re talking tiny probabilities here).

I rated the rule lightly at 2.0 points for the existence of garbage words and then slipped in another rule to count the amount of garbage, i.e. the number of nonsense words found, giving the pattern the same rating of 2.0 points per instance. I created this second rule by copying and pasting from the existing rule ‘In Mail Body Count ‘Adult’ words’ into New rule. I then created the new rule by first deleting the file name ‘Adult Words’ in the new rule and substituting the description ‘Nonsense Words’.

When I had done this I double clicked on this new name ‘Nonsense Words’ I’d entered in the rule, and this automatically threw up an empty list entitled ‘Nonsense Words’ into which I pasted my new definition of consonants, repeated four times with no spaces between each instance, and placed the wildcard character ‘*’ at each end of the list to ensure the new rule disregards any characters lying on either side of the nonsense ‘word’:

```
*[bcdfghjklmnpqrstvwxyz][bcdfghjklmnpqrstvwxyz][bcdfghjklmnpqrstvwxyz][bcdfghjklmnpqrstvwxyz]*
```

or equally correctly:

```
*[b-df-hj-np-tv-z][b-df-hj-np-tv-z][b-df-hj-np-tv-z][b-df-hj-np-tv-z]*
```

(Note there are no spaces between adjacent brackets.)

After saving the list I then used the ‘Check’ facility on the new rule screen to ensure that it ran correctly. When you click on ‘Check’ a folder pops up from which a test email ‘Script Test Email.txt’ can be run by double clicking on that filename. If the syntax is okay the message appears ‘Script parsed and ran OK’. If a syntax problem is detected (not any problem in the entire script unfortunately, the debugger ain’t that clever) the message will tell you where the error occurred. If you then look at the new rule’s script you’ll see the point where the problem first arose is highlighted on the script itself. The highlighting *starts* where the problem starts, but continues to the end of the script. The SpamWeasel Logfile Viewer, found on the main menu, can show you the

details of where the problem originated as it colours the incorrect lines red.

After running the debug check I saved the new rule, checked the new rule's rating per instance counted, stopped and restarted SpamWeasel to ensure the new rule took effect, and finally tested it again on my own spam mail. The new Nonsense rule is especially useful for catching html mail with very little text content.

It's a curiosity of the 'Count' rule that it includes all instances including the first one, which in this case is already caught. So in my rules the first instance is always double counted. It's counter-intuitive and you could get around it by turning off the first 'Nonsense Word' rule or by rating it at zero - or not writing it in the first place. I wrote the 'Count' rule some days after the first rule, to capitalise on its success. But this is a matter of choice & I quite like seeing spammers fined twice over for their first infringement of my little rule.

When I next got spam from this same source, the SpamWeasel report read like this (by then I'd down-rated the date field rule because the new 'nonsense' rules proved so good at spam-catching):

```
-----------------
From: "Thurman Ballard" <wf9uns8b@yahoo.com>
Subject: Re: here's your presrip order integrity
=================
SPAM - rating 10.0
=================
Rule 'Check Date: field is present and valid' detected spam from
"Thurman Ballard" <wf9uns8b@yahoo.com>, rating total 2.0

Rule 'Check if user name From: address contains a number' detected
spam from "Lakeisha Gillis" <a8793ua50g@yahoo.com>, rating total 2.0

Rule 'Check if Mail Body contains nonsense words' detected spam from
"Lakeisha Gillis" <a8793ua50g@yahoo.com>, rating total 4.0
      Pattern '*[bcdfghjklmnpqrstvwxyz] [bcdfghjklmnpqrstvwxyz]
[bcdfghjklmnpqrstvwxyz] [bcdfghjklmnpqrstvwxyz]*' from list 'Nonsense
Words' matched body

Rule 'Count Mail Body Nonsense Words' detected spam from " Lakeisha
Gillis" <a8793ua50g@yahoo.com>, rating total 10.0
      Pattern '*[bcdfghjklmnpqrstvwxyz] [bcdfghjklmnpqrstvwxyz]
[bcdfghjklmnpqrstvwxyz] [bcdfghjklmnpqrstvwxyz]*' from list 'Nonsense
Words' matched body

Max spam rating reached, skipping further rules
-----------------
```

Example 4.
Sometimes the Subject: field also includes nonsense text that’s rarely seen on the original mail as the spammer quite deliberately places it where he knows it will be outside the displayed field and invisible to the recipient. The nonsense text is a coded reference that, if the spam is replied to, will allow the spammer’s software automatically to classify and sort any response it receives.

Look at the ‘From’ address. In fact this mail had nothing at all to do with the navy or the military, as the domain name suggests, and instead contained personally unacceptable commercial material. Note there are numerous blank spaces in the subject field (49 is the total reported by SpamWeasel) followed by nonsense characters. So in putting these unseen nonsense characters a long way after the visible subject, the spammer has created another fingerprint and both help SpamWeasel classify the message as spam mail.

```
-----------------
From: "Bradley Townsend" <b.townsend@navy.mil>
Subject: thanks again                                                 i
djywe3kxq
=================
SPAM - rating 10.0
=================
Rule 'Check if message is only HTML text' detected spam from "Bradley
Townsend" <b.townsend@navy.mil>, rating total 2.0

Rule 'Check the Subject: for excessive spaces' detected spam from
"Bradley Townsend" <b.townsend@navy.mil>, rating total 4.0
      No of Spaces = 49

Rule 'Check if Subject contains nonsense words' detected spam from
"Bradley Townsend" <b.townsend@navy.mil>rating total 6.0

Rule 'Check Mail Body for 'Common Spam Message Phrases'' detected
spam from "Bradley Townsend" <b.townsend@navy.mil>, rating total 8.0
      Pattern '*\!*' from list 'Common Spam Message Phrases' matched
body
      Rule 'Check Mail Body for 'Common Spam Message Phrases''
detected spam from "Bradley Townsend" <b.townsend@navy.mil>, rating
total 10.0
      Pattern '*\%*' from list 'Common Spam Message Phrases' matched
body
Max spam rating reached, skipping further rules
-----------------
```

Ex. 5

```
-----------------
From: "Abe Mcneil" <3px3j5ivyy@firemail.de>
Subject: Having trouble finding reasonably priced Life Insurance
=================
SPAM - rating 10.0
=================
Rule 'Check Date: field is present and valid' detected spam from "Abe
Mcneil" <3px3j5ivyy@firemail.de>, rating total 2.0
```

```
Rule 'Check Mail Body for 'Common Spam Message Phrases'' detected
spam from "Abe Mcneil" <3px3j5ivyy@firemail.de>, rating total 4.0
      Pattern '*%*' from list 'Common Spam Message Phrases' matched
body

Rule 'In Mail Body - Count 'Common Spam Message Phrases'' detected
spam from "Abe Mcneil" <3px3j5ivyy@firemail.de>, rating total 10.0
      Pattern '*%*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*\!*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*\?*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*%*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*\!*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*\?*' from list 'Common Spam Message Phrases' matched
body
Max spam rating reached, skipping further rules

-----------------
```

You can see from the above example that use of the expressions ! and ? as well as percentages % and dollar symbols $ can be counted, allowing spam mail to be correctly identified. Spammers try to persuade, cajole, appeal to greed (they can't help themselves) and these kinds of symbols are therefore scattered across many spam mails. If you have a friend that writes to you in this spammy style, no problem: put their address into the 'Friendly Froms' list and it will bypass your spam checker and go straight to your inbox.

By the way, I put a backslash \ in front of these characters in the 'Common Spam Message Phrases' list to tell SpamWeasel that it is to search for the specific text character following. As ? and ! are themselves wildcards it is essential they have this backslash in front of them to avoid confusing SpamWeasel – see the section on wildcards in the SpamWeasel 'Help' files, (Technical Reference section – Using Wildcards).

Ex.6
Below is another good spam sample I particularly like because it is typical of how SpamWeasel catches spammers from different angles. The date field is missing; FREE (in caps) is a promotional device often found in spam; the word casino is a dead giveaway; and a liberal sprinkling of exclamation marks and question marks is typical of the artless persuasion techniques spammers use.

```
-----------------
From: "Rebecca Temple" <tpqpd8tww@netscape.com>
Subject: CC:Low cost cruise a first come first serve
=================
SPAM - rating 10.0
=================
Rule 'Check Date: field is present and valid' detected spam from
"Rebecca Temple" <tpqpd8tww@netscape.com>, rating total 2.0

Rule 'Check Mail Body for 'Common Spam Message Phrases'' detected
spam from "Rebecca Temple" <tpqpd8tww@netscape.com>, rating total 4.0
      Pattern '*[F][R][E][E]*' from list 'Common Spam Message
Phrases' matched body

Rule 'In Mail Body - Count 'Common Spam Message Phrases'' detected
spam from "Rebecca Temple" <tpqpd8tww@netscape.com>, rating total
10.0
      Pattern '*[F][R][E][E]*' from list 'Common Spam Message
Phrases' matched body
      Pattern '*\!*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*\?*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*casino*' from list 'Common Spam Message Phrases'
matched body
      Pattern '*[F][R][E][E]*' from list 'Common Spam Message
Phrases' matched body
      Pattern '*\!*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*\?*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*casino*' from list 'Common Spam Message Phrases'
matched body
Max spam rating reached, skipping further rules
-----------------
```

Ex. 7 This next example indicates what happens to a requested newsletter (and therefore not spam to me) if the From address is not in the ‘Friendly From’ address list. As mentioned earlier, spam is often a subjective matter and the only difference between the previous example, which I regard as pure spam because it was unsolicited, and this newsletter is that I subscribed to the service. Or at least, I did once, so I can probably trust them to unsubscribe me and not worry about having my address sold on if I follow their unsubscribe link.

```
-----------------

From: Forbes Newsletters <investingnewsletters@z2c.net>
Subject: Little Known Nanotech Stock Up 400%
=================
SPAM - rating 10.0
=================
Rule 'Check Mail Body for 'Common Spam Message Phrases'' detected
spam from , rating total 2.0
      Pattern '*$*' from list 'Common Spam Message Phrases' matched
body
Rule 'In Mail Body - Count 'Common Spam Message Phrases'' detected
spam from , rating total 10.0
      Pattern '*$*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*%*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*[F][R][E][E]*' from list 'Common Spam Message
Phrases' matched body
      Pattern '*\!*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*\?*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*subscribe*' from list 'Common Spam Message Phrases'
matched body
      Pattern '*$*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*%*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*[F][R][E][E]*' from list 'Common Spam Message
Phrases' matched body
      Pattern '*\!*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*\?*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*subscribe*' from list 'Common Spam Message Phrases'
matched body
      Pattern '*$*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*%*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*[F][R][E][E]*' from list 'Common Spam Message
Phrases' matched body
      Pattern '*\!*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*\?*' from list 'Common Spam Message Phrases' matched
body
      Pattern '*subscribe*' from list 'Common Spam Message Phrases'
matched body
Max spam rating reached, skipping further rules
----------------
```

The maximum rating total reported is 10 because 10/10 is the top ‘spamminess’ score SpamWeasel can award. Once any rule identifies sufficient spam to take the rating total to ten SpamWeasel stops rating the mail as it doesn’t effect the outcome of any Action settings. After all, once you’ve deleted spam what else can you do? You can’t set fire to it.

In fact, the cleverest thing an ordinary user can do with SpamWeasel is to identify as much spam mail as possible by using the header information alone. You can see what that involves by looking again through the example headers above and ensuring your rules pick up and rate any tell-tale information contained there with appropriate severity.

Then open the Ratings tab and after clicking on the top level (red) Action button untick both these checkboxes: ‘Allow message through’ and ‘Archive message to’ so they become clear and the contents of the fields below them are greyed out.

If SpamWeasel puts spam into this red Action category from the header information alone, SpamWeasel will delete the message at the main server or ISP and never download the spam message. So you neither get the spam in your archives nor do you see it in your inbox and as a bonus, you don’t lose any bandwidth downloading the spam mail itself. You can thus turn the tide against the spammer.

Ex. 8 – this is typical of the trend toward sending a pure graphics file in the form of an html web download request, hidden inside the spam mail. The entire mail comprises a single link to the website from which the full spam message is automatically downloaded. This spam is difficult to filter but easy for the spammer to disseminate, despite the size of the graphics download, as the web download occurs after the mail has arrived in your inbox.
In this instance, the mail size was 621 bytes while the automatically downloaded graphic file was 35763 bytes.

All the spam filter sees is something like this:

```
<html>
<body>
<br>
<center>
<img
src="http://6f8uziozb05zif6jjzi77zaj0za3zg.shoppersville.net/image.as
p?cmpid=hgh-100.gif&dvn=8mC210x671m8JJ199wJ6w_E" width="0"
height="0">
<br>
<a
href="http://6f8uziozb05zif6jjzi77zaj0za3zg.shoppersville.net/ctrack.
asp?cmpid=hgh-100&cvn=d89AF?is|F8d,,F==0,s0n@">
<img src="http://6f8uziozb05zif6jjzi77zaj0za3zg.stop-and-
shop.net/hgh-100.gif" border="0"></a>
<br>
<br>
<a
href="http://6f8uziozb05zif6jjzi77zaj0za3zg.shoppersville.net/remove/
remove.asp">
<img src="http://6f8uziozb05zif6jjzi77zaj0za3zg.stop-and-
shop.net/unsub.gif" border="0"></a>
```

```
</center>
</body>
</html>
```

SpamWeasel saw only the header, which contained this info (see SpamWeasel's remarks attached below the header info):

```
Envelope-to: info@mailgate.com
Delivery-date: Fri, 23 May 2003 00:09:27 +0100
Received: from smtpout-10-83.shoppersville.net ([157.156.170.83]
helo=pfkqrjox)
       by smtp.ky.net with smtp (Exim 4.14)
       id 19IzBd-0004sK-Er
       for info@mailgate.com; Fri, 23 May 2003 00:09:26 +0100
From: Lisa <Jessienv@shoppersville.net>
To: <info@mailgate.com >
Subject: 1Look and feel Younger!
Date: Thu, 22 May 2003 18:05:34 -0400
Mime-Version: 1.0
Content-Type: text/html
Message-Id: <6f8uioB05if6jji77Aj0A3w@shoppersville.net>
X-UIDL: cZ$"!5dR"!%UY!!hP6!!
X-SpamWeasel-Pro: Rule 'Check if message is only HTML text' detected spam
from Lisa <Jessienv@shoppersville.net>, rating total 2.0
```

But the viewer sees a whole page of pictures and written information which has been produced by a link in the mail being operated automatically by the email client after the mail has arrived in the inbox.

One way to stop the link being used by the spammer to make the client software display mail in this format is to disable this function and instead display mail in plain text only. The settings depend upon which client you use so a bit of investigation is needed to achieve this fix.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Tip: turn the tables on the spammer by hitting 'personalised' spam mail e.g. 'Subject: Richard, this is for you' by adding your first name (or your web name) to SpamWeasel's list of 'Common Spam Subject Words'

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Example 9.

```
-----------------
From: "Tropic-Inks"<yinfo-mmjsa@techoffers.us>
Subject: Printer Cartridges - Save up to 90% - Inkjet & Laser Toners
naqerj^bcranpprff(pb(hx
=================
SPAM - rating 10.0
=================
Rule 'Check user name in From: address for nonsense' detected spam
from "Tropic-Inks"<yinfo-mmjsa@techoffers.us> , rating total 2.0
Rule 'Check if a Subject: contains a number ' detected spam from
"Tropic-Inks"<yinfo-mmjsa@techoffers.us> , rating total 4.0
Rule 'Check the Subject: for excessive spaces' detected spam from
"Tropic-Inks"<yinfo-mmjsa@techoffers.us> , rating total 6.0
```

No of Spaces = 12
Rule 'Check if Subject field contains nonsense words' detected spam from "Tropic-Inks"<yinfo-mmjsa@techoffers.us> , rating total 8.0
Rule 'Check if message is only HTML text' detected spam from "Tropic-Inks"<yinfo-mmjsa@techoffers.us> , rating total 10.0
Max spam rating reached, skipping further rules

Example 10.

Finally, a near perfect piece of spam. It is a pure html document with no embellishments.

```
Received: from [ ] by mailgate.com (Mail 5.06.0014/WS0010.00.55e44a8a) with
ESMTP id duembaaa for info@mailgate.com; Sun, 1 Jun 2003 00:16:40 +0100
From: Annika Warkentin <Rachellekaor@mailme.dk>
To: < info@mailgate.com >
Subject: Avoid embarassment
Date: Sun, 01 Jun 2003 07:16:38 -0700
Mime-Version: 1.0
Content-Type: text/html
Message-Id: <wdbjsiblsxuxs@mailme.dk>
X-SpamWeasel-Pro: Rule 'Check if message is only HTML text' detected spam
from Annika Warkentin <Rachellekaor@mailme.dk>, rating total 2.0
```

The html document is a mailer containing only two click-links reading, ‘spotlight yourself’ and ‘don’t call again’ The rest is left to your curiosity (or lack of it). But it’s no surprise that while the first click-link takes you to a web page advertising aphrodisiacs, the second click-link leads to an entirely different and inaccessible site. So you’re unable to request ‘don’t call again’ to the spammer (not that you would be silly enough to let them know you exist save in the interest of scientific enquiry).

The source html program reads as follows.

```
<html>
<body>
<a href="http://www.get888.com/x17">spotlight yourself</a>
<BR>
<BR>
<BR>
<a href="http://www.easy-x10.com/1.html">don't call again</a>
</body>
</html>
```

Example 11.

This next example shows many different rules at work – including the new rule that catches deliberate misspellings such as ‘V1AGRA’ (the digit 1 replacing the letter I) as well as any subject that includes a number.

The new rules that look for exclamation marks and nonsense in the subject field have not been invoked as the mail hit its maximum spam rating first. Had I ordered the Subject field rules differently these rules might have come into play instead. The rule order can be easily changed by dragging and dropping any rule to a different position in the list displayed under the Rules tab.

```
-----------------
From: "Lynda Vernona" < mollymae76@aol.com>
Subject: Half Off V1AGRA!!       [ chvvv ufvbxmhf
=================
SPAM - rating 10.0
=================
Rule 'Check Date: field is present and valid' detected spam from ""
<mollymae76@aol.com>, rating total 2.0
Rule 'Check user name in From: address for nonsense' detected spam
from "" <mollymae76@aol.com>, rating total 4.0
Rule 'Check if user name From: address contains a number' detected
spam from "" <mollymae76@aol.com>, rating total 6.0
Rule 'Check if a Subject: contains a number ' detected spam from ""
<mollymae76@aol.com>, rating total 8.0
Rule 'Check the Subject: for excessive spaces' detected spam from ""
<mollymae76@aol.com>, rating total 10.0
      No of Spaces = 7
Max spam rating reached, skipping further rules
-----------------
```

Some more samples of New Rules

These are lifted from my Rules list. Each one is grouped with the relevant section in the list i.e. new subject field rules were placed amongst those rules already in place concerning the subject by dragging and dropping each completed new rule into its appropriate place.

Rule Name:
Spam if user name is in Subject: field
Comments:

```
// This rule checks the Subject: against the 'User Names' list.
// Placing the user part of the address is typical of unsolicited
// commercial mails.
// Mail is marked as Spam if a match is found.
//
```

Rule content:

```
field$ = HeaderFieldValue("Subject")
if MatchesListItem("User Names",field$) then
   IsSpam()
endif
```

Rule Name:
Check user name in From: address for nonsense
Comments:

```
// This rule checks the email address in the From: field for a
// user name that contains a nonsense user name by looking for
// string of four successive consonants. Spam mail quite
// often originates from mail accounts that are a jumble of letters,
// producing this style of account name. A visible name is displayed,
// usually of a fictitious female. This rule does not inspect the
// fictitious name, only the mailbox identifier.  Mail is marked as
// Spam if a match is found. This rule complements the rule checking
// for a mailbox user name containing a number.
//
```

Rule Content:
```
field$ = ParseAddress(HeaderFieldValue("From"))
if
WildcardMatch(field$,"*[bcdfghjklmnpqrstvwxyz][bcdfghjklmnpqrstvwxyz]
[bcdfghjklmnpqrstvwxyz][bcdfghjklmnpqrstvwxyz]*@*") then
   IsSpam()
endif
```

Rule Name:
Check if user name From: address contains a number
Comments:
```
// This rule checks the email address in the From: field for a
// user name that contains a number. Spam quite
// often originates from mail accounts with the large portals
// which use this style of account name but if you have friends
// using these types of mail account then you may need to disable
// this rule or alternatively, put their name in your Friendly
// From Addresses list. Mail is marked as Spam if a match is found.
//
```

Rule Content:
```
field$ = ParseAddress(HeaderFieldValue("From"))
if WildcardMatch(field$,"*[0-9]*@*") then
   IsSpam()
endif
```

Rule Name:
Check if a Subject: contains a number
Comments:
```
// This rule checks the Subject for any words listed in the
// 'Words containing numbers' list.
// Mail is marked as Spam if a match is found.
//
```
Rule Content:
```
field$ = HeaderFieldValue("Subject")
if WildcardMatch(field$,"*[0-9]*") then
   IsSpam()
endif
```

Rule Name:
Check if Subject: contains an exclamation!
Comments
```
// This rule checks the Subject for a '!' character.
// Note there are two checks made. The first looks for the '!' at
// the end of the subject and the second looks for a '!' followed
// by one space in case the exclamation mark is followed by another
// word.
// Marks the mail as spam if either occurrance is found.
//
```
Rule Content:
```
field$ = HeaderFieldValue("Subject")
if WildcardMatch(field$,"*\!*") then
   IsSpam()

   endif
```

Rule Name:

Check if Subject: contains a question?
Comments:
// This rule checks the Subject for a '?' character anywhere in the
// field, whether it lies within a word or at the end of a word.
// Marks the mail as spam if any '?' is found.
// Note that the ? is also a wildcard character and a backslash \ is
// placed immediately in front of it to snsure it is seen as a
// text character, not a program instruction.
//
Rule Content:
```
field$ = HeaderFieldValue("Subject")
if WildcardMatch(field$,"*\?*") then
   IsSpam()

endif
```

Rule Name:
Check if Subject: field_contains_an_underline_

Comments:
// This rule checks the Subject for an underline '_' character
// anywhere in the Subject field.
// Marks the mail as spam if an underline is found.
//

Rule Content:
```
field$ = HeaderFieldValue("Subject")
if WildcardMatch(field$,"*_*") then
   IsSpam()
endif
```
Rule Name:
Check if Subject field contains nonsense words

Comments:
// This rule checks the subject field for nonsense words.
// It detects a string containing four or more
// successive consonants.
// Mail is marked as Spam if a match is found.
//

Rule Content:
```
field$ = HeaderFieldValue("Subject")
if WildcardMatch(field$,"*[b-df-hj-np-tv-z][b-df-hj-np-tv-z][b-df-hj-
np-tv-z][b-df-hj-np-tv-z]*") then
   IsSpam()
endif
```

Rule Name:
Check if Mail Body contains nonsense words

Comments:
// This rule checks the mail body for nonsense words.
// It detects a string containing four or more
// successive consonants.
// Mail is marked as Spam if a match is found.
//

Rule Content:
```
seg% = 1
:loop
```

```
   if WildcardMatchBody("*[b-df-hj-np-tv-z][b-df-hj-np-tv-z][b-df-hj-
np-tv-z][b-df-hj-np-tv-z]*") then
      IsSpam()
        goto exit:
   endif
   seg% = seg% + 1
   if not SelectMsgSegment(seg%) then goto exit:
   goto loop:

:exit
```

Rule Name
In Mail Body count nonsense Words

Comments:
```
// This rule is an extension of "Check Mail Body for nonsense words
// and counts each example found.
// It is also an example of adjusting the rating according to the
// number of matches found.
// Mail is marked as Spam if any single match is found.
//
```

Rule Content:
```
RuleRating% = GetRuleRating()
NoHits% = 0

seg% = 1
:loop
  ThisHits% = 0
  ThisHits% = CountBodyPatternListHits("Nonsense Words")
  NoHits% = NoHits% + ThisHits%
  seg% = seg% + 1
  if not SelectMsgSegment(seg%) then goto endcheck:
  goto loop:

:endcheck

if NoHits% > 0 then
   SpamRating% = NoHits% * RuleRating%
   SetRuleRating(SpamRating%)
   IsSpam()
endif
```

Rule Name:
Check Mail body for words containing a number

Comments:
```
// This rule checks the mail body for nonsense words commonly used
// to deceive spam word lists by substituting a number for a letter.
// It detects a word beginning with an alphabetical character that
// has a number anywhere in it e.g. v1agra.
// Mail is marked as Spam if a match is found.
//
```

Rule Content:
```
seg% = 1
:loop
   if WildcardMatchBody("*[a-zA-Z][0-9]*") then
      IsSpam()
        goto exit:
```

```
   endif
   seg% = seg% + 1
   if not SelectMsgSegment(seg%) then goto exit:
   goto loop:

:exit
```

Rule Name:
In Mail body count words containing a number

Comments:
```
// This rule checks the mail body for nonsense words that
// deliberately substitute a number for a letter, a trick commonly
// used by spammers to thwart the use of spam word lists.
// The rule detects a word beginning with an alphabetical character
// that has a number anywhere in it e.g. v1agra.
// Mail is marked as Spam if a match is found and as this trick is a
// spammer's hallmark it can be given a high rating. Unless that is
// you expect mail using alphanumeric labelling or data.
// The pattern this rule looks for is found in the list
// "Word containing a number" and the pattern I've put in that list
// is *[a-zA-Z][0-9]*  which is interpreted to mean:
// * = a wildcard meaning 'anything' in front of the first character.
// [a-zA-Z] = shorthand for any alphabetical character in either
// lower case or upper case.
// [0-9] = shorthand for any digit between 0 and 9.
// * = anything after the last number detected (which could be an
// alphanumeric character or blanks).
```

Rule Content:
```
RuleRating% = GetRuleRating()
NoHits% = 0

seg% = 1
:loop
  ThisHits% = 0
  ThisHits% = CountBodyPatternListHits("Word containing a number")
  NoHits% = NoHits% + ThisHits%
  seg% = seg% + 1
  if not SelectMsgSegment(seg%) then goto endcheck:
  goto loop:

:endcheck

if NoHits% > 0 then
   SpamRating% = NoHits% * RuleRating%
   SetRuleRating(SpamRating%)
   IsSpam()
endif
```

Further reading

For further information on the subject try going to:

1. http://members.aol.com/emailfaq/

This is not for AOL members only, but is a public site for everyone, with lots of useful definitions and handy tips for beginners. Good FAQ’s on email abuse and a full description of how to mung the spammer (it’s the art of hiding your address from spammers but not from friends). Its main weakness is it appears to be a static site, so for smack up to date commentary, try the following sites:

2. http://www.samspade.org

SamSpade has spam FAQ’s, freeware spam tracing tools and many other useful utilities (not necessarily for beginners though), useful links to other anti-spam sites and much news and comment.

3. http://www.cauce.org

A voluntary US pressure group that voices the concerns of netizens to political and regulatory authorities on unsolicited commercial email. The site produces bulletins about ongoing official actions aimed at preventing spammers abusing the Internet.

Section 3.

Initial Configuration

Some Helpful information:

1. After making ANY changes to the SpamWeasel configuration – e.g. Rules ratings, settings, word & patterns lists – you need to stop and restart SpamWeasel for the new settings to take effect. If you don't do this, any new setting will only take effect the next time your computer is restarted.

2. All additional technical information you may need to use SpamWeasel is contained in the Help documentation, found by clicking on the Help item on the main menu. To find Help following initial installation, right click (control) on the SpamWeasel Pro icon in your system tray.

Configuring SpamWeasel Pro a) Using the Wizard and b) Manually

Note: It's best to set the maximum time ('timeout') SpamWeasel allows for each mail collection to less than the maximum collection time set for your email client. If you don't, collection looping may occur and fill your mailbox with duplicate mail. To guard against this, ***set SpamWeasel's maximum email collection timeout to 10 seconds less than the collection timeout in your email client****.*

Note: Setting SpamWeasel's maximum email collection time to 0 disables timeout.

a) Using the Wizard

If SpamWeasel Pro has not been configured, when you start the software you will be asked if you wish to use the Configuration Wizard. Choose this to guide you through the initial set-up of both SpamWeasel and your selected e-mail Client.

b) Manually

To manually configure SpamWeasel Pro:

1. If SpamWeasel is not running, double click on the Desktop icon or use your Start menu to start it.

2. Double click on the Desktop or System tray icon to open the Configuration Screens.

3. Click on the Settings Tab and review the settings for the Spam Report and Archive period. If you have other POP mail software which acts as a server/proxy (e.g. some Antivirus programs), you also may need to adjust the Listen on port (See Understanding Ports in the Help documentation for more detail.)

4. Click on the Rules Tab and review which rules you wish to be active, disabling any that are not required. You may also adjust the ratings for your chosen rules. Note - the default rule set is generally good for most basic configurations. For ideas on this, see the discussion below on Using Rules and Ratings

5. Click on the Patterns Tab and review any lists that should contain localised entries, for example - Friendly e-mail Addresses. For more ideas on this see the discussion below on Using SpamWeasel Pro.

6. Review any other settings as you wish, for example Anti-Virus scanning, then click OK to save your settings. Go to the Help documentation for more info on this subject.

7. Right click on the System Tray icon and select Exit to close the program, then re-start SpamWeasel Pro as above.

You should now configure your e-mail Client to work with SpamWeasel if you have not already done this. SpamWeasel Pro will remain inactive until you complete this step.

Configuring your E-mail Client.

In order for your e-mail client to work with SpamWeasel Pro, you need to change the settings for any POP collections (Incoming mail) you wish to be passed through the filter. Do not make any changes to your SMTP (Outgoing mail) settings.

To configure your client to work through SpamWeasel Pro you normally only need to make two changes:

1. POP3 Server - Note your current setting then change this to the localhost IP address setting of 127.0.0.1

2. POP3 Account Name - Change this by adding an @ and then your ISP's POP server address to the end of the current setting. You can also specify a non-standard port by further adding :

portnumber to the end of this setting.

Example

Your Current Settings were:

Incoming POP3 Server - pop.myisp.net

Account Name - jdoe

Password - pass

Change to:

Incoming POP3 Server - 127.0.0.1

Account Name - jdoe@pop.myisp.net

Password - pass

Notes to Configuring your Email Client.

i. Password - Do not change this setting.

ii. You may need to re-start your e-mail client for the changes to take effect.

iii. If you have multiple accounts configured in your e-mail client, you will need to repeat these changes for each account you wish to be passed through SpamWeasel.

iv. If ever you collect quite large numbers of emails when you connect to your ISP you may need to increase the timeout setting in your client software.

v. If your ISP requires Secure Password Authentication for incoming (POP3) mail, you must disable this and add authentication control to the account name in step 2. above. There are currently two options $APOP if your ISP uses APOP authentication and $MSN if you are using an MSN account which requires authentication. For example, for MSN your account name setting may look like jdoe@pop3.email.msn.com$MSN

vi. Some e-mail clients (some Netscape versions in particular) do not allow using '@' as the separator between the POP Account name and the ISP's server address. You can configure SpamWeasel to use a different character (we recommend '?') in place of the @ by changing the Separator Char setting on the Settings Tab.

vii. For advanced users - you may use any name in the POP3 server address, for example localhost, provided your TCP/IP configuration will correctly resolve this to 127.0.0.1

Different e-mail clients can use different names for the above settings. If you are not sure how to access these settings, please refer to your client software help or manual. You can also visit the SpamWeasel FAQ area on our web site www.mailgate.com where you will find details for configuring the more common e-mail packages and information on any special settings that may be required.

Written by R. A. Hall