# COMPARATIVE REVIEW

## ANTI-SPAM COMPARATIVE REVIEW JULY 2009

*Martijn Grooten*

*VB*’s first anti-spam comparative review and certification showed some interesting results (see *VB*, May 2009, p.S5), and the winners of VBSpam awards deserve full credit for doing so well. This month, the all-important question is: can they repeat their outstanding performance?

Prior to the first anti-spam comparative review we ran a trial, during which the licence for one of the participating products expired. This product was configured to continue to work, yet the anti-spam engine was no longer being updated – indeed, when we looked at its performance, the spam catch rate gradually decreased over time. Apart from acting as a reminder of how important it is to renew licences, this served as a demonstration that spam changes over time and that anti-spam vendors need to ensure their engines are up to date and work against the latest spam threats. This is why we run a new anti-spam test every two months.

Nine products participated in this month’s test, seven of which were commercial products, while the other two were free and open source. Of the seven commercial products, two were hosted solutions, two were hardware appliances, one ran on a *Windows Server* machine and the final two ran on *Linux*. Together they provide a good representation of the range of different options available when it comes to spam filtering within an organization.

### THE TEST SET-UP

The set-up of this test was more or less the same as for the last test. The methodology is recorded at http://www.virusbtn.com/vbspam/methodology.

Some changes to the MTAs enabled us to run a test with a much larger email corpus and more products running in parallel; however, these changes do not affect the test set-up itself. As in the previous test, the Message ID was used to uniquely identify emails, and if such a header was not already present, the gateway MTA added one. Email that did not reach the back-end server within an hour was assumed to have been classified as spam. During this hour, up to five redelivery attempts were made for emails that had caused an error during the SMTP transaction.

Unlike in the previous test, the two products that ran on the gateway MTA, *SpamAssassin* and *ClamAV*, were sent emails in real time.

The products that needed to be installed on a server were installed on a *Dell PowerEdge R200*, with a 3.0GHz dual core processor and 4GB of RAM.

## THE EMAIL CORPUS

As before, all of the emails sent to valid addresses on the virusbtn.com domain were sent through all of the products, in real time. This corpus consisted of 2,393 ham messages and 26,755 spam messages. The 'golden standard' for each email was decided upon by the recipient, with the exception of emails for which all products agreed: in these cases we assumed the products were correct. Emails that were reported as false positives were checked a second time, to make sure none had been misclassified by the recipient as ham.

To increase both the volume and the variety of spam seen by the products, we have been working closely with *Project Honey Pot*. The *Project Honey Pot* team manages a large distributed network of spam traps and thus receives not only a large amount of spam, but also spam that reflects the global variation in bogus email. For this month's test they sent us part of their feed; the emails were assigned to a random valid address on the virusbtn.com domain and then relayed to the products in real time.

So that the products would not have any information about which emails were part of this feed (all of which, of course, were assumed to be spam) the Received-headers were rewritten, so that it appeared as if the email had been received by our MTA. Moreover, many spam emails are 'personalized' and thus contained the local-part and/or the domain of the original spam trap; these were replaced by the local-part of the newly assigned recipient and virusbtn.com respectively.

The *Project Honey Pot* feed provided us with 716,256 additional spam emails, which meant an overall corpus of 745,404 emails, 2,393 of which were ham. As the test ran for a period of almost three weeks (starting at 16:45h on 5 June and finishing at 08:00h on 26 June) this meant that the products saw about 25 emails per minute or almost one email every two seconds.

It was interesting to see that all products performed better against the *Project Honey Pot* spam than against the *VB* spam; in most cases the difference in performance was rather significant. We can only guess the reason for this, but it could well be caused by the nature of the spam *VB* receives: like most companies, we receive a large amount of commercial email, unsolicited and sent in bulk, yet somehow targeted. Emails like these don't generally end up in spam traps and do not look like 'typical' spam, yet they are illegal and filters should be expected to block them.

## FALSE POSITIVES

Anyone comparing the false positives reported in our test with those reported by other anti-spam tests or by the vendors themselves, will notice that our numbers are significantly higher. There are several reasons for this. Firstly, the emails received by *VB* employees – many of which discuss malware and spam – are particularly hard to filter. For example, we may see a legitimate message in which a spam domain is being discussed, and this message is classified as spam by one or more products on the premise that it contains this very domain. While the mistake is understandable, these messages are ham and an ideal spam filter would not make this mistake.

Secondly, most products give the end-user and/or the system administrator the option to whitelist certain senders. While we encourage the use of such techniques in practice, we have not applied them in our tests: it would be hard, if not impossible, to perform whitelisting in the same way as an average end-user, plus it would put any products entering the test for the first time at a disadvantage.

This is one of the main reasons why our certification scheme emphasizes the relative performance of products compared to those of their competitors, rather than focusing on actual numbers.

In the results reported below, 'SC rate (total)' represents the overall spam catch rate over the entire corpus of 745,404 emails. 'SC rate (Project Honey Pot corpus)' represents the spam catch rate achieved within the Project Honey Pot corpus alone, and 'SC rate (VB spam corpus)' represents the spam catch rate achieved within the VB corpus alone. 'FP rate' represents the number of false positives as a proportion of the total number of ham messages, while the 'FP rate of total VB mail corpus' is the number of false positives as a proportion of the total number of messages contained in the VB mail corpus.

## BitDefender Security for Mail Servers 3.0.2

**SC rate (total):** 98.23%

**SC rate (Project Honey Pot corpus):** 98.56%

**SC rate (VB spam corpus):** 89.36%

**FP rate:** 2.55%

**FP rate of total VB mail corpus:** 0.209%

Romanian company *BitDefender* won a gold VBSpam award for its *Linux* server product in the May 2009 test and its developers were keen to see if the product could repeat its performance. A fresh installation of the product, again as an extension to *Postfix*, but this time installed on a server running *SUSE11*, was set up easily.

vb July 2009
silver
SPAM
virusbtn.com

The spam catch rate of more than 98% is a huge improvement compared to the previous test and even the product's performance on the *VB* spam corpus increased by more than five per cent. Unfortunately, the false positive rate also increased to a level that just pushed the product outside the limits for a gold award; instead it wins a VBSpam Silver award this month.

## ClamAV using Sanesecurity signatures

**SC rate (total):** 73.97%

**SC rate (Project Honey Pot corpus):** 75.04%

**SC rate (VB spam corpus):** 45.51%

**FP rate:** 0.38%

**FP rate of total VB mail corpus:** 0.031%

The signatures provided by *Sanesecurity* that can be plugged into the open source *ClamAV* anti-virus product were never meant to match the performance of dedicated anti-spam solutions. Rather, they are intended to work together with another solution to provide a layered spam filter. Still, the 23% spam catch rate measured in May was disappointing for the developer despite a zero false positive rate.

The developer will be happy to hear that this month the product's performance almost tripled to 74%, and even on the *VB* spam corpus it increased by 18% – and although these rates are insufficient to earn the product a VBSpam award, they indicate that the product is well up to its task as the first step in a multi-layered spam filter. There were a handful of false positives this time, all of which were caused by legitimate emails mentioning a malicious domain or email address – and it would be fair to say that it would be rare for the majority of end-users to receive such emails.

## FortiMail

**SC rate (total)**: 99.11%

**SC rate (Project Honey Pot corpus):** 99.28%

**SC rate (VB spam corpus):** 94.38%

**FP rate:** 2.63%

**FP rate of total VB mail corpus:** 0.216%

*Fortinet*, based in Vancouver B.C., is a regular participant in the VB100 anti-malware testing, so it came as little surprise that the company was eager to submit its *FortiMail* appliance for the anti-spam test. As with most hardware appliances, no software needed to be installed and after a short set-up process, the product was up and running.

vb July 2009 silver SPAM virusbtn.com

Further configuration, to fine tune the appliance, can be carried out via a web interface, while another web interface can be used by end-users to view their quarantined email or modify per-user settings. The web interface isn't used purely to click buttons however: one of the most important tasks of a system administrator running an anti-spam solution is to find out why certain emails were blocked and to prevent this from happening again. Those using *Fortinet* will have an easy task doing so, thanks to an extensive logging system: for every email received the logging system records which anti-spam tests were passed or failed.

With a stunning spam catch rate of 99.1%, *FortiMail* outperformed all of its competitors in this respect, and even on the *VB* corpus well over 94% of the spam was identified correctly. On the downside, the product's false positive rate was slightly higher than average and thus *FortiMail* debuts by winning a VBSpam Silver award.

## Kaspersky Anti-Spam 3.0

**SC rate (total):** 97.54%

**SC rate (Project Honey Pot corpus):** 98.17%

**SC rate (VB spam corpus):** 80.81%

**FP rate:** 0.04%

**FP rate of total VB mail corpus:** 0.003%

*Kaspersky* is another VB100 regular that has joined the anti-spam test this month. The Russian anti-malware giant has been active in the anti-spam field for a long time and the product we tested is the third generation of *Kaspersky Anti-Spam* (*KAS*).

vb July 2009 platinum SPAM virusbtn.com

A *Linux* product, we ran it on a *SUSE11* server as an extension to the *Postfix* MTA. Installation was smooth and painless and the product was running just a few minutes after the download had finished. After that, its performance gave so few reasons to worry that not until I started writing this review did I have a reason to search for log files, upon which I happily discovered that the decision made for each email is indeed stored.

Administrators running *KAS* will have little reason to look in these log files for false positives though: out of almost 2,400 ham messages sent to the product, only one was marked as spam. This unbelievably low false positive rate combined with a spam catch rate of over 97%, well above average, means that *Kaspersky Anti-Spam* is the deserving recipient of a VBSpam Platinum award.

## MessageStream

**SC rate (total):** 98.82%

**SC rate (Project Honey Pot corpus):** 99.21%

**SC rate (VB spam corpus):** 88.48%

**FP rate:** 1.59%

**FP rate of total VB mail corpus:** 0.130%

You would be forgiven for thinking that vendors submit their products for the anti-spam test purely for marketing reasons, but many developers are also keen to hear our feedback so that they can see how and in which areas their products can be improved. UK-based *Giacom*, which develops the hosted solution *MessageStream*, achieved a VBSpam Gold award in May, yet felt that its false positive rate could be reduced, resulting in its developers making some changes to its filtering for all of its customers.

These changes were rolled out halfway through this month's test, but the number of false positives was already reduced to about 1.5% – well below average. This, together with a spam catch rate of almost 99%, means that *MessageStream* has achieved its second VBSpam Gold award in a row.

## ModusGate

**SC rate (total):** 95.41%

**SC rate (Project Honey Pot corpus):** 95.63%

**SC rate (VB spam corpus):** 89.48%

**FP rate:** 6.64%

**FP rate of total VB mail corpus:** 0.545%

As someone with more than a decade of experience with various flavours of Unix and *Linux* I sometimes think that *Windows* software works against my intuition. This prejudice, however, was quickly proven wrong when I installed *ModusGate*, an anti-spam solution produced by Canadian company *Vircom*. The product is available as a hardware appliance, but tested here as a software solution installed on a *Windows Server 2003* machine. Set-up and installation were straightforward and the graphical interface was clear and easy to work with.

Unfortunately, despite a decent spam catch rate, the product's false positive performance was disappointing, with a score of more than 6%. The majority of the misclassifications were emails that had been sent as mass-mailings. Regrettably, the product's high false positive rate was enough to deny it a VBSpam award in this month's test.

## M+Guardian (Messaging Architects)

**SC rate (total):** 98.92%

**SC rate (Project Honey Pot corpus):** 99.12%

**SC rate (VB spam corpus):** 93.63%

**FP rate:** 0.79%

**FP rate of total VB mail corpus:** 0.065%

*Messaging Architects* was one of the first companies to submit its solution, *M+Guardian,* for the anti-spam test and the confidence its developers have shown in their product proved to be justified: the product was the sole winner of a VBSpam Platinum award in the first test. It was with interest that we assessed the results this month to see whether the product would be able to continue its excellent performance in the second test, with more spam, more products and stricter benchmarks.

Happily for *Messaging Architects* it did continue its excellent performance. A spam catch rate of almost 99% and a less than 1% false positive rate earn *M+Guardian* its second VBSpam Platinum award.

## SpamAssassin 3.2.5

**SC rate (total):** 64.26%

**SC rate (Project Honey Pot corpus):** 64.72%

| | | | | | Total spam | | | Project Honey Pot corpus | | | VB corpus | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | True negative | FP | FP rate | FP / total VB corpus | False negative | True positive | SC rate | False negative | True positive | SC rate | False negative | True positive | SC rate |
| BitDefender | 2332 | 61 | 2.55% | 0.209% | 13133 | 729878 | 98.23% | 10285 | 705971 | 98.56% | 2848 | 23907 | 89.36% |
| ClamAV | 2384 | 9 | 0.38% | 0.031% | 193391 | 549620 | 73.97% | 178813 | 537443 | 75.04% | 14578 | 12177 | 45.51% |
| FortiMail | 2330 | 63 | 2.63% | 0.216% | 6640 | 736371 | 99.11% | 5136 | 711120 | 99.28% | 1504 | 25251 | 94.38% |
| Kaspersky | 2392 | 1 | 0.04% | 0.003% | 18268 | 724743 | 97.54% | 13134 | 703122 | 98.17% | 5134 | 21621 | 80.81% |
| MessageStream | 2355 | 38 | 1.59% | 0.130% | 8752 | 734259 | 98.82% | 5670 | 710586 | 99.21% | 3082 | 23673 | 88.48% |
| ModusGate | 2234 | 159 | 6.64% | 0.545% | 34128 | 708883 | 95.41% | 31313 | 684943 | 95.63% | 2815 | 23940 | 89.48% |
| M+Guardian | 2374 | 19 | 0.79% | 0.065% | 8006 | 735005 | 98.92% | 6302 | 709954 | 99.12% | 1704 | 25051 | 93.63% |
| SpamAssassin | 2324 | 69 | 2.88% | 0.237% | 265516 | 477495 | 64.26% | 252664 | 463592 | 64.72% | 12852 | 13903 | 51.96% |
| Webroot | 2335 | 58 | 2.42% | 0.199% | 25659 | 717352 | 96.55% | 24310 | 691946 | 96.61% | 1349 | 25406 | 94.96% |

**SC rate (VB spam corpus):** 51.96%
**FP rate:** 2.88%
**FP rate of total VB mail corpus:** 0.237%

The ancient, but still actively developed, open source *SpamAssassin* product took part in the first anti-spam test, but failed to match the performance of most of its commercial competitors. It was suggested that this was because an old version of the product had been installed. With the latest version, 3.2.5, running on a fresh *SUSE11 Linux* server, we were keen to see if this would have a positive effect on its performance.

Unfortunately, the spam catch rate was still lower than 65%, while the false positive rate actually rose to a higher level than previously. While the *sa-update* command, which is run every hour, suggests that nothing is the matter, it seems likely that the poor results are caused by an incorrect configuration, one that can hopefully be fixed before the next test.

## Webroot E-Mail Security SaaS

**SC rate (total):** 96.55%
**SC rate (Project Honey Pot corpus):** 96.61%
**SC rate (VB spam corpus):** 94.96%
**FP rate:** 2.42%
**FP rate of total VB mail corpus:** 0.199%

*Webroot*'s hosted solution failed to win an award in the May test because its number of false positives was a lot higher than the benchmark – a problem caused by an incorrectly configured SPF test. The problem was easily fixed, however, and indeed, the false positive rate of 2.42% measured in this month's test is only slightly above the average.

vb July 2009 silver SPAM virusbtn.com

The product also achieved a very decent 96.5% spam catch rate, and deserves extra credit for having the highest spam catch rate on the *VB* spam corpus. A VBSpam Silver award is thus well deserved and only a slight improvement of the false positive rate would enable it to do even better next time.

## AWARDS

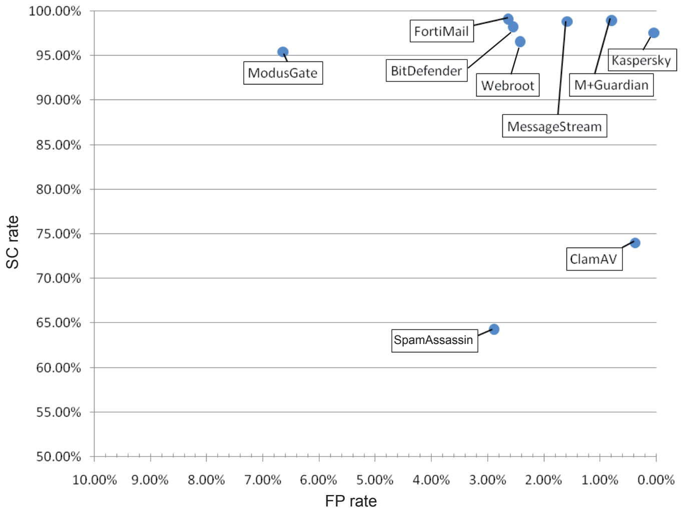
As in the previous test, the level of the awards earned by products are defined as follows:

- VBSpam Platinum for products with a total spam catch rate twice as high and a false positive rate twice as low as the average in the test.
- VBSpam Gold for products with a total spam catch rate at least as high and a false positive rate at least as low as the average in the test.
- VBSpam Silver for products whose total spam catch rate and false positive rates are no more than 50% worse than the average in the test.

To avoid the averages being skewed by one or more malperforming products, any product with a false positive rate of more than 10% and/or a spam catch rate of less than 70% is removed from the computation of the averages. In this case, the *SpamAssassin* scores were removed, because its spam catch rate was well below 70%. This month's benchmarks were then as follows:

- Platinum: SC 97.41%; FP 1.07%
- Gold: SC 94.82%; FP 2.13%
- Silver: SC 92.23%; FP 3.20%

The table opposite shows the scores for all of the products on test. The highlighted columns show the scores used for the benchmark calculations.



## CONCLUSIONS

It was a relief to see the bugs that had caused some stress during the first test had been solved. It is also good to see several of the products showing credible results and thus to get a better picture of which products really are the high performers in this field. Still, the question remains as to whether these products can continue their performance in the next test, and it will be interesting to see the effect of the improvements and tweaks that will undoubtedly be made to other products.

The next anti-spam comparative review will take place in August, with the results published in September 2009. The deadline for product submission will be 27 July. Any developers interested in submitting a product are encouraged to get in touch by emailing martijn.grooten@virusbtn.com.