COMPARATIVE REVIEW

VBSPAM COMPARATIVE REVIEW

Martijn Grooten

Of the many reviews of the 'noughties' we have seen in the media in recent weeks, few have mentioned spam as being something that defined the decade. Yet in the past ten years, spam has grown from a mere nuisance to Internet users into a major field of criminal activity.

Even the most optimistic will find little reason to believe that the spam problem will disappear any time soon, but thankfully those in the anti-spam world keep working hard to protect end-users' inboxes.

The first VBSpam comparative review of the new decade saw 15 products on the test bench: 14 full anti-spam products and one partial solution. Developers of three of the products that took part in previous tests decided to sit this one out in order to concentrate on new versions of their products; all of them hope to be back on board for the next test. However, four new products were included in this month's test.

THE TEST SET-UP

No major modifications were made to the test set-up and, as usual, the full methodology can be found at http://www.virusbtn.com/vbspam/methodology/.

As before, 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 *Linux* products ran on *SuSE Linux Enterprise Server 11*; the *Windows Server* products ran on either the 2003 or the 2008 version, depending on which was recommended by the vendor.

Some changes have, however, been made to the award criteria. First, we decided to stop using the combined average performance of the products to define the award thresholds – with the performance levels of all products continually improving, the thresholds were in danger of becoming too dependent on one or two products performing significantly more poorly than the rest. Secondly, with the thresholds for the three award levels edging ever closer to each other, the difference between the levels was becoming very small and almost more dependent on luck than on a significant difference in performance.

As a result, a product's performance will now be measured based on the value of its spam catch (SC) rate minus three times its false positive (FP) rate. A product will earn a VBSpam award if this value (referred to as the 'final score') is at least 96%:

$$\text{SC} - (3 \text{ x FP}) \geq 96\%$$

The simplification of the award structure should help to reduce confusion among end-users.

This does not mean we believe there is no difference in performance between the various products, and end-users are encouraged to compare the performance figures shown in the tables and to look at the relative positions of the products plotted in the VBSpam results graph.

Our intention is not to give an absolute value to the performance measured by us: a 98% catch rate in our test does not necessarily indicate the same as a 98% catch rate in another test, and does not mean that the product will catch 98% of a customer's spam. However, the catch rates (or false positive rates) of two products in our test can be compared against each other.

THE EMAIL CORPUS

The test ran from 1pm GMT on 14 December 2009 until 8am GMT on 4 January 2010 – a test period of almost three weeks, which included most of the Christmas holiday period (notorious for breaking spam records). The corpus contained 249,569 emails: 2,811 ham messages and 246,758 spam messages, where the latter consisted of 224,411 messages provided by Project Honey Pot and 22,347 messages sent to legitimate @virusbtn.com addresses.

As described in the previous VBSpam review (see *VB*, November 2009, p.22), the ham consisted of all legitimate messages sent to @virusbtn.com addresses, but with the senders of emails that regularly discuss spam- and malware-related topics (for example anti-spam discussion lists) excluded. Such emails regularly contain links to malicious and/or spamvertised URLs and we believe that not only are such emails unlikely to occur in the legitimate email stream of an average organization, but also that the recipients of such emails generally have the level of knowledge and technical ability required to whitelist these particular senders. To make up for these exclusions, we added to the corpus a number of email discussion lists on a variety of other topics.

In an attempt to make the test results more realistic, we decided to count no more than four false positives per sender for each product. This change should prevent a small mistake on a blacklist from having escalating effects if a certain sender sends many emails during a test period, but more importantly, it will reflect a real situation where legitimate senders whose emails keep being blocked are eventually whitelisted.

Another small change was that emails that claimed to have been sent from @virusbtn.com addresses were removed from the corpus: given the way our test is set up, products could have valid reasons for considering these emails to have been sent from a legitimate *VB* server. While this does not appear to have affected any product's past performance, we want to avoid the possibility of penalizing filters for making such assumptions.

A more showing change was the addition of two new categories: those of 'image spam' and 'large spam'. The former consisted of all spam emails that contained at least one embedded image, and the latter consisted of all spam emails with a body size of at least 50,000 bytes. Both types of emails are considered difficult to filter, especially using content scanning methods. We measured each product's performance on these sub-sets of the spam corpus, and while these measurements do not count towards the VBSpam award, they should give developers a better idea as to which part(s) of their filters can be improved upon.

RESULTS

Starting from this test we will distinguish between full solutions and partial solutions. The latter are anti-spam products that are unlikely to be deployed on their own but are intended to work together with other solutions. As such, the performance of these products should not be compared directly to other solutions. This test contained one such solution (*Spamhaus Zen*), but *SaneSecurity*, which participated in the previous tests, would also fall into this category.

BitDefender Security for Mail Servers 3.0.2

SC rate (total): 98.14%

SC rate (Project Honey Pot corpus): 98.86%

SC rate (VB spam corpus): 90.94%

SC rate (image spam): 97.53%

SC rate (large spam): 94.84%

FP rate: 0.605%

Final score: 96.33%

Having worked hard on their spam filter since the last test, *BitDefender*'s developers were eager to see the results of this month's test. Their hard work paid off: both the spam catch rate and the false positive rate improved a little, and in an area where the devil is in the details, this is no small achievement. *BitDefender*'s *Linux* server product thus wins its fifth VBSpam award in a row.

vb VERIFIED SPAM virusbtn.com

Fortinet FortiMail

SC rate (total): 98.40%

SC rate (Project Honey Pot corpus): 98.79%

SC rate (VB spam corpus): 94.57%

SC rate (image spam): 97.83%

SC rate (large spam): 94.91%

FP rate: 0.427%

Final score: 97.12%

One of the clear high achievers of the previous VBSpam test, *Fortinet*'s *FortiMail* appliance saw its performance levels drop slightly on both fronts. However, this was not enough to prevent the product from earning a VBSpam award – the company's fourth in a row – and its developers will no doubt be extra motivated to improve its score during the next round of testing.

vb VERIFIED SPAM virusbtn.com

Kaspersky Anti-Spam 3.0

SC rate (total): 95.94%

SC rate (Project Honey Pot corpus): 97.15%

SC rate (VB spam corpus): 83.71%

SC rate (image spam): 97.54%

SC rate (large spam): 93.67%

FP rate: 0.071%

Final score: 95.73%

It is hard not to feel that the anti-spam developers at *Kaspersky* are a bit unlucky: while their *Linux* server product was the only one to miss out on a VBSpam award in this test, it had fewer false positives than any other full solution. An improved spam catch rate should see the product winning an award again next time around.

M86 MailMarshal SMTP

SC rate (total): 99.60%

SC rate (Project Honey Pot corpus): 99.86%

SC rate (VB spam corpus): 97.01%

SC rate (image spam): 99.60%

SC rate (large spam): 98.25%

FP rate: 0.142%

Final score: 99.17%

M86's *MailMarshal SMTP* spam filter, which runs on *Windows Server 2003*, made its debut in the VBSpam test in November with commendable results, but did even better in this test: it saw its false positive

vb VERIFIED SPAM virusbtn.com

rate reduced significantly, while barely compromising on the spam catch rate, and it was the only product in this test with a final score of more than 99%. Moreover, neither large spam emails nor those containing images proved a problem for the product.

McAfee Email Gateway (formerly IronMail)

SC rate (total): 99.59%
SC rate (Project Honey Pot corpus): 99.84%
SC rate (VB spam corpus): 97.11%
SC rate (image spam): 99.46%
SC rate (large spam): 97.95%
FP rate: 0.640%
Final score: 97.67%

For the third time in a row, the *McAfee Email Gateway* hardware appliance caught more than 99% of all spam and its performance in the various categories shows that this product is a good all-round filter. The product's false positive rate is slightly on the high side, but certainly not too high for it to win another VBSpam award.

McAfee Email and Web Security Appliance

SC rate (total): 98.92%
SC rate (Project Honey Pot corpus): 99.49%
SC rate (VB spam corpus): 93.23%
SC rate (image spam): 98.84%
SC rate (large spam): 94.86%
FP rate: 0.462%
Final score: 97.53%

Another of the high achievers of the previous two tests, *McAfee*'s *Email and Web Security Appliance* demonstrated a very good spam catch rate once again – a small improvement compared to the previous test even – but also saw its false positive rate increase. While certainly not a bad performance, the developers will no doubt be eager to show that the rise in false positives was a one-off incident.

MessageStream

SC rate (total): 99.14%
SC rate (Project Honey Pot corpus): 99.61%
SC rate (VB spam corpus): 94.38%
SC rate (image spam): 99.15%
SC rate (large spam): 97.55%
FP rate: 0.605%
Final score: 97.33%

The *MessageStream* hosted solution is another product whose performance dropped slightly compared to the previous test (in particular, it missed more legitimate emails than during previous tests), but this didn't stop it from performing well enough to earn a fifth VBSpam award in a row.

Microsoft Forefront Protection 2010 for Exchange Server

SC rate (total): 99.06%
SC rate (Project Honey Pot corpus): 99.32%
SC rate (VB spam corpus): 96.49%
SC rate (image spam): 99.24%
SC rate (large spam): 98.07%
FP rate: 0.249%
Final score: 98.31%

The publication of the previous VBSpam test report almost coincided with the official release of *Microsoft*'s *Forefront Protection 2010 for Exchange Server* but the developers certainly weren't too busy to make improvements to their product. This test saw improvements in both the spam catch rate and the false positive rate, and with a final score of over 98%, *Forefront* was among the top performers in this test.

MXTools Reputation Suite

SC rate (total): 97.65%
SC rate (Project Honey Pot corpus): 98.81%
SC rate (VB spam corpus): 85.97%
SC rate (image spam): 98.28%
SC rate (large spam): 94.86%
FP rate: 0.178%
Final score: 97.12%

MXTools sells three anti-spam solutions, each of which can be used as an add-on to improve an existing solution, but the three can also be used together to form a standalone spam filter. Apart from *Spamhaus ZEN plus DBL*, which is described below, the suite also contained *SURBL* and *Server Authority*.

SURBL is a DNS blacklist against which any domains contained in the body of an email can be checked: most spam contains a link to a website, and by looking at the domain part of the URL and checking this against a database of known bad domains, a lot of spam can easily be identified. Using the DNS protocol, the *SURBL* database can be queried repeatedly, with very short response times.

Server Authority also checks for bad domains, but rather than checking the domain itself, it looks up the name server associated with the domain: identifying domains associated with name servers used by spammers is a proactive way of blocking email containing bad domains. *Server Authority* was not only applied to URLs but also to the EHLO/HELO domain, the reverse DNS of the sending IP address and the domain part of the MAIL FROM address.

It should be noted that, when it comes to finding domains in emails, there is no unique way of doing so. We searched the bodies of emails for strings matching certain regular expressions, but it is possible to use less strict regular expressions that would catch more URLs, to follow redirects, or even to search URLs contained inside images. This may have improved the spam catch rate, but at the cost of a higher server load, longer processing times and, possibly, more false positives.

Even with the settings used, the suite's spam catch rate was better than some traditional anti-spam solutions. Like those, however, it was not without fault and an apparently incorrectly listed *SURBL*-domain, as well as a small mistake in the way domains were read from emails, caused a total of five false positives. Still, with a final score that is higher than around half of the full solutions tested, it easily won a VBSpam award.

(Note: A small error in the way the *SURBL* server was queried, for which *VB* and *MXTools* share responsibility, meant that the suite's performance over the first four days of the testing period was slightly lower than it could have been; without this error, the final score could have been a few hundredths of a per cent higher.)

SPAMfighter Mail Gateway

SC rate (total): 97.60%

SC rate (Project Honey Pot corpus): 98.17%

SC rate (VB spam corpus): 91.85%

SC rate (image spam): 97.15%

SC rate (large spam): 92.44%

FP rate: 0.427%

Final score: 96.32%

SPAMfighter's developers made use of the feedback we gave them after the previous two tests not just to review the product's settings, but also to make some changes to the solution itself. These changes certainly had a positive effect: the product's performance improved and it earned another VBSpam award. There is still room for improvement though, and with the product's relatively poor performance on both large spam and image spam, the developers might want to look into these areas.

SpamTitan

SC rate (total): 99.65%

SC rate (Project Honey Pot corpus): 99.90%

SC rate (VB spam corpus): 97.13%

SC rate (image spam): 99.60%

SC rate (large spam): 98.59%

FP rate: 0.356%

Final score: 98.58%

SpamTitan, which runs as a virtual machine under *VMware*, had the highest spam catch rate in the last test and repeated that achievement in this test. The detailed results show that neither large spam nor spam containing images are a problem for the product and, as there were few false positives, it earns a VBSpam award with the second highest final score.

Sunbelt VIPRE Email Security

SC rate (total): 98.77%

SC rate (Project Honey Pot corpus): 99.08%

SC rate (VB spam corpus): 95.65%

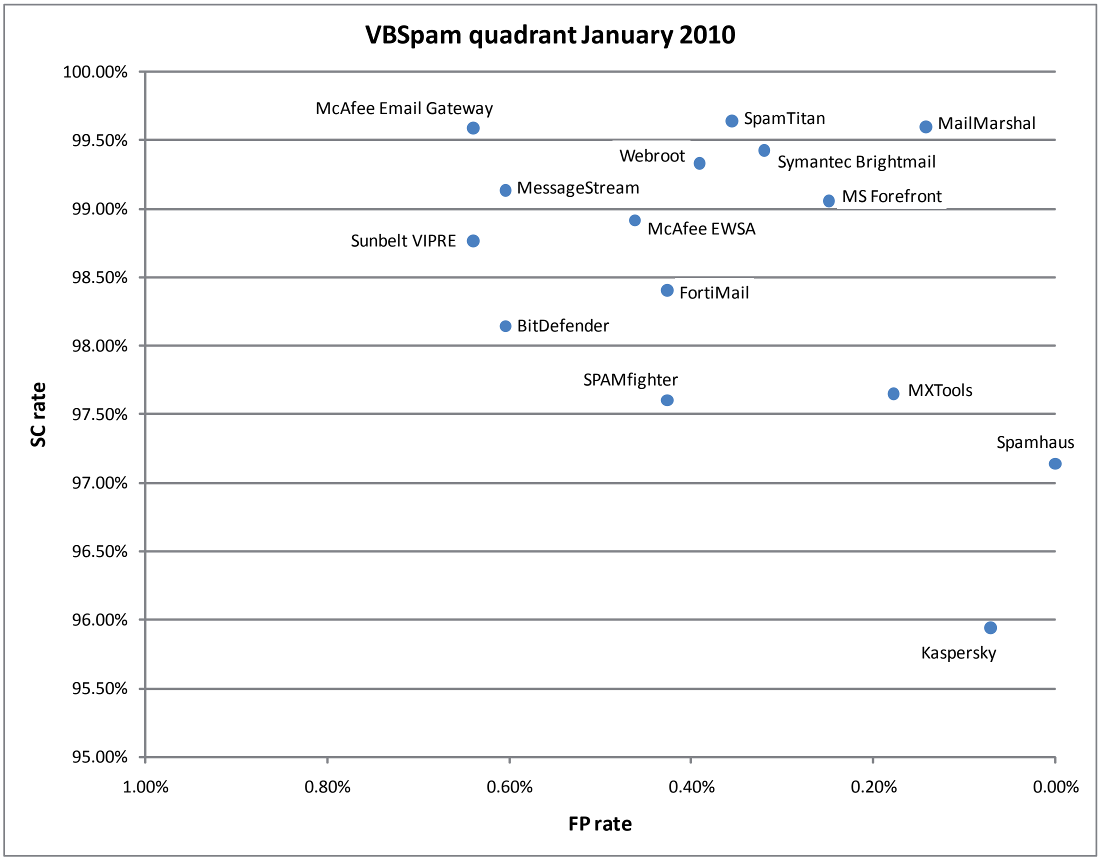
SC rate (image spam): 97.34%

SC rate (large spam): 94.56%

FP rate: 0.640%

Final score: 96.85%

Over the last few years, *Sunbelt* has become a big name in the world of computer security. Until recently, its anti-spam solution was known as *Ninja*, but, like its anti-malware solution, it is now known as *VIPRE*. The product runs alongside *Microsoft Exchange 2007* (which we ran on a *Windows Server 2003* machine), and once that has been installed,



the product is easy to set up and works almost immediately. While we ran the product mostly using its default settings, administrators have plenty of options to add, modify and disable anti-spam rules.

The product certainly had a good spam catch rate, although large spam and image spam are areas where there is some room for improvement. Its false positive rate was on the high side, but as the product was new to the test, this may well be the result of teething problems that may easily be solved by some modifications to the settings. In any case, the product won a VBSpam award with relative ease, and this should motivate the developers to perform even better next time.

Symantec Brightmail Gateway

SC rate (total): 99.43%

SC rate (Project Honey Pot corpus): 99.88%

SC rate (VB spam corpus): 94.88%

SC rate (image spam): 99.39%

SC rate (large spam): 96.66%

FP rate: 0.320%

Final score: 98.47%

As the world's largest vendor of security software, it is not surprising that *Symantec* offers a range of anti-spam solutions. One of these is *Brightmail*, which was acquired by *Symantec* in 2004.

Brightmail Gateway is available both as a hardware appliance and as a *VMware* virtual appliance; we tested the latter.

The product works well using its default settings, but it comes with an easy-to-use web interface where it can be fine tuned to meet the needs of an organization. Like more and more spam products, it can also be used for outbound filtering and company policies can be enforced on outgoing email: given the importance of email reputation this certainly seems a good idea.

In our test, we only looked at inbound filtering and *Brightmail* certainly does an excellent job there, catching all but just over 0.5% of spam. A few mailing list emails and some newsletters were incorrectly blocked, but that didn't stop the product from debuting with a VBSpam award and the third highest final score.

				Total spam			Final score
	True negative	False positive	FP rate	False negative	True positive	SC rate	
BitDefender	2794	17	0.605%	4581	242177	98.14%	96.33%
Fortinet FortiMail	2796	12	0.427%	3937	242821	98.40%	97.12%
Kaspersky	2809	2	0.071%	10026	236732	95.94%	95.73%
M86 MailMarshal	2807	4	0.142%	987	245771	99.60%	99.17%
McAfee Email Gateway	2789	18	0.640%	1001	245757	99.59%	97.67%
McAfee EWSA	2795	13	0.462%	2667	244091	98.92%	97.53%
MessageStream	2782	17	0.605%	2130	244628	99.14%	97.33%
MS Forefront	2804	7	0.249%	2318	244440	99.06%	98.31%
MXTools	2804	5	0.178%	5803	240955	97.65%	97.12%
SPAMfighter	2797	12	0.427%	5920	240838	97.60%	96.32%
SpamTitan	2801	10	0.356%	873	245885	99.65%	98.58%
Sunbelt VIPRE	2793	18	0.640%	3043	243715	98.77%	96.85%
Symantec Brightmail	2798	9	0.320%	1404	245354	99.43%	98.47%
Webroot	2796	11	0.391%	1639	245119	99.34%	98.17%
Spamhaus	2811	0	0.000%	7064	239694	97.14%	97.14%

Webroot E-Mail Security SaaS

SC rate (total): 99.34%

SC rate (Project Honey Pot corpus): 99.55%

SC rate (VB spam corpus): 97.23%

SC rate (image spam): 99.26%

SC rate (large spam): 97.31%

FP rate: 0.391%

Final score: 98.17%

Webroot's hosted solution saw its false positive rate reduced significantly in this test, while it also caught more spam. Its performance on the difficult-to-filter *VB* spam corpus was especially striking, and with a final score of well over 98%, the product earns another well-deserved VBSpam award.

vb VERIFIED SPAM virusbtn.com

Spamhaus ZEN plus DBL

SC rate (total): 97.14%

SC rate (Project Honey Pot corpus): 98.50%

SC rate (VB spam corpus): 83.47%

SC rate (image spam): 98.20%

SC rate (large spam): 94.64%

FP rate: 0.00%

Final score: 97.14%

Spamhaus (officially known as *The Spamhaus Project*) has been active for well over a decade and provides several DNS blacklists – databases of IP addresses known to be used by spammers. *Spamhaus ZEN* combines all three of the DNSBLs the organization provides and in this test, we combined it with *Spamhaus DBL*, which uses various heuristics to identify domains used by spammers. This DBL was checked for the domain part of every URL that appeared in the body of the emails – using the same method as used for *SURBL* and *Server Authority* – and also for the EHLO/HELO domain and the reverse DNS of the sending IP address.

vb VERIFIED SPAM virusbtn.com

Spamhaus has a rather conservative approach when it comes to adding IP addresses and domains to blacklists in order to minimize the number of false positives and, indeed, we did not see any false positives in this test. At the same time, the solution caught over 97% of the spam in this test, giving it a very good final score.

Still, the low catch rate for the *VB* spam corpus suggests that using *Spamhaus* on its own would lead to a fairly large number of spam messages reaching users' inboxes. This is why this is only a partial solution, the performance of

	Project Honey Pot spam		VB spam corpus		Image spam*		Large spam*	
	False negative	SC rate	False negative	SC rate	False negative	SC rate	False negative	SC rate
BitDefender	2556	98.86%	2025	90.94%	419	97.53%	209	94.84%
Fortinet FortiMail	2724	98.79%	1213	94.57%	369	97.83%	206	94.91%
Kaspersky	6386	97.15%	3640	83.71%	417	97.54%	256	93.67%
M86 MailMarshal	319	99.86%	668	97.01%	68	99.60%	71	98.25%
McAfee Email Gateway	355	99.84%	646	97.11%	91	99.46%	83	97.95%
McAfee EWSA	1154	99.49%	1513	93.23%	197	98.84%	208	94.86%
MessageStream	874	99.61%	1256	94.38%	144	99.15%	99	97.55%
MS Forefront	1534	99.32%	784	96.49%	129	99.24%	78	98.07%
MXTools	2668	98.81%	3135	85.97%	292	98.28%	208	94.86%
SPAMfighter	4098	98.17%	1822	91.85%	483	97.15%	306	92.44%
SpamTitan	232	99.90%	641	97.13%	68	99.60%	57	98.59%
Sunbelt VIPRE	2072	99.08%	971	95.65%	452	97.34%	220	94.56%
Symantec Brightmail	259	99.88%	1145	94.88%	104	99.39%	135	96.66%
Webroot	1021	99.55%	618	97.23%	125	99.26%	109	97.31%
Spamhaus	3370	98.50%	3694	83.47%	305	98.20%	217	94.64%

* There were 16,970 spam messages containing images and 4,047 considered large; the two are not mutually exclusive.

which should not directly be compared to that of full solutions. Still, even as a partial solution, it easily earns a VBSpam award.

CONCLUSION

This test saw several changes both to the way in which we measure results and to the make up of the email corpus. It is hoped that these changes will make it easier to translate the results to a real-world situation. We are working on some changes to the test set-up to make the next test even more realistic. In particular, we will be able to emulate a real situation where filters receive emails directly from the senders.

To achieve this, we will be able to send extra SMTP commands prior to the DATA command that inform the filter of the original sender's IP address and of their HELO/EHLO domain. For instance, this is possible in the Postfix MTA using the little known XCLIENT extension (http://www.postfix.org/XCLIENT_README.html), but we will be able to send different commands to different products. Using these commands, products will be able to block email pre-DATA (that is, before the actual email is sent) and the spam catch rate will be split into a pre-DATA rate and a post-DATA rate.

It should be noted that even in the current set-up, products have access to the original IP address and original HELO/EHLO domain. It will therefore not be mandatory for products to make use of these extended SMTP commands; we are well aware that for some products it may be harder, or even impossible, to change the way SMTP commands are dealt with. What will matter for the earning of a VBSpam award, as previously, are the total spam catch rate and the total false positive rate, regardless of how much (if anything) is blocked pre-DATA. It should, however, be an excellent opportunity for those products who want to boost their ability to block a large percentage of spam 'at the gate'.

The next VBSpam comparative review is set to run throughout February. The deadline for product submission will be 28 January 2010; any developers interested in submitting a product should contact martijn.grooten@virusbtn.com.