virus
BULLETIN

MARCH 2014 VBSPAM COMPARATIVE REVIEW

INTRODUCTION

One of the biggest security stories of the last few months was the data breach at US retail giant *Target*, which saw the credit card details of some 110 million customers end up in the hands of cybercriminals.

The breach involved a compromise of the company's internal system. It turns out that a security product *did* alert IT staff that something had gone amiss – however, the warnings were ignored. This proved to be a very serious mistake. Yet there is an important question that should be asked: how often did the security product give *false* alerts?

False positives aren't just bad because they can cause genuine events to be terminated: if they happen too often, people will fail to recognize legitimate alerts, and may even ignore them. (I once worked in a university building where we had so many false fire alarms that people would just shrug and continue with what they were doing when they heard the alarm, despite the fact that chemistry labs were located in the same building.)

It is no different for email – the odd false positive means missing the odd legitimate email, or at least receiving it with significant delay. However, it also increases the risk of true positives (spam correctly marked as such) being 'retrieved from quarantine', and this too can have expensive consequences – as demonstrated by the $60m breach of *RSA* back in 2011 (which involved a malicious email that, allegedly, had originally been blocked).

We sometimes refer to the VBSpam tests as 'anti-spam tests', because the prevalence of spam is the reason we run them, and the reason why people apply spam filters. But in fact, this is just as much a test of which products are best at avoiding quarantining legitimate emails.

In this month's test we focused a little more on products' ability (or inability) to recognize legitimate emails as such, by also taking into consideration how products handle 'newsletters' – updates and notifications from various companies and organizations that we had explicitly subscribed to. These emails have long been part of the VBSpam test, but this is the first time we have incorporated performance against this corpus into the final score.

Fifteen full solutions participated in this test. As in the last test, spam catch rates were high, but false positives were still an issue for some products, and two of them failed to achieve a VBSpam award (and would also have done under the old rules). Achieving a VBSpam+ award has become a little harder under the new rules, yet four products managed to do so.

THE TEST SET-UP

The VBSpam test methodology can be found at http://www.virusbtn.com/vbspam/methodology/. As usual, emails were sent to the products in parallel and in real time, and products were given the option to block email pre-DATA (that is, based on the SMTP envelope and before the actual email was sent). Four products chose to make use of this option.

For those products running on our equipment, we used *Dell PowerEdge* machines. As different products have different hardware requirements (not to mention those running on their own hardware, or those running in the cloud) there is little point comparing the memory, processing power or hardware the products were provided with; we followed the developers' requirements and note that the amount of email we receive is representative of that received by a smaller organization.

To compare the products, we calculate a 'final score' and, as indicated above, this month sees the introduction of a new, slightly modified formula. We have introduced a 'weighted

false positive' (WFP) rate, defined as the false positive rate of the ham and newsletter corpora taken together, with emails from the latter corpus having a weight of 0.2. Moreover, no more than 20% of the newsletters were counted as such.

In a formula:

WFP rate = (#false positives + 0.2 * min(#newsletter false positives , 0.2 * #newsletters)) / (#ham + 0.2 * #newsletters)

The final score is then defined as the spam catch (SC) rate minus five times the weighted false positive rate. Products earn VBSpam certification if this value is at least 98:

SC - (5 x WFP) ≥ 98

Meanwhile, products that combine a spam catch rate of 99.50% or higher with a lack of false positives *and* no more than 2.5% false positives among the newsletters earn a VBSpam+ award.

THE EMAIL CORPUS

The test ran for 16 days: from 12am on Saturday 22 February to 12am on Monday 10 March 2014.

Email as a protocol is designed to work under sub-optimal circumstances: temporarily shutting a mail server off from the Internet won't cause any long-term effects. Our test is no different, and when there is an issue with the network, or with the MTAs, it typically has no significant effect on the participating products.
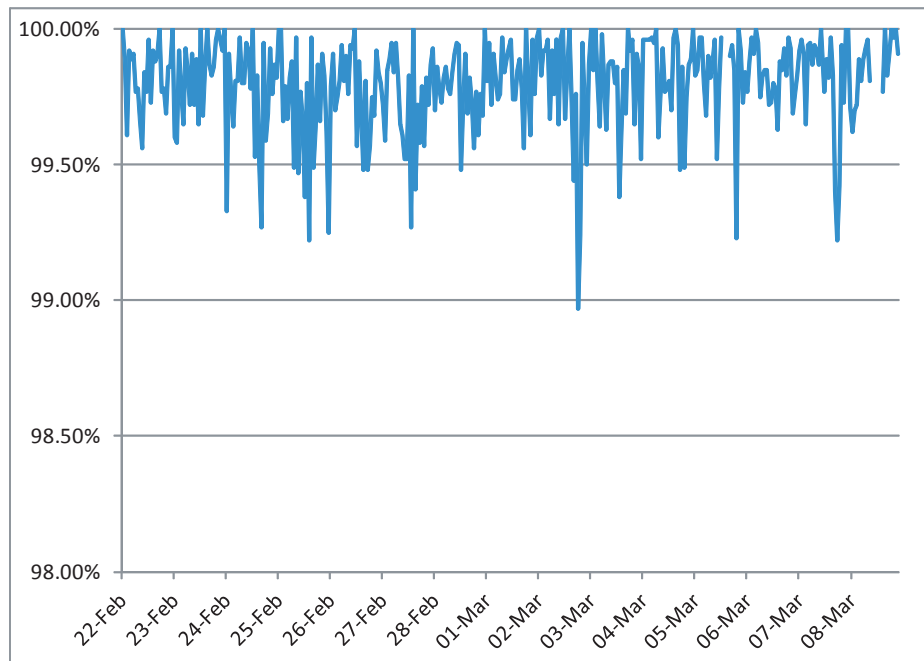
However, because we want the test to be run under optimal circumstances, we decided to exclude two periods from the test: a three-hour block on 5 March and a five-hour block on 8 March. During these periods, we couldn't guarantee that sub-optimal test conditions didn't affect the test, so they were excluded to err on the side of caution.

The corpus of emails sent during the test period consisted of 97,985 emails, 85,408 of which were spam. 72,953 of these were provided by *Project Honey Pot*, with the remaining 12,455 emails provided by *spamfeed.me*, a product from *Abusix*. They were all relayed in real time, as were the 12,225 legitimate emails ('ham') and 352 newsletters.

Figure 1 shows the catch rate of all full solutions throughout the test. To avoid the average being skewed by poorly performing products, the highest and lowest catch rates have been excluded for each hour.

As the graph shows, spam catch rates were very high once again. The average catch rate of those products that participated in both this and the last test decreased by 0.04%, but at 99.76% this is hardly an issue. The average false positive rate remained the same.

A NEW LOOK AT THE NEWSLETTER CORPUS

The newsletter corpus was introduced two and a half years ago, when it was described as containing 'non-personal emails that are sent with certain regularity after the recipient has subscribed to them'.

Many of these emails are of a direct commercial nature and provide product offers, or are of an indirect commercial nature and provide updates on websites that make money through advertisements. Other emails are sent with an ideological motive, or have the sole purpose of spreading information.

A lot of these newsletters are written in English, but many others aren't – to emphasize the global nature of the test, newsletters in more than two dozen languages have been subscribed to. Some of these newsletter services send emails

Figure 1: Spam catch rate of all complete solutions throughout the test period.

daily; others only send emails once a month or even less frequently[1].

But there is one feature the newsletters share: they have all explicitly been subscribed to. After all, *unsolicited* bulk emails are spam, no matter how much a lawyer may point out that some of them don't break any laws.

Now, there are subscriptions and there are subscriptions. It has long been considered good practice to confirm subscriptions by sending an email to the subscriber and only adding the address to the list if the subscriber confirms the action. In email circles, such newsletters are called 'confirmed opt-in'[2].

We'd love only to include such newsletters in the corpus, but unfortunately, that wouldn't be an accurate reflection of the real world: we found that around two-thirds of the newsletters we subscribed to didn't confirm subscriptions. We also noticed that those that do confirm subscriptions have a slightly smaller chance of being blocked.

As of this test, newsletters count towards the final score and determine the position of products in the VBSpam chart. In both, the 'weighted false positive' rate, introduced in the Test Set-Up section of this review, replaces the false positive rate: in the weighted rate, newsletters are considered part of the ham corpus, but with a weight of 0.2.

There are two reasons behind our decision to include newsletters. The first is that in recent tests, we have noticed the differences between products' performances becoming very small. This isn't a problem: there is more than one right solution for blocking spam. Yet we have noticed that small differences in performance are a strong motivation for products to work towards performing even better.

No one would disagree that, of two products with the same spam catch rate and the same false positive rate, the product which blocks the fewest *wanted* newsletters performs better, even if the difference is marginal.

The second reason is that newsletter false positives don't stand on their own: typically, a product that blocks more newsletters also blocks more spam, and vice versa. Especially among the harder to block spam, we tend to find many emails that look like newsletters – and possibly are newsletters, except that they were sent to an address that did not subscribe to them.

Thus, by including the newsletters in the calculation of the final score, a product that blocks more legitimate and illegitimate newsletters won't automatically see its final score increase.

---

[1] To avoid results being skewed, in each test corpus, we include no more than three newsletters from the same subscription.
[2] These are sometimes called 'double opt-in' – but this term is also used in a derogatory sense by those who oppose the idea of confirmed opt-ins.

Having said all this, we stress that, ultimately, the spam catch rate, the false positive rate and the newsletter false positive rate are three separate measures. While they are linked, combining them in a single formula always means making a choice. The choice that we made may not be that of a customer who wants to avoid false positives at any cost, or another who is willing to accept some false positives as long as a large amount of spam is blocked.

We thus encourage readers of the report to consider the various rates independently and draw their own conclusions about which products performed best according to their own criteria.

RESULTS

To allow readers of the report to compare results of this test to that of the last test, we have included the 'old-style final score'. The 'final score' as included here and in subsequent reports should not be compared with those in previous tests.

Axway MailGate 5.3.1

SC rate: 99.51%
FP rate: 0.60%
Final score: 96.22
Old-style final score: 96.52
Project Honey Pot SC rate: 99.46%
Abusix SC rate: 99.81%
Newsletters FP rate: 11.1%

I had hoped that *Axway* would score a lower false positive rate this time than in the last test, when it failed to win its second VBSpam award. Unfortunately, that wasn't to be: the *MailGate* product actually missed more legitimate emails than in the last test.

Looking at the legitimate emails that were missed, with many emails coming from the same senders, it looks like it should be possible for the product to find ways to improve its false positive rate. *MailGate* also missed one in nine newsletters and thus we weren't able to give the product a VBSpam award – noting that it wouldn't have achieved one under the old rules either.

Bitdefender Security for Mail Servers 3.1.2

SC rate: 99.95%
FP rate: 0.00%
Final score: 99.94
Old-style final score: 99.95
Project Honey Pot SC rate: 99.95%
Abusix SC rate: 99.93%
Newsletters FP rate: 0.3%

	True negatives	False positives	FP rate	False negatives	True positives	SC rate	Final score
Axway	12152	73	0.60%	418	84990	99.51%	96.22
Bitdefender	12225	0	0.00%	45	85363	99.95%	99.94
ESET	12224	1	0.01%	462	84946	99.46%	99.40
FortiMail	12224	1	0.01%	66	85342	99.92%	99.85
GFI	12224	1	0.01%	344	85064	99.60%	99.47
IBM	12223	2	0.02%	173	85235	99.80%	99.69
Kaspersky LMS	12225	0	0.00%	215	85193	99.75%	99.75
Libra Esva	12225	0	0.00%	26	85382	99.97%	99.92
Net At Work NoSpamProxy	12169	56	0.46%	357	85051	99.58%	97.06
Netmail Secure	12224	1	0.01%	339	85069	99.60%	99.51
OnlyMyEmail	12225	0	0.00%	1	85407	99.999%	99.93
Scrollout	12202	23	0.19%	70	85338	99.92%	98.41
Sophos	12210	15	0.12%	343	85065	99.60%	98.98
SpamTitan	12225	0	0.00%	128	85280	99.85%	99.76
ZEROSPAM	12222	3	0.02%	87	85321	99.90%	99.38
Spamhaus ZEN+DBL[*]	12224	1	0.01%	6276	79132	92.65%	92.59

[*] *Spamhaus* is a partial solution and its performance is not to be compared with that of other products.

(*Please refer to the text for full product names.*)

Bitdefender hasn't missed a single legitimate email in our tests since November 2012. This test was no different, though the new focus on newsletters shows that *Bitdefender* does make mistakes: it accidentally blocked a single newsletter, aimed at foodies.

With a spam catch rate that increased to 99.95%, *Bitdefender* achieves this month's highest final score. Just as impressively, the Romanian company is awarded its 30th VBSpam award (it being the only product to have been submitted to all 30 VBSpam tests) and its eighth VBSpam+ award in succession.

ESET Mail Security for Microsoft Exchange Server

SC rate: 99.46%
FP rate: 0.01%
Final score: 99.40
Old-style final score: 99.42
Project Honey Pot SC rate: 99.39%
Abusix SC rate: 99.86%
Newsletters FP rate: 0.6%

In this test, *ESET*'s spam catch rate took a bit of a dive. The product missed 462 spam emails and a quick glance at those it missed didn't reveal an obvious pattern – though no doubt the product's developers will be keen to spot one to improve performance in the next test.

However, a catch rate of 99.46% remains decent, and since the product only missed one legitimate email and two newsletters, it had no problem achieving its tenth VBSpam award in as many tests.

Fortinet FortiMail

SC rate: 99.92%
FP rate: 0.01%
Final score: 99.85
Old-style final score: 99.88
Project Honey Pot SC rate: 99.92%
Abusix SC rate: 99.94%
Newsletters FP rate: 1.1%

Fortinet's *FortiMail* appliance saw a nice improvement in what was already a good spam catch rate: in this test, the product missed only 66 spam emails. The false positive rate

	Newsletters		Project Honey Pot		Abusix		pre-DATA‡		STDev‡
	False positives	FP rate	False negatives	SC rate	False negatives	SC rate	False negatives	SC rate	
Axway	39	11.1%	394	99.46%	24	99.81%			0.48
Bitdefender	1	0.3%	36	99.95%	9	99.93%			0.26
ESET	2	0.6%	445	99.39%	17	99.86%			0.41
FortiMail	4	1.1%	58	99.92%	8	99.94%			0.28
GFI	11	3.1%	313	99.57%	31	99.75%			0.47
IBM	3	0.9%	170	99.77%	3	99.98%			0.41
Kaspersky LMS	0	0.0%	143	99.80%	72	99.42%			0.23
Libra Esva	6	1.7%	19	99.97%	7	99.94%	72585	84.99%	0.10
Net At Work NoSpamProxy	30	8.5%	299	99.59%	58	99.53%	42844	50.16%	0.54
Netmail Secure	6	1.7%	314	99.57%	25	99.80%	72545	84.94%	0.45
OnlyMyEmail	8	2.3%	0	100.00%	1	99.99%			0.03
Scrollout	99	28.1%	33	99.95%	37	99.70%			0.33
Sophos	1	0.3%	331	99.55%	12	99.90%			0.39
SpamTitan	11	3.1%	109	99.85%	19	99.85%			0.75
ZEROSPAM	49	13.9%	79	99.89%	8	99.94%			0.25
Spamhaus ZEN+DBL*	2	0.6%	5789	92.06%	487	96.09%	71111	83.26%	4.15

* *Spamhaus* is a partial solution and its performance is not to be compared with that of other products.

‡ pre-DATA filtering was optional and was applied on the full corpus. One of *NoSpamProxy*'s false positives occurred pre-DATA, all others occurred post-DATA.

‡ The standard deviation of a product is calculated using the set of its hourly spam catch rates.

(*Please refer to the text for full product names.*)

was reduced as well – *FortiMail* missed only one legitimate email – and thus it ended up with a final score that was significantly higher than in the last test. The product's 29th VBSpam award in as many tests is well deserved.

GFI MailEssentials

SC rate: 99.60%
FP rate: 0.01%
Final score: 99.47
Old-style final score: 99.56
Project Honey Pot SC rate: 99.57%
Abusix SC rate: 99.75%
Newsletters FP rate: 3.1%

GFI's *MailEssentials* saw a small decline in its spam catch rate, though at 99.60% there is little reason to worry. As in the previous test, a single legitimate email was missed, this one in German.

That false positive, along with 11 missed newsletters, prevented the product from winning its fourth VBSpam+ award. Interestingly, a *Facebook* notification was among the 'newsletter' emails missed. Regardless, *GFI* achieves its 18th VBSpam award after three full years of participating in these tests.

IBM Lotus Protector for Mail Security

SC rate: 99.80%
FP rate: 0.02%
Final score: 99.69
Old-style final score: 99.72
Project Honey Pot SC rate: 99.77%
Abusix SC rate: 99.98%
Newsletters FP rate: 0.9%

In this test, *IBM* demonstrated that its 99.80% spam catch rate in the last test wasn't a one-off. Moreover, the false

Hosted solutions	Anti-malware	IPv6	DKIM	SPF	DMARC	Multiple MX-records	Multiple locations
OnlyMyEmail	Proprietary (optional)		√	√	*	√	√
ZEROSPAM	ClamAV			√		√	√

*OnlyMyEmail verifies DMARC status, but doesn't provide feedback at the moment.

(*Please refer to the text for full product names.*)

Local solutions	Anti-malware	IPv6	DKIM	SPF	DMARC	Interface			
						CLI	GUI	Web GUI	API
Axway MailGate	Kaspersky; McAfee	√	√	√				√	
Bitdefender	Bitdefender	√				√		√	
ESET	ESET Threatsense					√	√		
FortiMail	Fortinet	√	√	√		√		√	
GFI	Five anti-virus engines	√		√				√	
IBM	Sophos; IBM Remote Malware Detection			√		√		√	
Kaspersky LMS	Kaspersky	√		√		√		√	
Libra Esva	ClamAV; others optional		√	√		√		√	
Net At Work NoSpamProxy	Commtouch			√			√		√
Netmail Secure	Proprietary	√	√	√		√		√	
Scrollout	ClamAV			√		√		√	
Sophos	Sophos							√	
SpamTitan	Kaspersky; ClamAV	√	√	√		√		√	√

(*Please refer to the text for full product names.*)

positive rate was reduced a little, with just two (related) legitimate emails having been blocked.

The product thus comes very close to achieving a VBSpam+ award, but for now, the industry giant achieves its 14th VBSpam award.

Kaspersky Linux Mail Security 8.0

SC rate: 99.75%

FP rate: 0.00%

Final score: 99.75

Old-style final score: 99.75

Project Honey Pot SC rate: 99.80%

Abusix SC rate: 99.42%

Newsletters FP rate: 0.0%

In the criteria for the VBSpam+ award, we allow for products to miss up to one in 20 newsletters; we feel that requiring a product not to block *any* legitimate newsletters would be too strict a criterion. However, *Kaspersky*'s *Linux Mail Security* achieved exactly that.

The spam catch rate dropped a little compared to the last test, but with only one in 400 spam emails missed, it remains high – and some similarities among the missed spam messages

suggest that the product may have been a little unlucky. In any case, the product's fourth VBSpam+ award is well deserved.

Libra Esva 3.2

SC rate: 99.97%

FP rate: 0.00%

Final score: 99.92

Old-style final score: 99.97

Project Honey Pot SC rate: 99.97%

Abusix SC rate: 99.94%

Pre-DATA SC rate: 84.99%

Newsletters FP rate: 1.7%

Libra Esva has earned a VBSpam+ award in each of the last four tests. As of this test, we set the bar a little higher, but the Italian product wasn't caught off its guard.

Yet again, it combined an impressive catch rate (99.97%) with a lack of false positives. The six false positives among the newsletters, written in three different languages, fall well below the 2.5% threshold and thus the product easily achieves yet another VBSpam+ award.

Complete solutions sorted by final score	
Bitdefender	99.94
OnlyMyEmail	99.93
Libra Esva	99.92
FortiMail	99.85
SpamTitan	99.76
Kaspersky LMS	99.75
IBM	99.69
Netmail Secure	99.51
GFI	99.47
ESET	99.40
ZEROSPAM	99.38
Sophos	98.98
Scrollout	98.41
Net At Work NoSpamProxy	97.06
Axway	96.22

(*Please refer to the text for full product names.*)

Net At Work NoSpamProxy

SC rate: 99.58%

FP rate: 0.46%

Final score: 97.06

Old-style final score: 97.29

Project Honey Pot SC rate: 99.59%

Abusix SC rate: 99.53%

Pre-DATA SC rate: 50.16%

Newsletters FP rate: 8.5%

Compared to the previous test (when it failed to win a VBSpam award), *Net At Work*'s *NoSpamProxy* saw its false positive rate increase by 0.08%. That's certainly a good thing, but unfortunately it wasn't good enough – the 56 legitimate emails it missed (all but one were written in English) pushed the final score below the VBSpam threshold. It would also have missed out on a VBSpam award under the 'old rules'.

Netmail Secure

SC rate: 99.60%

FP rate: 0.01%

Final score: 99.51

Old-style final score: 99.56

Project Honey Pot SC rate: 99.57%

Abusix SC rate: 99.80%

Pre-DATA SC rate: 84.94%

Newsletters FP rate: 1.7%

Yet again, *Netmail Secure* missed just a single legitimate email: a German-language message that a few other products also blocked erroneously. Once again, that prevented the virtual appliance from achieving a VBSpam+ award. The 300-odd missed spam emails in a variety of languages marked a small decrease compared to the last test though, and for now, *Netmail* earns its 19th VBSpam award.

OnlyMyEmail's Corporate MX-Defender

SC rate: 99.999%

FP rate: 0.00%

Final score: 99.93

Old-style final score: 99.999

Project Honey Pot SC rate: 100.00%

Abusix SC rate: 99.99%

Newsletters FP rate: 2.3%

The spam catch rates of *OnlyMyEmail*'s hosted *MX-Defender* solution never cease to impress: out of more than 85,000 spam emails sent during the test, it missed just one – a loan offer made in German. At the same time, the product didn't miss any legitimate email, which under the old rules would have resulted in a near-perfect final score.

Things have become a little more challenging under the new rules, but *OnlyMyEmail* was still able to earn a VBSpam+ award with ease. With eight missed newsletters it also achieved the second highest final score.

Scrollout F1

SC rate: 99.92%

FP rate: 0.19%

Final score: 98.41

Old-style final score: 98.98
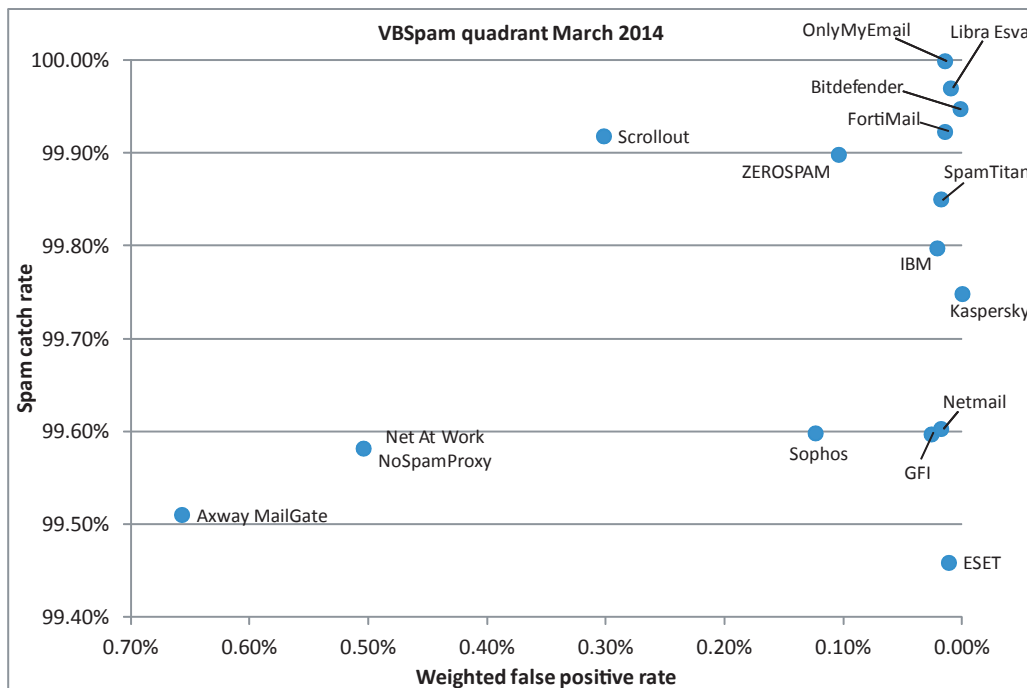
Project Honey Pot SC rate: 99.95%

Abusix SC rate: 99.70%

Newsletters FP rate: 28.1%

Open-source virtual appliance *Scrollout F1* saw its spam catch rate improve to a very decent 99.92% – the product missed just 70 spam emails, a significant number of which were finance-related tips or warnings.

False positives have always been a bit of a concern for *Scrollout* though, and in this test it missed 23 legitimate emails. It also missed 28.1% of the newsletters (the impact of this on the final score is somewhat dampened because we don't count more than 20% of the newsletters – a rule introduced to allow products to deliberately block newsletters without failing the test). *Scrollout* still achieved a decent final score however, and as a result earns its fifth VBSpam award.

VBSpam quadrant March 2014

(Please refer to text for full product names.)

Sophos Email Appliance

SC rate: 99.60%

FP rate: 0.12%

Final score: 98.98

Old-style final score: 98.98

Project Honey Pot SC rate: 99.55%

Abusix SC rate: 99.90%

Newsletters FP rate: 0.3%

Sophos's developers must have been looking forward to the inclusion of the newsletters in the final score, seeing as their appliance didn't block any of them in the last test. Things were a little different this time, though: not only did the product block a single newsletter (nothing to be ashamed of), but it also blocked 15 legitimate emails, although these were sent by only a handful of senders.

Perhaps it was just one of those months – as *Sophos* also missed twice as much spam as it did in the last test. But we shouldn't only consider things relative to performance in the last test: in absolute terms, the product's performance remained very decent and *Sophos* had little problem attaining its 25th VBSpam award in as many tests.

SpamTitan 6.00

SC rate: 99.85%

FP rate: 0.00%

Final score: 99.76

Old-style final score: 99.85

Project Honey Pot SC rate: 99.85%

Abusix SC rate: 99.85%

Newsletters FP rate: 3.1%

SpamTitan may be seen as the first 'victim' of the new rules. If it hadn't been for the 3.1% false positive rate among the newsletters (it missed 11 – of which, interestingly, only two were in English, the rest being written in Russian, Polish or Japanese), it would have earned another VBSpam+ award.

Instead, *SpamTitan* earns a standard VBSpam award this month – even though its newsletter false positive rate actually improved compared with the last test, as did both its normal false positive rate and its spam catch rate.

ZEROSPAM

SC rate: 99.90%

FP rate: 0.02%

Final score: 99.38

Old-style final score: 99.78

Project Honey Pot SC rate: 99.89%

Abusix SC rate: 99.94%

Newsletters FP rate: 13.9%

ZEROSPAM missed only one in 1,000 spam emails – a continuation of the hosted solution's record of high spam catch rates. It missed three legitimate emails, in as many languages. That was three too many, of course, but not too a big deal.

ZEROSPAM blocked more newsletters than all but one other product. Certainly, the ability to deliver legitimate newsletters isn't something that all customers want, and while this decreased the product's final score, it still easily achieved a VBSpam award.

Spamhaus ZEN+DBL

SC rate: 92.65%

FP rate: 0.01%

Final score: 92.59

Old-style final score: 92.61

Project Honey Pot SC rate: 92.06%

Abusix SC rate: 96.09%

Pre-DATA SC rate: 83.26%

Newsletters FP rate: 0.6%

Spamhaus's IP- and domain-based blacklists take an approach that is focused on generating very few false positives – and the solution (which is only a partial solution, and its performance thus shouldn't be compared to that of other participating solutions) has indeed had very few of them historically.

In this test, however, *Spamhaus* showed that very few false positives isn't the same as none, and the product erroneously blocked a single legitimate email and a newsletter. More positively, it saw its spam catch rate increase by 1.7 percentage points.

CONCLUSION

This test marks the end of our fifth year of VBSpam testing, during which time we've had more than 40 different anti-spam solutions on our test bench. The most important thing we have learned is that spam filters generally run very well – but of course, the devil is in the detail, and in the details some products are better than others. Having added some extra focus on false positives in this test, it was interesting to see which products found it easier to step up to the mark than others. We will continue to look for ways to add more information to these tests.

Here's to the next five years!

The next VBSpam test will run in April 2014, with the results scheduled for publication in May. Developers interested in submitting products should email martijn.grooten@virusbtn.com.